Through the eyes of the teacher

Mandy Klatt[1], Dr. Gregor Kachel[1, 2], Dr. Christin Lotz[1], & Prof. Dr. Anne Deiglmayr[1]

[1] Leipzig University

[2] Max-Planck University for Evolutionary Anthropology

Author Note

The Ethics Advisory Board of Leipzig University has dealt with the research project and has come to the conclusion that there are no objections to the implementation of this research project. The Ethics Advisory Board points out that the scientific and ethical responsibilty for the implementation of the project remains with the project director.

Correspondence concerning this article should be addressed to Mandy Klatt, Egelstraße 2a 04103 Leipzig. E-mail: mandy.klatt@uni-leipzig.de

Abstract

This document is a supplement to the paper and shows first graphs findings from the pilot study.

Keywords: Professional Vision, Expert-Novice-Paradigm, Eye-Tracking

Word count: 1949

Through the eyes of the teacher

State of research

Teaching and classroom management are multidimensional settings in which teachers have to respond immediately to events as they develop (Barnes, 2004). The different interests and abilities of students must be managed in a way that maximizes the active learning time of students and minimizes disruptions whilst teaching. Learning to develop such classroom management skills and to teach effectively is a complicated and complex process (Wolff, Jarodzka, & Boshuizen, 2017).

During teaching, teachers must be able to select from a variety of visual and acoustic impressions to focus their attention on the essential and to distinguish between relevant and irrelevant events. This ability is called professional vision and is a key component of teacher expertise and successful teaching (Barth, 2017). Eye tracking technology has become a reliable means to study teachers' visual focus of attention (Bogert, 2016; Pouta, Lehtinen, & Palonen, 2020; Wolff, Jarodzka, & Boshuizen, 2017)

Educational research has repeatedly shown that there are differences between experienced and novice teachers in terms of perception and behavioral competencies (Barth, 2017; Bogert, 2016; Wolff, Jarodzka, & Boshuizen, 2017). For example, experts direct their attention more often and more evenly to all students, whereas novices only direct their attention to some students. The frequency and duration of fixations as eye movement are decisive (Stuermer, Seidel, Mueller, Häusler, & Cortina, 2017). Mobile eye-tracking technology has also shown that experienced teachers distribute their focus more efficiently to solve tasks (Jarodzka, Scheiter, Gerjets, & Van Gog, 2010). Furthermore, in contrast to novices, experts are able to focus their attention on the entire class and guide the class while giving feedback to individual students and answering questions (Cortina, Miller, McKenzie, & Epstein, 2015).

Research questions

The aim of the pilot study was to investigate whether there are differences in how expert and novice teachers manage scripted classroom disruptions. The disruptions were experimentally varied using a previously written script. Thus, our aim was to find out whether differences in the allocation of attention between expertise groups can be detected in this controlled context.

In order to answer this question, the hypothesis was formulated that teachers with more professional experience not only notice more disruptions but also notice them faster. In the hypothesis, therefore, it is necessary to check what has already been shown in the research literature: In complex teaching situations, experts have a more structured and elaborate professional knowledge than novices in order to perceive and interpret relevant events and to act appropriately (Berliner, 2001; Lachner, Jarodzka, & Nückles, 2016).

Methods

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

Participants

For the sample recruitment of the subjects ($N = 8$, experts $n = 2$, novices $n = 6$), schools in the city of Leipzig in Saxony were contacted. The institutions as well as the subjects were informed in detail about the aim and intention of the study in advance. Participation in the study was voluntary and only took place after written consent has been given.

The selection of the subjects was based on extreme groups, whereby professional experience is the crucial criterion for the selection of experts or novices. Novices were recruited as teachers who have been working in the teaching profession for no more than 3

Table 1

Demographic Informationand and Teaching Experience

group	N	Male	M age	Min age	Max age	SD age	M exp.	Min exp.	Max exp.	SD exp.
expert	2	1	47.50	44	51	4.95	20.00	15.00	25.00	7.07
novice	6	2	25.67	20	33	4.89	0.68	0.00	1.50	0.68

years, whereas experts were considered to have professional experience of 10 years or more (Messner & Reusser, 2000).

Procedure/ Data collection

Set up. For this study, scripted mini-lessons with n = 2 experts and n = 6 novices were recorded in the mobile Lab of the Empirical School and Classroom Research at the University of Leipzig. The subjects were divided into groups of four, so the study was conducted on two different sessions. All participants were asked to hold a 10-minute lesson. The duration of each appointment was approximately 2h: per group 10min briefing, 4 x 10min mini-lessons, 10min technical preparation and follow-up and 4x 10min transition points between the lessons and answering questionnaires.

One person from the group of 4 acted as a teacher, the other three subjects acted as the class. The subjects, who represented the class, were given behavioral instructions in a pre-written script to simulate typical events and disruptions in the classroom (e.g. putting their heads on the table, chatting, looking at their mobile phones, etc.).

The lesson disruptions were displayed as instructions during the lesson for all “students” but not the teacher. In order to avoid learning effects, the disruptions in each lesson were distributed pseudo-randomly over the short teaching phase. In addition, the order of the data collection was taken into account in the analyses and variance caused by order was controlled.

Figure 1. Example for set up during a mini-lesson

Questionnaire data. After each mini-lesson, the students answered items on the teaching quality using a validated questionnaire (Helmke et al., 2014) and scales on the teacher's presence behavior (students n = 24). In addition, the teacher was asked to give a self-assessment on his/her classroom management by completing the questionnaire after each mini-lesson (teachers n = 8).

Behavioral data. The speech, sounds and voices were recorded with an audio recorder installed in the middle of the Lab. Movements, facial expressions and gestures of the subjects were recorded by four cameras from different angles. One camera was installed to film the class from the side. Two more cameras were installed on the blackboard and at the end of the Lab to film the teacher and class from the front and back. Furthermore, the fourth camera was installed in such a way that only facial expressions and gestures of the teacher were recorded, which enables a semi-automated analysis of the movement sequences.

Eyetracking data. A binocular Tobii Pro Glasses 2 eye-tracker consisting of a wearable head unit and a recording unit was used to record the eye movements of all 8 participants. The head unit is a measuring device with different sensitive sensors. A high-definition scene camera captures a full HD video and an integrated microphone records the surrounding sounds. Infrared light illuminators support the eye tracking sensors which record the eye orientation. The videos were recorded with a sampling rate of 50 Hz and a video resolution with 1920 x 1080 at 25 frames per second. The scene camera has a field of view of 90 deg. in 16:9 format (82 deg. horizontal and 52 deg. vertical) and has a frame dimension of 179 x 159 x 57mm (width x depth x height). The Tobii Pro Glasses Controller software was used to record and calibrate the eye movements.

Coding/ Data preparation/ Reliability

Questionnaire Data. The evaluation after each mini-lesson was conducted using paper questionnaires. Time needed to complete the questionnaire was about 5 minutes. The scales on the quality of teaching are a validated questionnaire (Helmke et al., 2014). Whereas the scales on the teacher's presence behavior were derived from the research literature (Brophy, 1986; Kiel, Frey, Weiß, & Weiss, 2013; Kounin, 2006; Marzano, 2007; Nolting, 2012) and were used in the pilot for the first time. The questionnaire is 4-point Likert scale (1 = Strongly Disagree; 2 = Disagree; 3 = Agree; 4 = Strongly Agree). Data was obtained from N = 32 subjects (students n = 24, teachers n = 8).

