# Vorschlag 1:

Eyes on the Classroom - How Expertise Shapes Gaze Behavior and Classroom Management Measures

# Vorschlag 2:

Who Masters Classroom Disruptions - Expertise Differences in Gaze Behavior and Classroom Management Measures in a Laboratory Setting

# Abstract

*Keywords:* classroom management, professional vision, eye-tracking, expertise differences, managing classroom disruptions

# Introduction14201

## Classroom Management

## Teachers’ Professional Knowledge

## Expertise Differences in Teachers’ Professional Vision

## Eye-Tracking to Assess Teachers’ Professional Vision

Fixation Number: This metric reflected the frequency of visual attention shifts, with higher fixation numbers indicating a more dynamic scanning behavior across the classroom environment (Grub et al., 2020).

Average Duration Fixation: This metric provided a measure of cognitive processing, with longer durations suggesting more time spent processing visual information (Negi & Mitra, 2020).

The GRI provided a standardized measure of visual scanning efficiency, with smaller GRI values indicating a combination of shorter fixation durations and higher fixation frequencies, which is typically associated with more dynamic and efficient scanning behavior (Gegenfurtner et al., 2020).

## Present Study

The present study investigated differences in gaze behavior between experienced and inexperienced teachers during a micro-teaching unit that involved classroom disruptions. To examine these differences, multimodal data were analyzed from both in-service (experienced) and pre-service (inexperienced) teachers who participated in the laboratory-based study *Professional Vision of Novice and Expert Teachers (ProVisioNET).*

Participants individually attended a laboratory session, where they conducted a brief micro-teaching unit lasting approximately 15 minutes. The “class” consisted of three trained actors who portrayed students and simulated typical classroom disruptions. During the micro-teaching unit, teachers’ gaze patterns were recorded using eye-tracking technology. Additionally, self-reports on classroom management and strategic knowledge were assessed through a questionnaire, an interview, and a test.

The study addressed five primary aims:

One aim was to examine how teachers allocate their visual attention during classroom interactions in a lab-based micro-teaching unit. We expected students to be the primary focus of teachers’ attention. This exploration was descriptive.

Another aim was to investigate differences in gaze behavior related to noticing abilities between experienced and inexperienced teachers. We hypothesized that, compared to inexperienced teachers, experienced teachers would demonstrate more efficient gaze patterns throughout the micro-teaching unit (**Hypothesis 1a**), focus their attention more frequently on students (**Hypothesis 1b**), and notice disruptions more quickly (**Hypothesis 1c**).

Additionally, we examined whether the type of disruption (i.e., verbal disruptions, physical disruptions, and indicators of lack of eagerness to learn) influenced the speed at which teachers noticed disruptions. We hypothesized that verbal and physical disruptions would be more salient and, therefore, noticed more quickly than indicators of a lack of eagerness to learn (**Hypothesis 2a**). Furthermore, we explored whether teaching expertise had an impact on the disruption noticing speed.

Another aim was to determine whether differences in teaching expertise were reflected in classroom management measures, including self-reports on classroom management (e.g., self-evaluations of classroom management, disruptiveness, and confidence ratings) as well as strategic knowledge of classroom management. We expected experienced teachers to outperform inexperienced teachers (**Hypothesis 3a**).

Finally, we examined the relationship between gaze behavior and classroom management measures, expecting these variables to be correlated.

# Method

## Participants

We recruited a total of 84 teachers from Germany (42 pre-service teachers and 42 in-service teachers) through personal contacts, email lists, and flyers. Pre-service teachers were required to be actively enrolled in a teacher education program and to have completed their first internship, while in-service teachers needed to have completed both the teacher training program (2. Staatsexamen) and to be currently working in the teaching profession. Data from two in-service teachers were excluded due to low-quality eye-tracking data, resulting in a final sample of 82 teachers, comprising 42 pre-service teachers and 40 in-service teachers.

The pre-service teachers (*n*= 29 women, *n*= 13 men) had a mean age of 22.80 years (*SD* = 1.90; range: 19–27). On average, they were in their 7th semester (*M* = 6.70, *SD* = 2.60; range: 3–11) and had an average of 9.60 hours (*SD* = 7.20; range: 1–36) of teaching experience through internships completed during their studies.

The in-service teachers (*n*= 24 women, *n*= 16 men) had a mean age of 39.10 years (*SD* = 10.60; range: 26–60) and an average of 11.60 years (*SD* = 11.30; range: 1–38) of teaching experience.

The study adhered to ethical guidelines and received approval from the University’s Institutional Review Board. Participants were fully informed about the study’s objectives before testing. Their participation was voluntary, without incentives, and commenced only after written consent.

## Setting and Procedure

Participants individually attended the lab for approximately two hours, following a standardized procedure for which a seminar room was transformed into a classroom. Upon arrival, they were welcomed by the experimenter, introduced to the procedure, and asked to sign the data protection agreement. Participants were then fitted with a binocular Tobii Pro Glasses 2 eye-tracker, adjusted for comfort and vision. After performing an initial calibration of the glasses (for details of the eye-tracker and the calibration, see Appendix A), the experimenter activated and synchronized the recording devices (eye-tracking glasses, four cameras, and an audio recorder) using an auditory signal. This setup phase included a brief introductory game to acclimate participants to the eye-tracking equipment and the three students, which took approximately 10-15 minutes.