The following scales were assessed:

(1) Classroom management
(2) Positive climate and motivation
(3) Clarity and structuredness

(4) Activation and support

(5) Presence: posture/gaze

(6) Presence: voice

(7) Presence: verbal and non-verbal intervention

(8) Natural behaviour

The table provides an overview over the mean, the range and standard deviation of all scales.

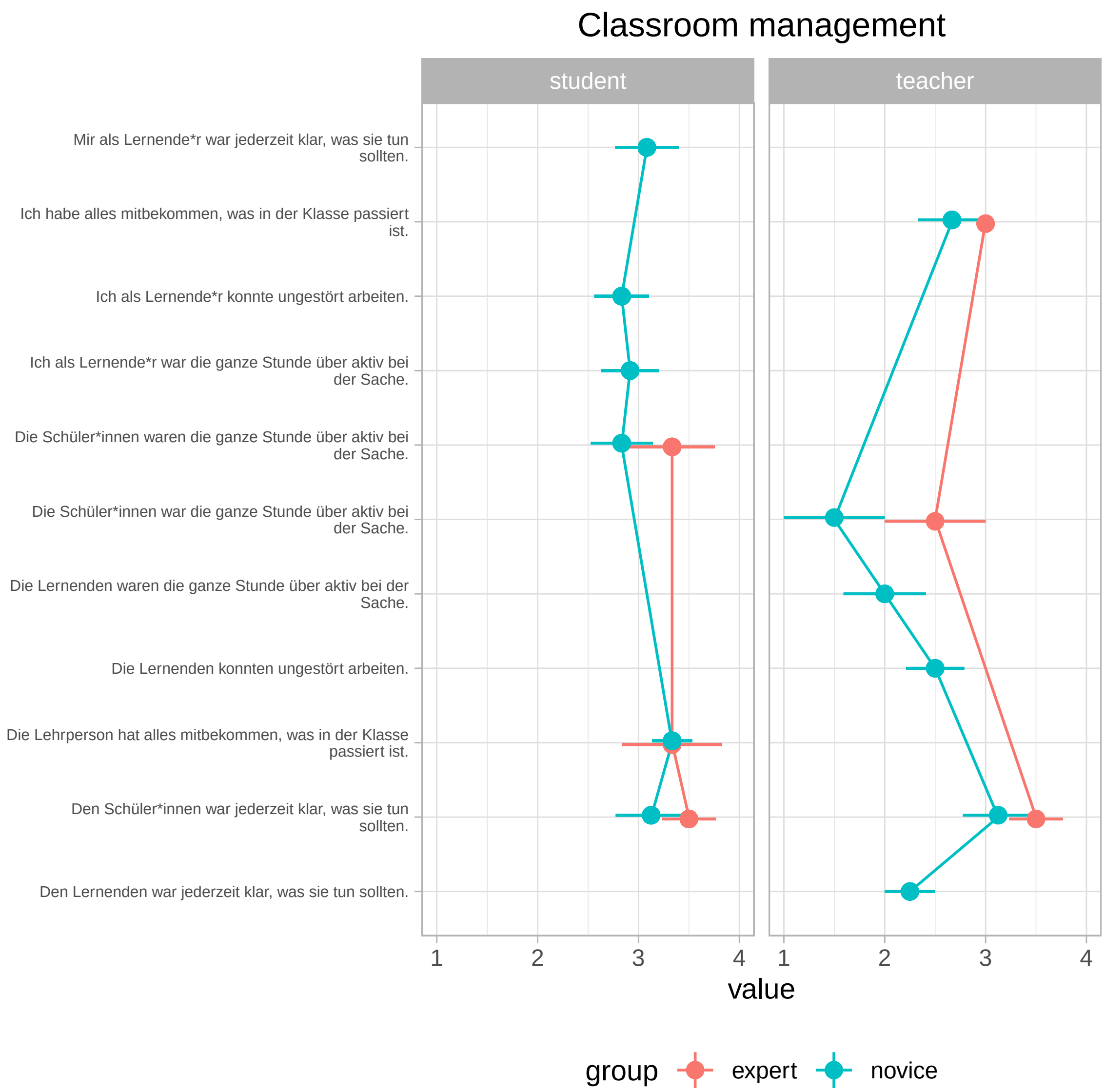
Table 2

Mean values of all scales

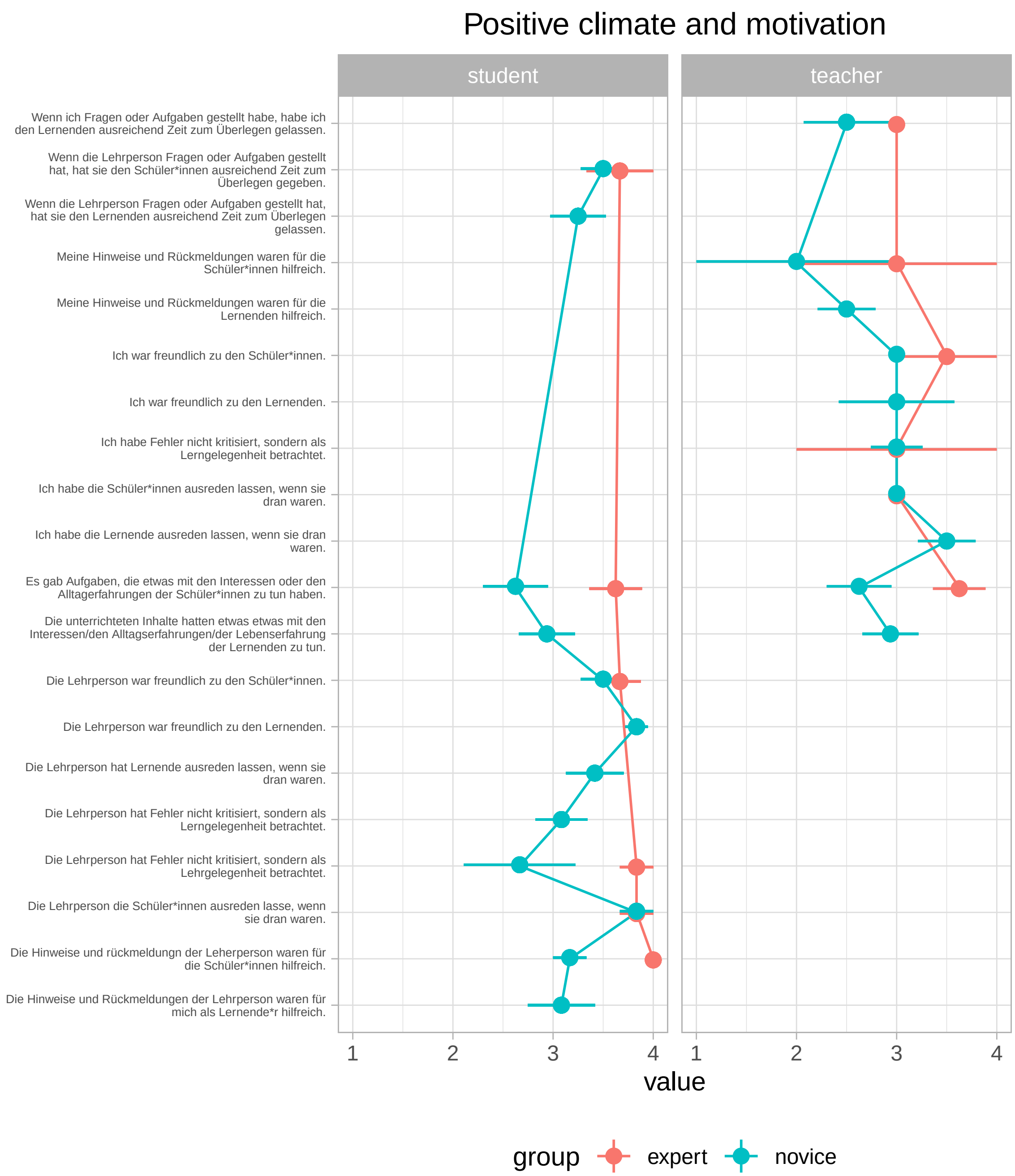
scale	M scale	SD scale	min scale	max sclae
Activation and support	2.45	1.37	0.00	4.00
Clarity and structuredness	3.41	0.81	1.00	4.00
Classroom management	2.97	0.94	1.00	4.00
Natural behaviour	3.28	0.74	2.00	4.00
Positive climate and motivation	3.27	0.89	0.00	4.00
Presence: posture/gaze	3.10	0.96	0.00	4.00
Presence: verbal and non-verbal intervention	3.07	0.82	1.00	4.00
Presence: voice	3.40	0.72	2.00	4.00

The individual items of a scale are further represented in graphs.

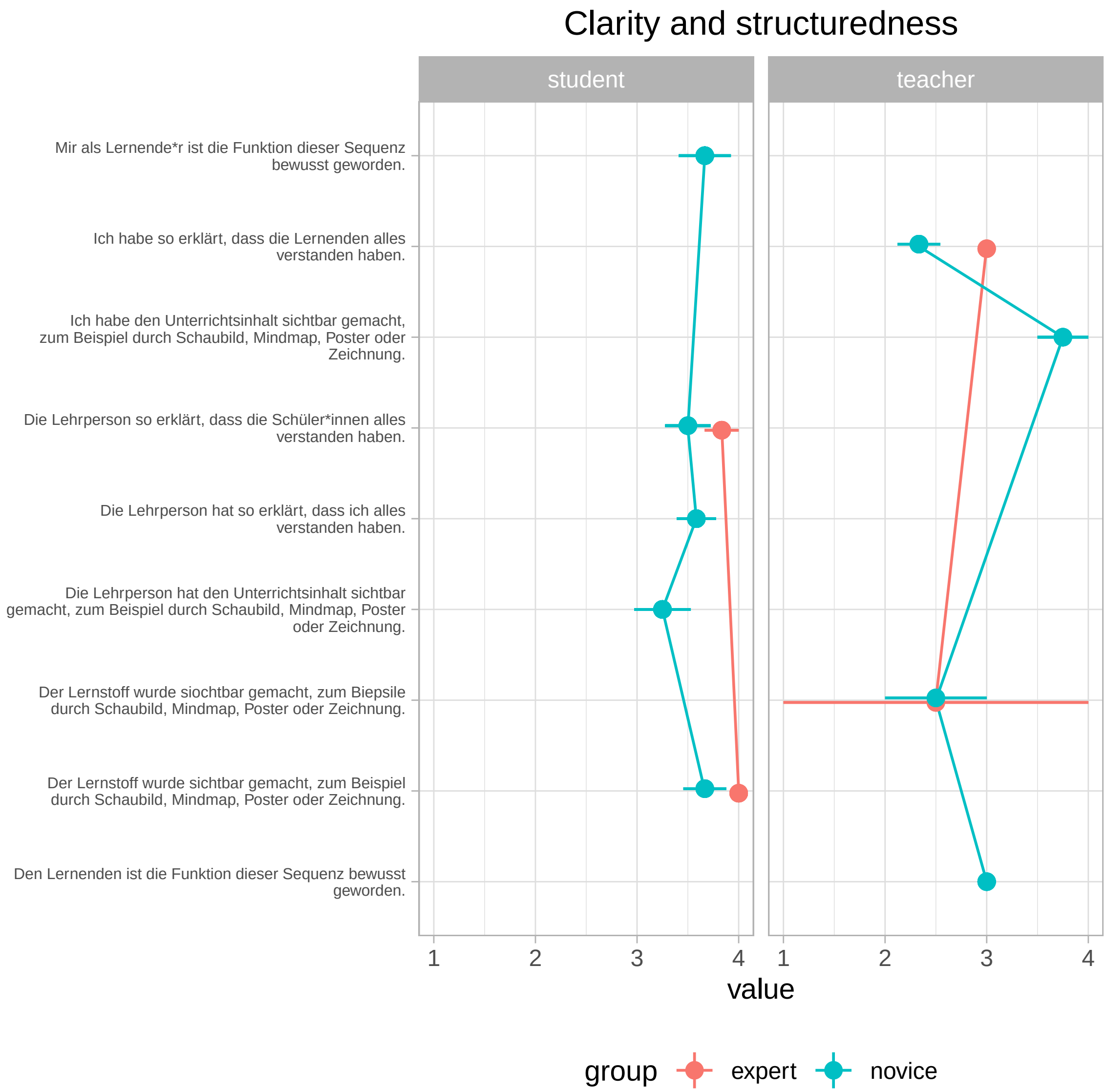
(1) Classroom management