After the initial setup, a second calibration was done in a separate room. As soon as the teacher re-entered the classroom, the micro-teaching unit started. Participants were asked to prepare a 15-minute micro-teaching unit on a topic and grade level of their choice. The only requirement was that the unit had to be an introductory micro-teaching unit, and had to consist of supervised individual work and/or frontal teaching. During the unit, three trained actors (playing students) performed scripted classroom disruptions. The students followed prompts that appeared every 1.5 minutes on a screen only visible to them (e.g., chatting with a neighbor, heckling, looking at the phone; see Table A1 in the supplementary material for an overview and categorization of all disruptions; and Figure B1 and B2 in the supplementary material for a depiction of the laboratory setting of the micro-teaching unit). The order of the disruptions and the performing students were fully balanced using Latin Squares. The whole micro-teaching unit was recorded using eye-tracking glasses to capture teachers’ gaze patterns. The micro-teaching unit lasted about 15-20 minutes.

After the teaching session, participants underwent a third calibration. Following this, all recording devices were stopped, and participants filled out a brief computer-based questionnaire (~10-15 minutes) assessing sociodemographic data and a self-evaluation of their classroom management during the micro-teaching unit. A ten-minute break followed, concluding the first part of data collection.

In the study’s second part, participants engaged in a Stimulated Recall Interview (SRI). They watched a video of their own teaching session, recorded through the eye-tracking glasses, while the experimenter paused the video at each classroom disruption. Participants answered five open-ended questions and three rating questions for each disruption, including self-evaluated *disruptiveness* and *confidence ratings* (see Measures). The SRI lasted approximately 45-60 minutes. Finally, participants completed a Situational Judgment Test (SJT) online, assessing their *strategic knowledge of classroom management*. The questionnaire took approximately 15 minutes to complete, marking the end of the study.

## Measures

***Gaze Behavior Measures***

Gaze behavior was analyzed using predefined Areas of Interest (AOI) to examine how participants allocated their visual attention during the micro-teaching unit[[1]](#footnote-2). Two types of AOIs were defined to structure the data meaningfully: global AOI and event-based AOI. Global AOI were used to assess gaze behavior over the entire video duration and included the following: the *students*, representing gaze points focused on the group of three students; *teacher material*, capturing gaze points directed toward instructional materials such as the board, screen, or other teaching aids; *student desks*, representing gaze points on elements related to students’ desks, including name tags and student materials; and *classroom/others*, which encompassed gaze points directed toward other areas of the classroom that were not associated with the students or teacher. In contrast, the event-based AOI was coded only during specific classroom events, i.e., disruptions caused by students. The AOI labeled *disruptive student* captured gaze behavior directed toward one of the three students who performed a classroom disruption.

#### Average Fixation Duration on AOI.

The average fixation duration was calculated to determine the mean time participants spent focusing on individual elements. This was computed by dividing the total duration of all fixations, measured in milliseconds, by the total number of fixations recorded during the session. Fixation data for these calculations were aggregated across all global AOI, offering a comprehensive overview of participants’ monitoring behavior throughout the micro-teaching unit for each person and again, summarized for both groups.

#### Fixation Number per Minute on AOI *Students*.

The fixation number per minutewas calculated as the total number of fixations recorded during the session, standardized as fixations per minute to account for varying session durations. Specifically, the total fixation count was summed across all global AOI, including students, teacher material, student desks, and other classroom areas. This sum was then divided by the session duration in minutes to yield the fixation number per minute for each person and was then summarized for both groups (experienced and inexperienced teachers).

#### Gaze Relational Index.

The Gaze Relational Index (GRI) was calculated as a composite metric to evaluate the efficiency of participants’ scanning behavior during the micro-teaching unit. The GRI was derived by dividing the average fixation duration (in milliseconds) by the fixation number per minute.

#### Time to First Fixation on AOI *Disruptive Student*.

Thetime to first fixation (TTFF) was used to assess gaze behavior within the AOI *disruptive student* and to measure the disruption noticing speed, e.g., the latency, in seconds, for participants to direct their first fixation to the AOI after the onset of the disruption. Therefore, values were extracted from the eye-tracking data by identifying the timestamp when the first fixation on the AOI *disruptive student* occurred relative to the onset of the disruption. Data were filtered to include only valid fixation times, excluding instances where the TTFF was zero[[2]](#footnote-3) or exceeded 30 seconds.[[3]](#footnote-4) These raw TTFF values were then converted from milliseconds to seconds for interpretability. To calculate participant-level averages, the TTFF values were aggregated across all disruptions per participant. This yielded an average TTFF in seconds for each participant, reflecting their overall responsiveness to classroom disruptions.

### Self-Evaluation of Classroom Management

After the micro-teaching unit, teachers answered a questionnaire using five items from a validated questionnaire (Helmke et al., 2013) and eleven self-developed items derived from the research literature (Kiel et al., 2013; Kounin, 2006; Marzano, 2007) to assess teachers’ self-evaluated classroom management competencies. The questionnaire was a 4-point Likert scale (1 = strongly disagree; 4 = strongly agree).

### Subjective Disruptiveness and Confidence Ratings

The subjective disruptiveness and confidence ratings were assessed during the SRI on an 11-point rating scale, ranging from 0 (not at all disrupting/confident) to 10 (extremely disrupting/confident). Ratings were averaged across the nine classroom disruptions for each participant to capture a general sense of how disruptive the classroom disruptions were during the micro-teaching unit and how confident participants felt in handling them.