(2) Positive climate and motivation

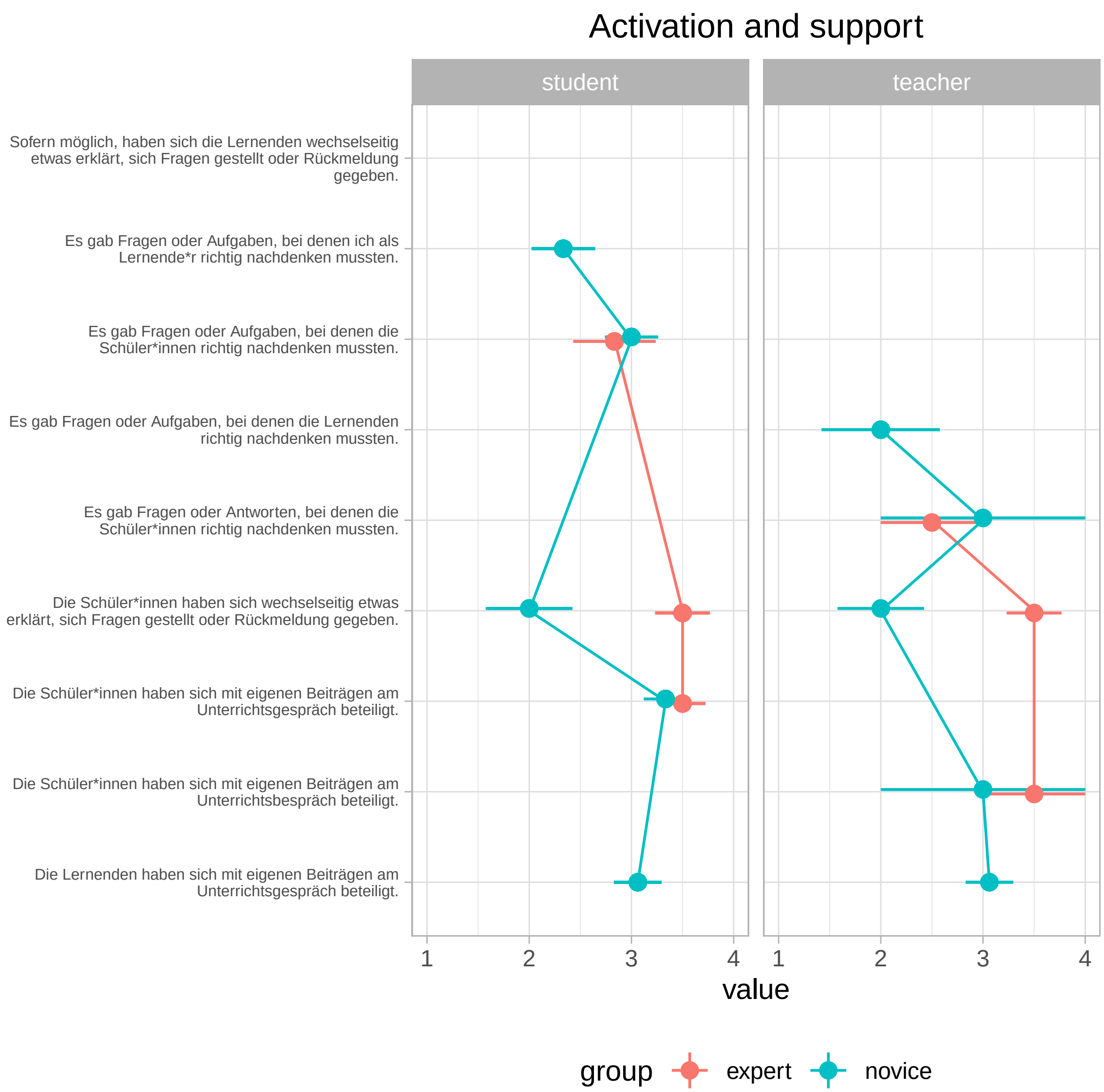


133 (3) Clarity and structuredness



134

(4) Activation and support



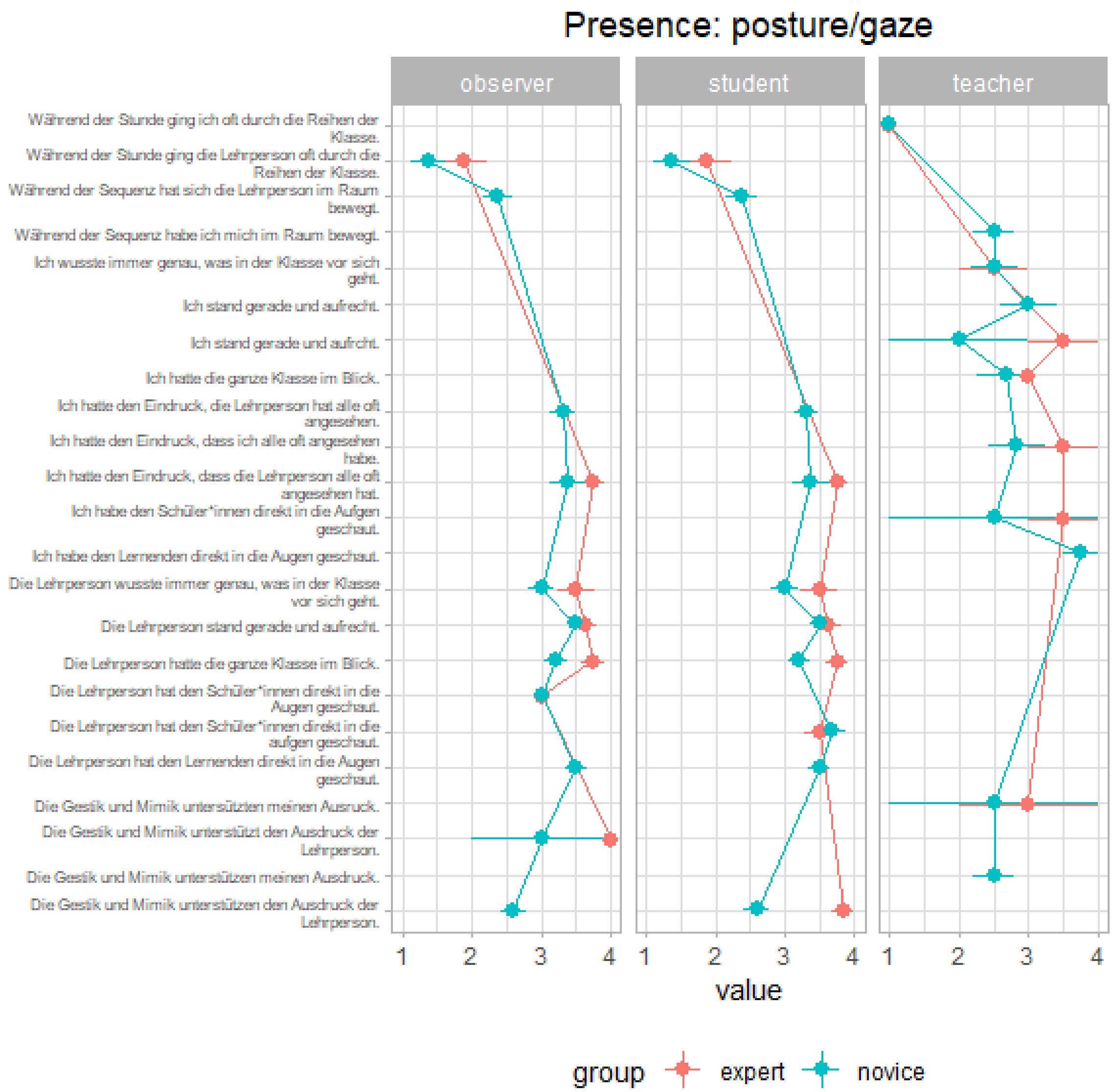


Figure 2. (5) Presence: posture/gaze

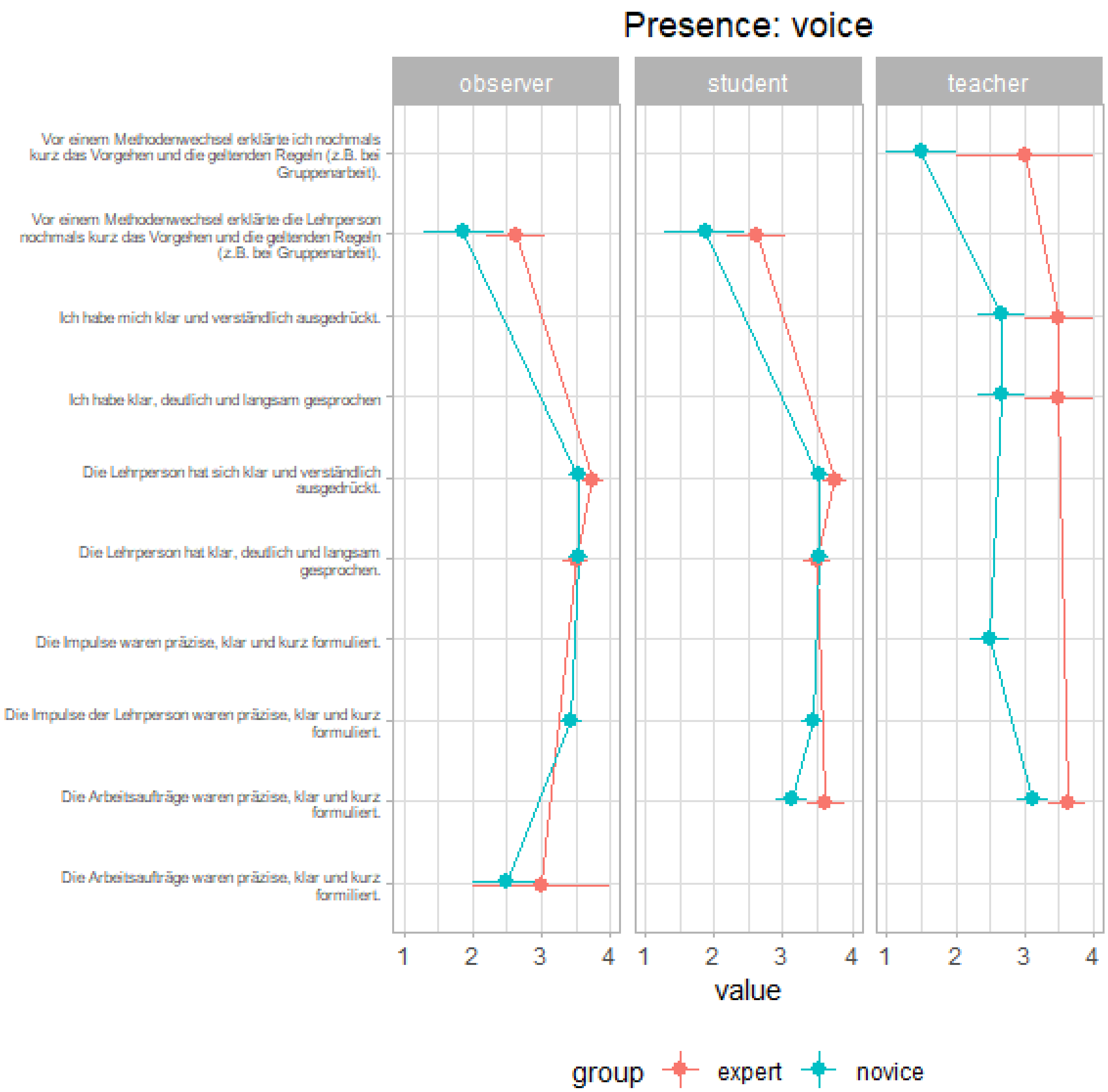