### Strategic Knowledge of Classroom Management

Teachers’ strategic knowledge of classroom management was assessed using a Situational Judgment Test (SJT; Gold & Holodynski, 2015) via an online questionnaire on SoSci Survey. Participants graded five to six action alternatives for twelve teaching scenarios in which classroom disruptions were discussed on a six-point Likert scale (grade 1 = “very good” to grade 6 = “unsatisfactory”). As the SJT was originally designed for primary schools, adjustments were made to enable the use of the SJT for all types of schools in the *ProVisioNET* study. Due to their general applicability, all twelve scenarios and answer options were adopted and only the names of the class levels were removed from the questions - except for scenario 6, where this information was essential.

## Data analysis

The data were analyzed using R (RStudio Team, 2020, Version 2024.12.0) and IBM SPSS Statistics (Version 29). Graphics were created using ggplot2 (Wickham, 2016).

To address our first research question, we descriptively analyzed the fixation durations[[4]](#footnote-5) for each AOI (*Students*, *Disruptive Student*, *Teacher Material*, *Student Desks*, and *Classroom/Others*) during the micro-teaching unit. Proportions of fixation time were calculated relative to total fixation duration, with the AOI *Disruptive Student* analyzed separately for proportional comparisons. Mean percentages were computed for experienced and inexperienced teachers and visualized using bar graphs. Independent-sample *t*-tests assessed statistical differences for each AOI, with effect sizes reported as Cohen’s *d* (Cohen, 1988).

To test **Hypothesis 2**, which examined global and event-related gaze efficiency, we calculated for experienced and inexperienced teachers the GRI for the entire micro-teaching unit (**Hypothesis 2a**), the number of fixations per minute on the AOI *Students* (**Hypothesis 2b**), and the TTFF for the AOI *Disruptive Student* to assess the disruption noticing speed (**Hypothesis 2c**). Independent-sample *t*-tests were used to compare the group means and Cohen’s *d* (Cohen, 1988) was calculated to determine the effect size.

To test **Hypothesis 3**, which stated that the disruption category affects teachers’ noticing speed, a 2 × 3 repeated-measures Analysis of Variance was conducted. The dependent variable, TTFF (representing the disruption noticing speed), was converted to seconds and log-transformed to correct skewness and enhance interpretability. Expertise (experienced vs. inexperienced teachers) served as a between-subject factor, while disruption category (verbal, physical, and lack of eagerness) was a within-subject factor. The analysis examined main effects of disruption category (**Hypothesis 3a**) and expertise (**Hypothesis** **3b**), as well as their interaction (**Hypothesis 3c**). Generalized eta-squared (*η²*) quantified effect sizes and Bonferroni-adjusted post-hoc comparisons were conducted for significant effects. Cohen’s *d* (Cohen, 1988) was calculated for effect size estimation.

To examine differences in conventional classroom management measures between experienced and inexperienced teachers (**Hypothesis 4**), independent-sample *t*-tests were conducted. Mean differences were analyzed for self-evaluations of disruption prevention and management competencies (**Hypothesis 4a**), ratings of how disruptive the disruptions were (**Hypothesis 4b**), how confident teachers felt in handling them (**Hypothesis 4c**), and their strategic knowledge of classroom management (**Hypothesis 4d**). For each comparison, Cohen’s *d* (Cohen, 1988) was calculated.

To address the fifth research goal, we analyzed the relationship between global and event-related gaze efficiency and classroom management measures using Pearson correlation coefficients. Specifically, we examined correlations between GRI values (**Hypothesis 5a**), fixation number per minute on students (**Hypothesis 5b**), and disruption noticing speed (**Hypothesis 5c**) and classroom management measures, including self-evaluations of disruption prevention and management competencies, ratings of how disrupted teachers felt by classroom disruptions, confidence in handling them, and their strategic knowledge of classroom management. Significance was assessed using *p*-values, and effect sizes were interpreted based on correlation coefficients.