Figure 3. (6) Presence: voice

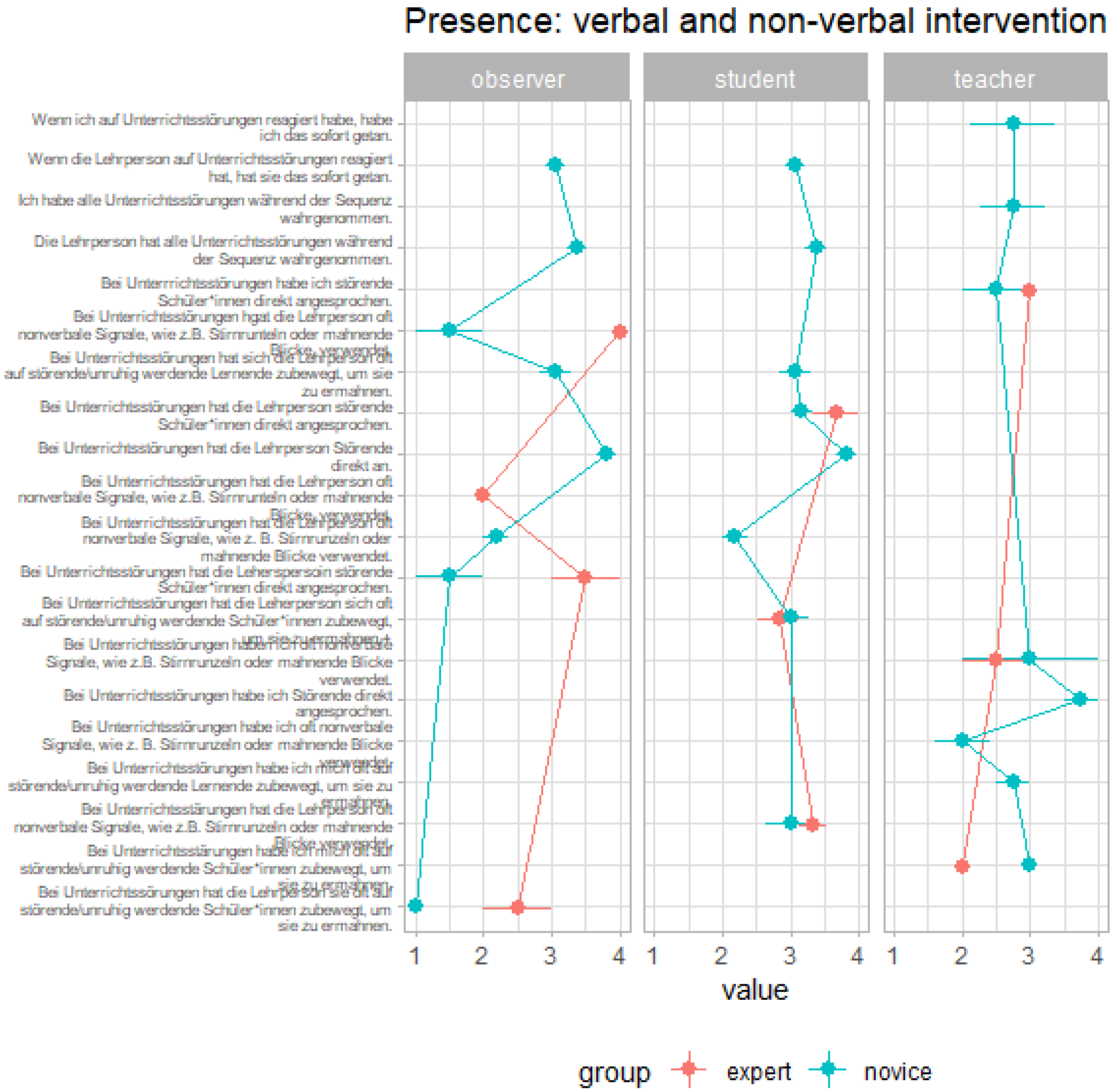
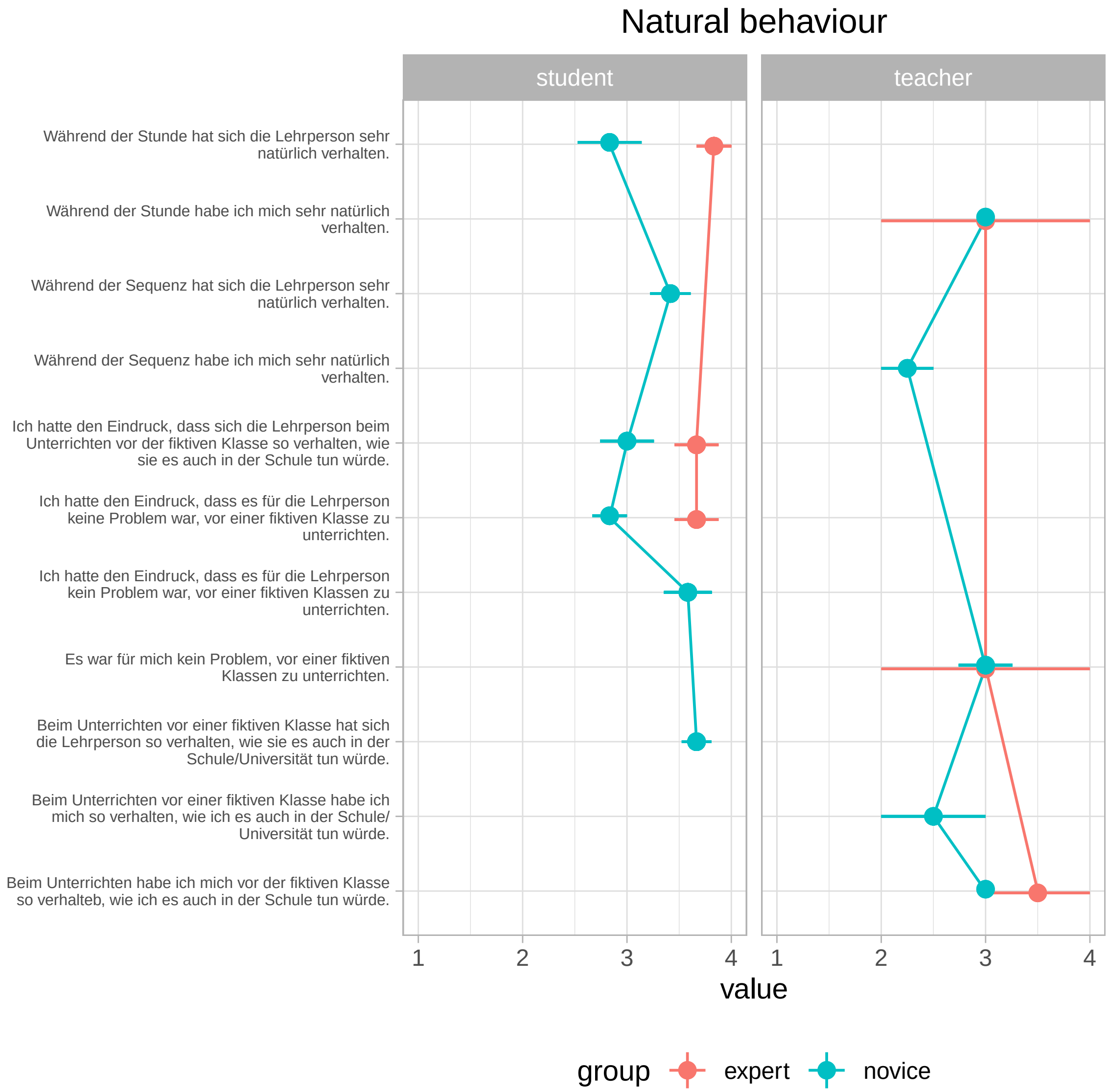
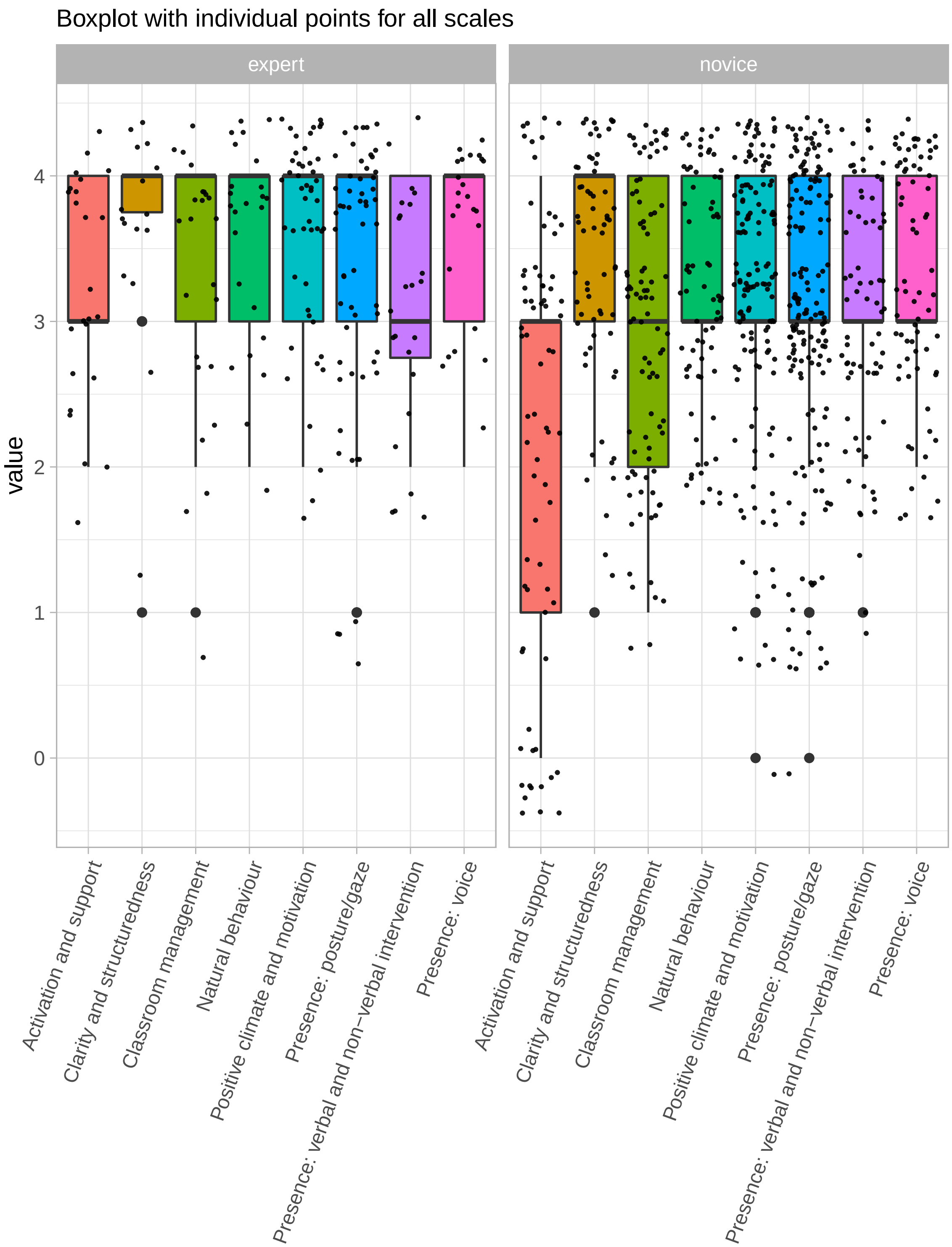


Figure 4. (7) Presence: verbal and non-verbal intervention

(8) Natural behaviour



In addition, we plotted all scales. Graph provides boxplots and individual data for experts and novices.



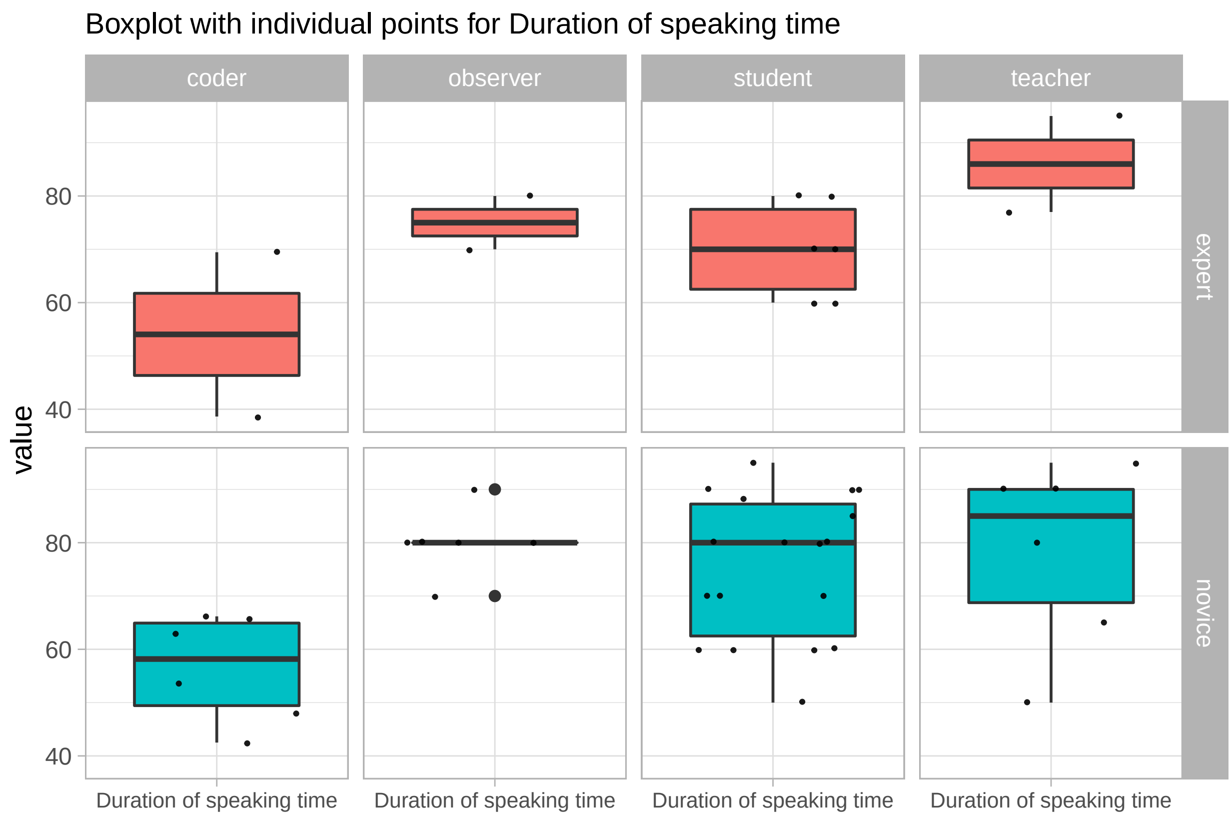
Behavioral Data. The recorded lessons were coded in a post-hoc procedure with the coding software MAXQDA by previously trained raters (Kuckartz and Rädiker (2019)). The following coding scheme was developed:

- phase - lesson begin, state event: teacher starts the lesson with a noise, talk, taking a position in class
- phase - lesson end, state event: teacher finishes the lesson with a noise, talk, taking a position in class
- phase - organization/transition points, state event: any situation that does not imply effective learning time (fetching chalk, working material, organizing desks, opening windows, printing work results etc.)
- phase - single, state event: any individual student activity on a given task (reading, writing, drawing etc.)
- phase - group, state event: any student activity on a given task together in a group of at least 3 students (reading, writing, drawing etc.)
- phase - class discussion, state event: discussion in class, teacher talks to class/individual/group
- phase - pair: state event: any student activity on a given task together in a team of 2 students (reading, writing, drawing etc.)
- phase - teachers lecture, state event: any teacher's presentation on a certain topic which maybe supported by a PPP, PREZI, notes on board, OHP etc.
- phase - other, state event: not categorizable
- phase - break, state event: e.g. drinking, relaxation exercises

- phase - external interruption, state event: external interruptions (e.g. fire alarm, technical problems, other teachers coming into the room)
- speaking time - teacher, state event
- speaking time - students, state event
- disruption - chatting with neighbor, state event (perceived/ not perceived, reacted: verbal, non-verbal/ not reacted)
- disruption - asking a question, state event (perceived/ not perceived, reacted: verbal, non-verbal/ not reacted)
- disruption - yelling, state event (perceived/ not perceived, reacted: verbal, non-verbal/ not reacted)
- disruption - looking at phone, state event (perceived/ not perceived, reacted: verbal, non-verbal/ not reacted)
- disruption - staring out of window, state event (perceived/ not perceived, reacted: verbal, non-verbal/ not reacted)
- disruption - drawing, state event (perceived/ not perceived, reacted: verbal, non-verbal/ not reacted)
- disruption - head on table, state event (perceived/ not perceived, reacted: verbal, non-verbal/ not reacted)
- disruption - clicking pen, state event (perceived/ not perceived, reacted: verbal, non-verbal/ not reacted)
- disruption - drumming hands, state event (perceived/ not perceived, reacted: verbal, non-verbal/ not reacted)

- disruption - walking around, state event (perceived/ not perceived, reacted: verbal, non-verbal/ not reacted)

First, we coded the speaking time of the teacher and the students to compare all perspectives: coder, observer, students, teacher. The graph below shows the result of the coded speaking duration compared to the estimated speaking duration assessed with the questionnaire.