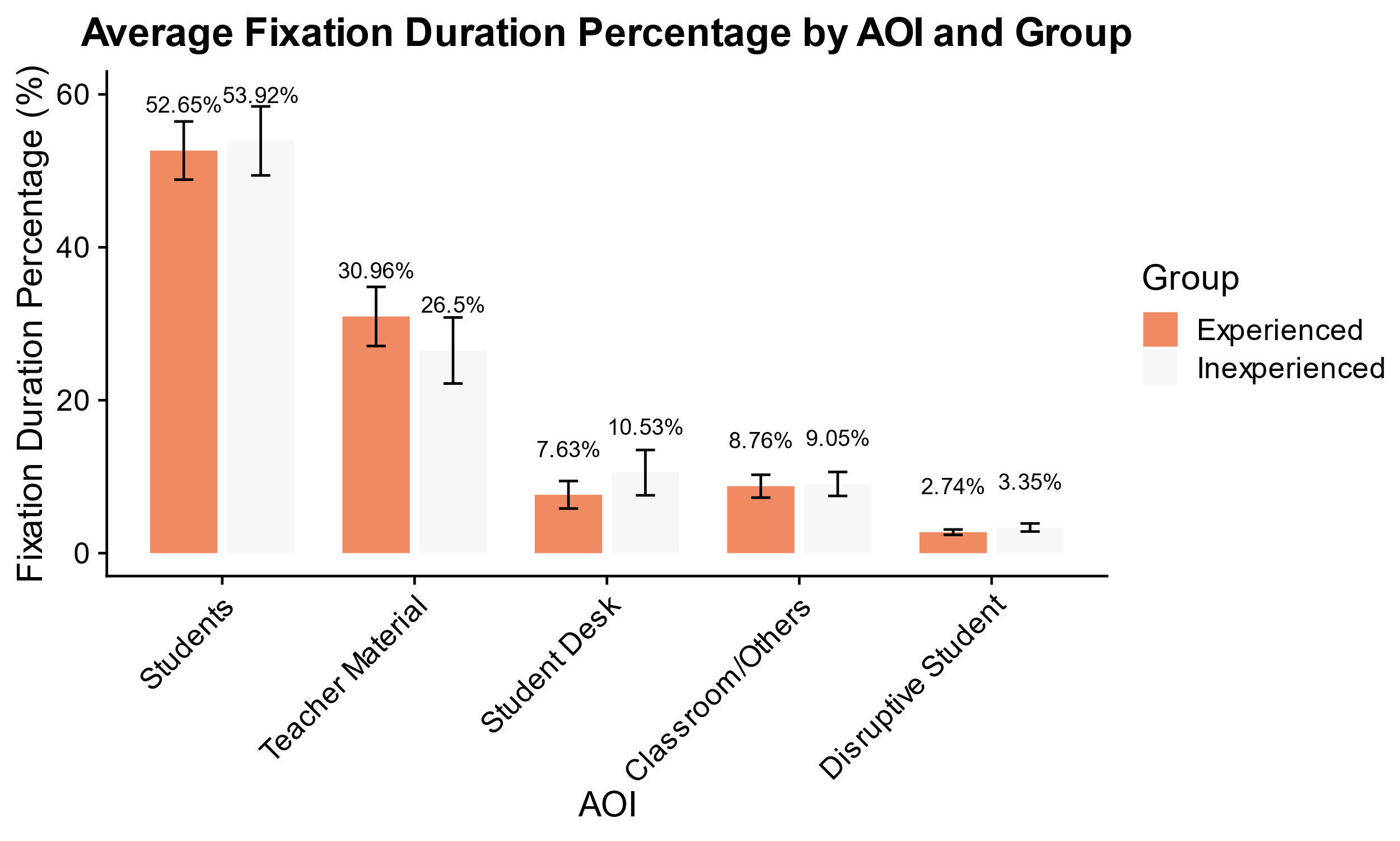
# Results

## Gaze Behavior Across Areas of Interest

To investigate teachers’ gaze behavior, we calculated the percentage of fixation durations directed at each AOI during the micro-teaching unit. The descriptive results for experienced and inexperienced teachers are displayed in Figure 1.

**Figure 1**

*Fixation Duration Percentages by Area of Interest (AOI) and Teacher Experience Group for the Entire Micro-Teaching Unit with 95% Confidence Intervals*

******

*Note.* The bar chart displays the percentage of fixation durations across five AOIs (Students, Teacher Material, Student Desk, Classroom/Others, and Disruptive Student) during the micro-teaching unit. Data are presented separately for experienced and inexperienced teachers, with error bars indicating 95% confidence intervals.

As suggested, both groups exhibited the highest percentages of fixation durations in the AOI *Students*, while the AOI *Disruptive Student* received the lowest values. No statistically significant differences were observed between experienced and inexperienced teachers across any AOI (see Appendix C, Table C2 for *t*-test results and effect sizes).

## Global and Event-related Gaze Efficiency

Means, standard deviations, and range of experienced and inexperienced teachers’ global and event-related gaze efficiency measures are shown in Table 1.

**Table 1**

*Means, Standard Deviations, And Range of Experienced and Inexperienced Teachers’ Global and Event-related Gaze Efficiency Measures*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Gaze Efficiency Measures | Experienced | | | | Inexperienced | | | |
| *M* | *SD* | Min | Max | *M* | *SD* | Min | Max |
| GRI for micro-teaching unit | 5.32 | 2.21 | 2.22 | 11.72 | 6.08 | 2.15 | 2.14 | 10.80 |
| Fixation Numbers in AOI *Students* | 98.53 | 19.21 | 49.07 | 135.81 | 91.21 | 14.52 | 68.99 | 123.37 |
| TTFF (log) in AOI *Disruptive Student* | 0.28 | 0.69 | –1.81 | 1.51 | 0.30 | 0.65 | –1.06 | 2.21 |
| *Note*. GRI = Gaze Relational Index, AOI = Area of Interest, TTFF = Time to First Fixation (log-transformed). Values represent the mean (*M*), standard deviation (*SD*), and range (minimum and maximum values) for each measure. Fixation numbers in AOI *Students* refer to the total number of fixations directed at students per minute. TTFF values represent the log-transformed time (in seconds) until teachers first fixated on a disruptive student. | | | | | | | | |

First, regarding gaze efficiency, experienced teachers exhibited more frequent but shorter fixations, leading descriptively to a lower GRI than inexperienced teachers across the entire micro-teaching session (**Hypothesis 2a**). However, this difference was not statistically significant, *t*(80) = –1.57, *p* = .12.

Second, experienced teachers directed their gaze toward AOI *Students* more often than inexperienced teachers (**Hypothesis 2b**), and this difference was statistically significant, *t*(80) = 1.96, *p* = .05, *d* = 0.43 (small effect).

Third, in terms of event-related gaze efficiency, experienced teachers detected disruptions descriptively slightly faster, as reflected in their shorter noticing speed for the AOI *Disruptive Student*[[5]](#footnote-6) (**Hypothesis 2c**), but with no statistically significant difference, *t*(80) = –0.14, *p* = .89.

## Effect of Disruption Category and Expertise on Disruption Noticing Speed

A 2 × 3 repeated-measures ANOVA tested **Hypothesis 3**, examining the effect of disruption category and expertise on log-transformed TTFF.