Eyetracking Data. The Tobii Pro Lab 2 software was used to analyze the teachers' visual attetion during each mini-lesson. The software allows for non-screen based recordings of a participants' attention while moving in real-world settings. The recordings of the glasses contain both HD-video from the subjects' perspective as well as the respective gaze data mapped onto the video. In order to map multiple recordings to AOIs, we first imported the eye-tracking recordings into the Tobii Pro Analyzer software. Second, we created dynamic Areas of Interest (AOI) manually to plot the gaze data. Once the

AOIs are created, the gaze recordings of multiple recordings can be mapped and analyzed in aggregated form. Tobii Pro does not allow to do AOI based analyses within Pro Lab. So we exported a tsv. file to do further analyses in the software R.

Gaze relational index (GRI).

The GRI is a measure of visual expertise in information processing. This metric is calculated as the ratio of mean fixation duration to fixation count. The GRI is higher for novices than for experts. (Gegenfurtner et al., 2020)

Table 3

Number and Duration (in msec) of Fixations

Participant	Variable	Fixation Number	Fixation Duration	M Duration Fixation	TOI	GRI
01_01_D	Expert	803.00	316,571.00	394.00	781,978.00	0.49
01_02_A	Expert	1,070.00	385,812.00	361.00	838,026.00	0.34
01_03_B	Novice	617.00	374,315.00	607.00	744,444.00	0.98
01_04_C	Novice	769.00	384,537.00	500.00	723,922.00	0.65
02_01_A	Novice	569.00	101,541.00	178.00	729,762.00	0.31
02_02_B	Novice	1,140.00	520,431.00	457.00	730,565.00	0.40
02_03_C	Novice	1,048.00	469,018.00	448.00	737,604.00	0.43
02_04_D	Novice	613.00	438,655.00	716.00	747,729.00	1.17

Data analysis

We used R [Version 4.0.3; R Core Team (2019)] and the R-packages *}dplyr* [@}R-dplyr], *forcats* [Version 0.5.0; Wickham (2020a)], *ggplot2* [Version 3.3.2; Wickham (2016)], *papaja* [Version 0.1.0.9997; Aust and Barth (2020)], *papayar* (Muschelli, 2016), *psych* [Version 2.0.12; Revelle (2020)], *purrr* [Version 0.3.4; Henry and Wickham (2020)], *readr* [Version 1.4.0; Wickham, Hester, and Francois (2018)], *stringr* [Version 1.4.0; Wickham (2019)], *tibble* [Version 3.0.4; Müller and Wickham (2021)], *tidyr* [Version 1.1.2; Wickham (2020b)], and *tidyverse* [Version 1.3.0; Wickham et al. (2019)] for all our analyses.

Table 4

Number and Duration (in msec) of Fixations during calibration

Participant	Variable1	Fixation Number	Fixation Duration	M Duration Fixation	TOI	GRI
01_01_D	Expert	9.00	14,372.00	1,597.00	16,470.00	177.44
01_02_A	Expert	10.00	10,194.00	1,019.00	13,335.00	101.90
01_03_B	Novice	17.00	9,234.00	543.00	10,615.00	31.94
01_04_C	Novice	14.00	15,311.00	1,094.00	17,224.00	78.14
02_01_A	Novice	13.00	5,157.00	397.00	17,902.00	30.54
02_02_B	Novice	12.00	10,654.00	888.00	12,325.00	74.00
02_03_C	Novice	18.00	14,151.00	786.00	16,494.00	43.67
02_04_D	Novice	14.00	19,128.00	1,366.00	20,964.00	97.57

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Questionnaire Data.

Behavioral Data.

Eyetracking Data.

Results

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Questionnaire Data.

Behavioral Data.

Eyetracking Data.

Discussion

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

References

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown.* Retrieved from https://github.com/crsh/papaja

Barnes, D. (2004). The significance of teachers' frames for teaching. In *Teachers and teaching* (pp. 16–38). Routledge.

Barth, V. L. (2017). *Professionelle wahrnehmung von störungen im unterricht.* Springer.

Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, *35*(5), 463–482.

Bogert, N. J. van den. (2016). *On teachers' visual perception and interpretation of classroom events using eye tracking and collaborative tagging methodologies.* Technische Universiteit Eindhoven.

Brophy, J. (1986). Classroom management techniques. *Education and Urban Society*, *18*(2), 182–194.

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.

Gegenfurtner, A., Boucheix, J.-M., Gruber, H., Hauser, F., Lehtinen, E., & Lowe, R. K. (2020). The gaze relational index as a measure of visual expertise. *Journal of Expertise. East Lansing, Michigan: Michigan State University, 2020, Vol. 3, Iss. 1.*

Helmke, A., Helmke, T., Lenske, G., Pham, G., Praetorius, A.-K., Schrader, F.-W., & AdeThurow, M. (2014). Unterrichtsdiagnostik mit EMU. *Aus-Und Fortbildung Der Lehrkräfte in Hinblick Auf Verbesserung Der Diagnosefähigkeit, Umgang*

Mit Heterogenität Und Individuelle Förderung, 149–163.

Henry, L., & Wickham, H. (2020). *Purrr: Functional programming tools.* Retrieved from https://CRAN.R-project.org/package=purrr

Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, *20*(2), 146–154.

Kiel, E., Frey, A., Weiß, S., & Weiss, S. (2013). *Trainingsbuch klassenführung* (Vol. 3992). UTB.

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

Kuckartz, U., & Rädiker, S. (2019). *Analyzing qualitative data with MAXQDA.* Springer.

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.

Marzano, R. J. (2007). *The art and science of teaching: A comprehensive framework for effective instruction.* Ascd.

Messner, H., & Reusser, K. (2000). Die berufliche entwicklung von lehrpersonen als lebenslanger prozess. *Beiträge Zur Lehrerinnen-Und Lehrerbildung*, *18*(2), 157–171.

Muschelli, J. (2016). *Papayar: View medical research images using the papaya JavaScript library.* Retrieved from https://CRAN.R-project.org/package=papayar

Müller, K., & Wickham, H. (2021). *Tibble: Simple data frames.* Retrieved from https://CRAN.R-project.org/package=tibble

Nolting, H.-P. (2012). *Störungen in der schulklasse.* Beltz.

Pouta, M., Lehtinen, E., & Palonen, T. (2020). Student teachers' and experienced teachers' professional vision of students' understanding of the rational number concept.

R Core Team. (2019). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Revelle, W. (2020). *Psych: Procedures for psychological, psychometric, and personality research.* Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Stuermer, 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.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string operations.* Retrieved from https://CRAN.R-project.org/package=stringr

Wickham, H. (2020a). *Forcats: Tools for working with categorical variables (factors).* Retrieved from https://CRAN.R-project.org/package=forcats

Wickham, H. (2020b). *Tidyr: Tidy messy data.* Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., Hester, J., & Francois, R. (2018). *Readr: Read rectangular text data.* Retrieved from https://CRAN.R-project.org/package=readr

Wolff, C. E., Jarodzka, H., & Boshuizen, H. P. (2017). See and tell: Differences between expert and novice teachers' interpretations of problematic classroom management events. *Teaching and Teacher Education*, *66*, 295–308.