For **Hypothesis 3a**, the main effect of disruption category was significant, *F*(1.94, 141.49) = 68.05, *p* < .05, *η²* = .34. Post-hoc comparisons revealed that verbal disruptions were detected significantly faster than both physical disruptions, *t*(73) = 6.33, *p* < .05, *d* = 0.62 (large effect), and lack of eagerness disruptions, *t*(73) = 11.09, *p* < .05, *d* = 1.23 (large effect). Additionally, physical disruptions were noticed faster than lack of eagerness disruptions, *t*(73) = -5.72, *p* < .05, *d* = 0.62 (large effect).

For **Hypothesis 3b**, the main effect of expertise was not significant, *F*(1, 73) = 0.03, *p* = .86, indicating no difference in noticing speed between experienced and inexperienced teachers.

For **Hypothesis 3c**, the interaction effect between expertise and disruption category was not significant, *F*(1.94, 141.49) = 1.28, *p* = .28, suggesting that the disruption category affected both groups similarly.

## Expertise Differences in Classroom Management Measures

Table 2 presents a summary of the classroom management measures, including group means, standard deviations, range, and internal consistency reliabilities (ω), for each variable across experienced and inexperienced teachers.

**Table 2**

*Means, Standard Deviations, Range and Internal Consistency Reliability (McDonalds’ Omega, ω) of Experienced and Inexperienced Teachers’ Classroom Management Measures*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classroom Management Measures | Experienced | | | | Inexperienced | | | | ω |
| *M* | *SD* | Min | Max | *M* | *SD* | Min | Max |
| Self-Evaluation | 3.25 | 0.31 | 2.38 | 3.88 | 3.06 | 0.30 | 2.56 | 3.62 | .82 |
| Disruption Rating | 4.81 | 2.90 | 0 | 10 | 5.55 | 2.81 | 0 | 10 | .84 |
| Confidence Rating | 8.44 | 1.68 | 2 | 10 | 7.18 | 2.04 | 0 | 10 | .88 |
| Strategic Knowledge Test | 0.75 | 0.08 | 0.54 | 0.88 | 0.72 | 0.13 | 0.10 | 0.91 | .89 |
| *Note*. Values represent the mean (*M*), standard deviation (*SD*), and range (minimum and maximum values), along with McDonald’s Omega (ω) coefficient, which indicates the internal consistency reliability for each measure. Self-evaluation refers to participants’ assessment of their disruption prevention and management competencies. Disruption Rating indicates the extent to which participants judged the behaviors as disruptive, while Confidence Rating measures their confidence in managing these disruptions. The Strategic Knowledge Test measures knowledge of classroom management strategies. | | | | | | | | | |

The *t*-test for self-evaluation of disruption prevention and management competencies (**Hypothesis 4a**) revealed a significant difference, *t*(80) = 2.78, *p* < .05, *d* = 0.62 (large effect).

Results of the disruption ratings (**Hypothesis 4b**) showed a significant difference, *t*(80) = –2.57, *p* < .05, *d* = –0.57 (medium effect), as well as the confidence ratings (**Hypothesis 4c**; *t*(80) = 5.63, *p* < .05, *d* = 1.24 (large effect).

The situational judgment test (SJT) scores (**Hypothesis 4d**), assessing strategic knowledge, did not show a significant difference between groups, *t*(80) = 1.00, *p* = .32.

## Correlation Between Gaze Efficiency Measures and Classroom Management Measures

Table 3 presents Pearson correlation coefficients between the Gaze Relational Index (GRI) as a key gaze efficiency measure and classroom management measures, including self-evaluation of disruption prevention and management competencies, disruption ratings, confidence ratings, and strategic knowledge test scores, separately for experienced (below the diagonal) and inexperienced teachers (above the diagonal).[[6]](#footnote-7)

**Table 3**

*Correlations Between the Gaze Efficiency Measure and Classroom Management Measures for Experienced (Below Diagonal) and Inexperienced (Above Diagonal) Teachers*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| (1) GRI | – | –.44\* | .11 | –.01 | .16 | .15 | –.14 |
| (2) Fixation Number Per Minute on AOI *Students* | –.53\* | – | –.39\* | .39\* | –.49\* | –.22 | .14 |
| (3) TTFF (log) on AOI *Disruptive Student* | .05 | –.29 | – | –.20 | .32\* | –.01 | .01 |
| (4) Disruption Rating | .03 | –.20 | .02 | – | –.02 | .01 | –.15 |
| (5) Confidence Rating | 0 | .06 | –.06 | –.51\* | – | .48\* | –.19 |
| (6) Self-Evaluation | .11 | 0 | –.18 | –.11 | .33\* | – | –.21 |
| (7) Strategic Knowledge | –.36\* | .16 | –.34\* | –.19 | –.03 | .12 | – |
| *Note.* The table presents Pearson correlation coefficients between gaze efficiency measures (average fixation duration, fixation number per minute, GRI (= Gaze Relational Index), fixation number per minute on AOI *Students,* and TTFF (= log-transformed time to first fixation on AOI *Disruptive Student*), and classroom management measures (self-evaluation of teachers’ classroom management, disruptiveness ratings, confidence ratings, and strategic knowledge) for inexperienced teachers (above diagonal) and experienced teachers (below diagonal).  Statistically significant correlations (*p* < .05) are marked with an asterisk (\*).  Negative values indicate inverse relationships.  GRI = Gaze-Relational Index; TTFF = Time to First Fixation on AOI *Disruptive Student* (log-transformed). | | | | | | | | |

For inexperienced teachers, GRI was not significantly correlated with classroom management measures. Confidence ratings showed a significant positive correlation with self-evaluation. For experienced teachers, GRI was significantly negatively correlated with strategic knowledge test scores. Confidence ratings were significantly negatively correlated with disruption ratings, and significantly positively correlated with self-evaluation.

# Discussion

## Key Findings

Our study investigated …

Overall, our findings indicate that …

Our findings are consistent with prior research that illustrates the …

## Limitations and future directions

While the laboratory setting of the study allowed for a controlled implementation of stressors and high internal validity, it was not an authentic classroom environment, raising questions about its external validity. Most importantly, the teacher and their students did not have a shared history, and only a very thin basis for establishing a positive teacher-student relationship, which is a core characteristic of effective classroom management (Beaty-O’Ferrall et al., 2010; Rüedi, 2014).

In addition, the micro-teaching unit was only about 15 minutes long, and thus much shorter than a regular school lesson, providing less opportunities for experienced teachers to build up an engaging lesson.

Finally, the onset of disruptive student behavior was scripted, following an experimental time schedule, which was not affected by the behavior of the teacher. Thus, the setting may have masked effects of teaching experience by providing too little opportunities of experienced teachers to demonstrate their true classroom management skills.

In subsequent studies, it would therefore be insightful to …

## Conclusion

This study investigated …

In summary, our study contributes to the understanding …

# References

Beaty-O’Ferrall, M. E., Green, A., & Hanna, F. (2010). Classroom Management Strategies for Difficult Students: Promoting Change through Relationships. *Middle School Journal (J1)*, *41*(4), 4–11.

Cohen, J. (1988). Statistical power for the behavioural sciences. Hilsdale. *NY: Lawrence Erlbaum*, *58*(1), 7–19.

Gegenfurtner, A., Boucheix, J.-M., Gruber, H., Lehtinen, E., & Lowe, R. K. (2020). *The Gaze Relational Index as a Measure of Visual Expertise*. *3*.

Gold, B., & Holodynski, M. (2015). Development and Construct Validation of a Situational Judgment Test of Strategic Knowledge of Classroom Management in Elementary Schools. *Educational Assessment*, *20*(3), 226–248. https://doi.org/10.1080/10627197.2015.1062087

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*. https://doi.org/10.25656/01:21187

Helmke, A., Schrader, F.-W., Helmke, T., Lenske, G., Pham, G., Praetorius, A.-K., & Ade-Thurow, M. (2013). *Basisfragebogen EMU - Evidenzbasierte Methoden der Unterrichtsentwicklung*. Unterrichtsdiagnostik. http://www.unterrichtsdiagnostik.info/downloads/fragebogen/

Kiel, E., Frey, A., & Weiß, S. (2013). *Trainingsbuch Klassenführung*. Verlag Julius Klinkhardt.

Kounin, J. S. (2006). *Techniken der Klassenführung*. Waxmann Verlag.

Lohmann, G., & Meyer, H. (2003). *Mit Sch"ulern klarkommen: Professioneller Umgang mit Unterrichtsst"orungen und Disziplinkonflikten*. Cornelsen-Scriptor.

Marzano, R. J. (2007). *The Art and Science of Teaching: A Comprehensive Framework for Effective Instruction*. ASCD.

Onkhar, V., Dodou, D., & de Winter, J. C. F. (2024). Evaluating the Tobii Pro Glasses 2 and 3 in static and dynamic conditions. *Behavior Research Methods*, *56*(5), 4221–4238. https://doi.org/10.3758/s13428-023-02173-7

RStudio Team. (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC.

Rüedi, J. (2014). Zur Bedeutung positive Beziehungen für die Klassenführung und den Umgang mit Unterrichtsstörungen. *Beziehungen in Schule Und Unterricht. Teil*, *3*, 105–126.

Tobii AB. (2024). *Tobii Pro Lab User Manual v 24.21*. https://go.tobii.com/tobii\_pro\_lab\_user\_manual

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://doi.org/10.1007/978-0-387-98141-3

# Appendix

# Appendix A

**Eye-tracking apparatus and calibration**

Teachers wore a binocular Tobii Pro Glasses 2 eye-tracker during the micro-teaching unit to record eye-tracking data. The system consisted of a wearable head unit and a recording unit. The head unit was a measuring device with different sensors. A high-definition scene camera captured a full HD video of the teacher’s field of vision. An integrated microphone recorded surrounding sounds. Infrared light illuminators supported the eye-tracking sensors which recorded the eye orientation to capture the teacher’s gaze point. The videos were recorded with a sampling rate of 50 Hz in a video resolution of 1920 x 1080 at 25 frames per second. The scene camera had a field of view of 90 degrees in 16:9 format (82 degrees horizontal and 52 degrees vertical) and a frame dimension of 179 x 159 x 57 mm (width x depth x height). The recording unit is a compact computer that manages the head unit. It captures and saves eye-tracking data, audio, and scene camera footage on a removable SD memory card.

The evaluation of the calibration process followed the guidelines outlined in the Manual of Tobii AB (2024) and Onkhar et al. (2024) for assessing calibration quality. Participants’ gaze was calibrated using a bullseye card that the participant held at arm’s length. A successful calibration was achieved when the participant’s gaze marker sufficiently overlapped with the bullseye for a specified time, based on criteria internally determined by the manufacturer’s software (Tobii AB, 2024). All participants achieved successful calibration, and no participants were excluded due to calibration failure. The robustness of the calibration was further verified through a secondary nine-point calibration. During this step, participants were asked to read numbers from one to nine aloud and direct their gaze at specific fields corresponding to each number. The initial calibration was performed before and the verification before and after each data collection session.

**Table A1**

*Classification of Nine Typical Classroom Disruptions According to Lohmann & Meyer (2003) Performed in The Micro-Teaching Unit by Actors*

|  |  |  |
| --- | --- | --- |
| Verbal disruptions | Physical disruptions | Lack of eagerness to learn |
| Heckling | Clicking pen | Looking at phone |
| Chatting | Snipping hands | Drawing |
| Whispering | Drumming hands | Head on table |

*Note.* Disruptions were classified based on the typology provided by Lohmann & Meyer (2003)*.* Categories include verbal, physical, and disengagement-related behaviors performed during the micro-teaching unit. The order of the performing actors and the disruptions was fully balanced using Latin squares.

# Appendix B

# Laboratory Setting of The Study

**Figure B1**

*Laboratory Setting of The Micro-Teaching Unit. Ein Bild, das Mobiliar, Stuhl, Kleidung, Schuhwerk enthält.

Automatisch generierte Beschreibung*

*Note*. The setting included three actors as the class (left) and a teacher (participant, right).

**Figure B2**

*Laboratory Setting of The Interview.*

Ein Bild, das Mobiliar, Zeichnung, Entwurf, Tisch enthält.

Automatisch generierte Beschreibung

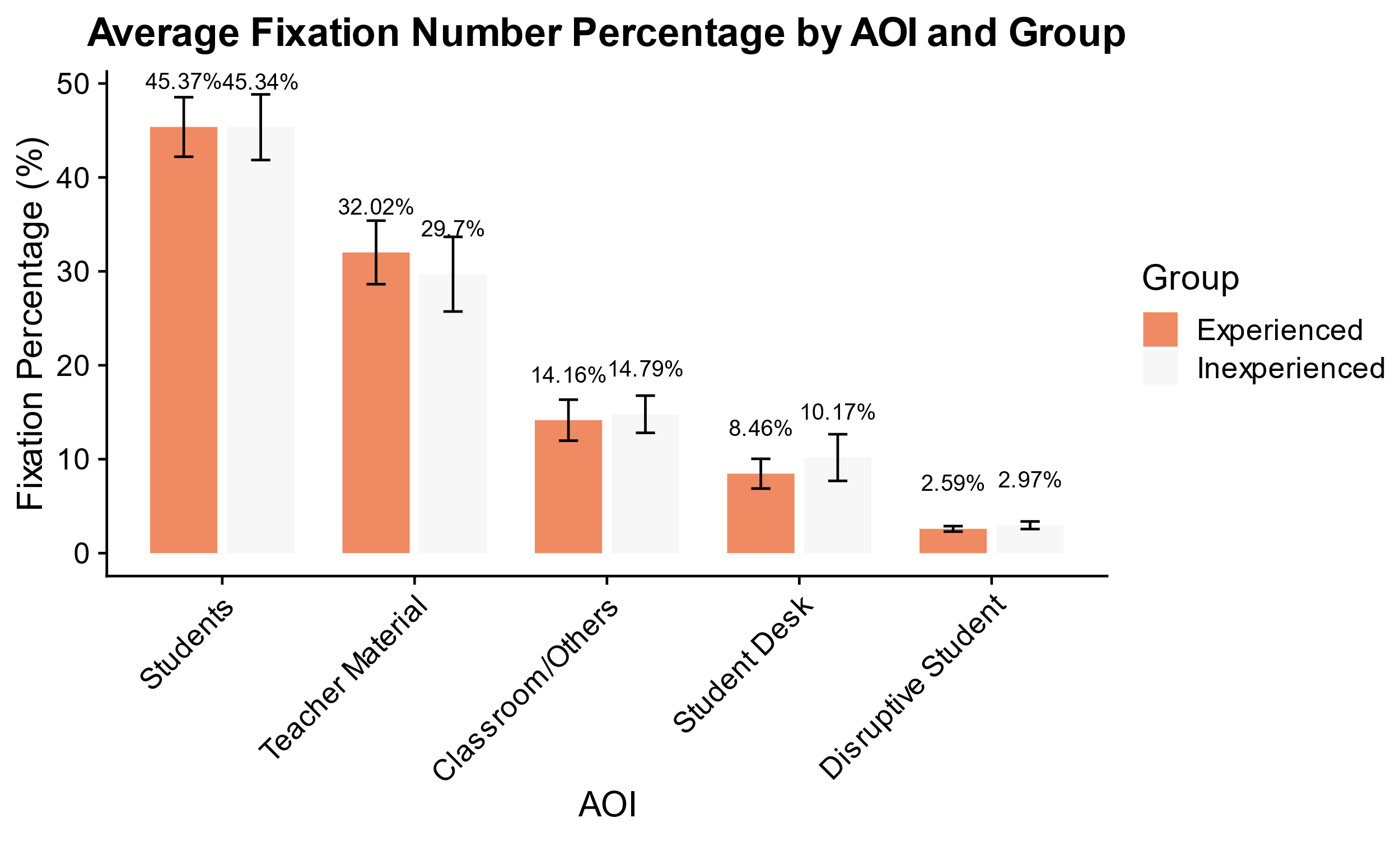
*Note*. The experimenter and participant watched the previously taught micro-teaching unit on video.

# Appendix C

**Additional Results for Gaze Behavior Across AOI**

**Figure C1**

*Average Fixation Number Percentages by Area of Interest (AOI) and Teacher Experience Group for the Entire Micro-Teaching Unit with 95% Confidence Intervals*



*Note.* The bar chart illustrates the average fixation number percentage directed at five AOIs (Students, Teacher Material, Classroom/Others, Student Desk, and Disruptive Student) during the micro-teaching unit. Results are presented separately for experienced and inexperienced teachers, with error bars indicating 95% confidence intervals.

**Table C2**

t*-Test Results and Effect Sizes for Fixation Number Percentages (FNP) and Fixation Duration Percentages (FDP) Across AOIs Between Experienced and Inexperienced Teachers*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Measures | AOI | *t-*value | df | *p*-value | *M* Difference | Cohen's *d* | 95% CI |
| FNP a | Students | 0.01 | 80 | 0.99 | 0.03 | 0.00 | [–4.63, 4.69] |
| FNP | Disruptive Student | –1.32 | 80 | 0.19 | –0.27 | –0.29 | [–0.67, 0.14] |
| FNP | Teacher Material | 0.89 | 80 | 0.37 | 2.32 | 0.20 | [–2.84, 7.48] |
| FNP | Student Desk | –1.17 | 80 | 0.25 | –1.72 | –0.26 | [–4.65, 1.21] |
| FNP | Classroom/Others | –0.43 | 80 | 0.67 | –0.63 | –0.10 | [–3.53, 2.27] |
| FDP b | Students | –0.43 | 80 | 0.67 | –1.26 | –0.10 | [–7.1, 4.58] |
| FDP | Disruptive Student | –1.90 | 80 | 0.06 | –0.48 | –0.42 | [–0.98, 0.02] |
| FDP | Teacher Material | 1.55 | 80 | 0.13 | 4.45 | 0.34 | [–1.28, 10.18] |
| FDP | Student Desk | –1.67 | 80 | 0.1 | –2.89 | –0.37 | [–6.35, 0.56] |
| FDP | Classroom/Others | –0.28 | 80 | 0.78 | –0.30 | –0.06 | [–2.43, 1.84] |
| *Note*. This figure displays the results of *t*-tests and effect sizes (Cohen’s *d*) for fixation number percentages (FNP) and fixation duration percentages (FDP) across Areas of Interest (AOIs) between experienced and inexperienced teachers. AOIs included Students, Disruptive Student, Teacher Material, Student Desk, and Classroom/Others. Positive mean differences (*M* Difference) indicate higher percentages for experienced teachers. Confidence intervals (95%) for the mean differences are presented in brackets.  *p*-value < .05 is considered statistically significant.  a FNP = Fixation Number Percentages  b FDP = Fixation Duration Percentages | | | | | | | |

**Appendix D**

**Correlation Between Gaze Efficiency Measures and Classroom Management Measures**

**Table D1**

*Correlations Between the Gaze Efficiency Measure and Classroom Management Measures for Experienced and Inexperienced Teachers*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| (1) GRI | – | –.44\* | .11 | –.01 | .16 | .15 | –.14 |
| (2) Fixation Number Per Minute on AOI *Students* | –.53\* | – | –.39\* | .39\* | –.49\* | –.22 | .14 |
| (3) TTFF (log) on AOI *Disruptive Student* | .05 | –.29 | – | –.20 | .32\* | –.01 | .01 |
| (4) Disruption Rating | .03 | –.20 | .02 | – | –.02 | .01 | –.15 |
| (5) Confidence Rating | 0 | .06 | –.06 | –.51\* | – | .48\* | –.19 |
| (6) Self-Evaluation | .11 | 0 | –.18 | –.11 | .33\* | – | –.21 |
| (7) Strategic Knowledge Test | –.36\* | .16 | –.34\* | –.19 | –.03 | .12 | – |
| *Note.* The table presents Pearson correlation coefficients between gaze efficiency measures (average fixation duration, fixation number per minute, GRI (= Gaze Relational Index), fixation number per minute on AOI *Students,* and TTFF (= log-transformed time to first fixation on AOI *Disruptive Student*), and classroom management measures (self-evaluation of teachers’ disruption prevention and management competencies, disruption ratings, confidence ratings, and strategic knowledge test) for inexperienced teachers (above diagonal) and experienced teachers (below diagonal). Statistically significant correlations (*p* < .05) are marked with an asterisk (\*). Negative values indicate inverse relationships. GRI = Gaze-Relational Index; TTFF = Time to First Fixation on AOI Disruptive Student (log-transformed). | | | | | | | | |

1. To code and analyze the gaze behavior, we used the software Tobii Pro Lab Analyzer (Version 1.241.54542). A fixation filter was applied, with a threshold set at 30°/sec to identify fixations, as this default fixation filter is recommended for mobile eye-tracking data in the Tobii Lab Analyzer Software (Tobii AB, 2024). [↑](#footnote-ref-2)
2. As our focus was on gaze behavior directed toward the disruptive student, instances where the gaze was already fixated on the disruptive student, when the event started, were excluded from the analysis. [↑](#footnote-ref-3)
3. Fixation times beyond 30 seconds were excluded, as they exceeded the duration of the scripted disruptions. [↑](#footnote-ref-4)
4. The percentage of fixation duration was used as the primary measure, as it offered a more intuitive visualization of attention distribution over time (see Appendix C, Figure C1 for the percentage for fixation numbers). [↑](#footnote-ref-5)
5. Before conducting the analysis, 21.33% of data points were excluded because the participant either did not fixate on the disruption or had fixation times of zero or longer than 30 seconds, exceeding the defined threshold (see Data Analysis). Removing them ensured a focus on meaningful and interpretable gaze detection times. [↑](#footnote-ref-6)
6. See Appendix D (Table D1) for a full correlation table including all gaze efficiency measures. [↑](#footnote-ref-7)