



HAL
open science

Improving pre-service teachers' diagnostic judgments regarding task difficulties in the domain of fractions and angles -A process-oriented investigation with Eye-Tracking

Saskia Schreiter, Markus Vogel, Markus Rehm, Tobias Dörfler

► To cite this version:

Saskia Schreiter, Markus Vogel, Markus Rehm, Tobias Dörfler. Improving pre-service teachers' diagnostic judgments regarding task difficulties in the domain of fractions and angles -A process-oriented investigation with Eye-Tracking. Twelfth Congress of the European Society for Research in Mathematics Education (CERME12), Feb 2022, Bozen-Bolzano, Italy. hal-03744869

HAL Id: hal-03744869

<https://hal.science/hal-03744869v1>

Submitted on 3 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Improving pre-service teachers' diagnostic judgments regarding task difficulties in the domain of fractions and angles – A process-oriented investigation with Eye-Tracking.

Saskia Schreiter¹, Markus Vogel¹, Markus Rehm¹ and Tobias Dörfler¹

¹Heidelberg University of Education, Germany; schreiter@ph-heidelberg.de

In this ongoing study, the focus is put on the potential influence of specific knowledge on pre-service teachers' judgment processes when diagnosing the difficulty of mathematics tasks. It is assumed that pre-service teachers use their knowledge in order to identify difficulty-generating task characteristics and to evaluate them in terms of their difficulty for students. To examine this assumption, an experimental group, in which participating pre-service teachers acquire knowledge about typical student difficulties and specific difficulty-generating task characteristics is compared with a control group. Using eye-tracking technology and stimulated recall interviews, indicators for diagnostic judgment processes are collected and compared between conditions. First results suggest that specific knowledge leads to a more efficient judgment process and enables pre-service teachers to identify and correctly evaluate difficulty-generating task characteristics more frequently.

Keywords: Diagnostic judgment processes, knowledge, eye-tracking, stimulated recall interview.

Introduction

For effective mathematics instruction, the use of tasks that are adapted to the students' ability level in terms of difficulty is considered central (Leuders & Prediger, 2016). A prerequisite for this is that teachers are able to identify task difficulties and evaluate them adequately. The difficulty of mathematics tasks is influenced by mathematical task characteristics (e.g., in the domain of fractions: like vs. unlike fractions; Padberg & Wartha, 2017) and instructional characteristics (according to cognitive load theory (CLT): e.g., split-attention vs. integrated task design; Sweller et al., 2011), among other characteristics. However, research findings show that teachers often make inadequate judgments about task difficulties (Karing & Artelt, 2013) and do not sufficiently consider the task's instructional design (Schreiter et al., 2021). Specific knowledge of difficulty-generating task characteristics was found to play a significant role for the accuracy of teachers' diagnostic judgments regarding task difficulty (Ostermann et al., 2017). However, most of the research on teachers' diagnostic judgments has so far focused on judgment accuracy. Therefore, little is known about how teachers get to their result and what role knowledge plays in the judgment process (Loibl et al., 2020). To gain insight into internal judgment processes, eye-tracking in combination with eye-tracking stimulated recall interviews has proven to be an effective method (Schindler & Lilienthal, 2019).

Theoretical background: Diagnostic judgment processes

It is assumed that in the genesis of diagnostic judgments, teachers use the information available in a diagnostic situation and process it on the basis of their knowledge to get to their result (Herppich et al., 2018). When judging the difficulty of a task for students, the relevant information constitutes the task characteristics that hold information about the tasks' difficulty ("difficulty-generating task characteristics"; Leuders & Prediger, 2016). In the judgment process (Loibl et al., 2020), these

difficulty-generating task characteristics should be identified by teachers and adequately evaluated in terms of difficulty for students. Diagnostic judgment processes, such as identifying and evaluating task characteristics, constitute internal cognitive processes that cannot directly be observed. Research showed that eye-tracking is an effective method to investigate prospective teachers' diagnostic judgment processes (Brunner et al., 2021). However, as the interpretation of eye-tracking data can be challenging and ambiguous (Moreno-Esteva et al., 2017), a triangulation with qualitative data, such as eye-tracking stimulated recall interviews (ET SRI), is recommended (Schindler & Lilienthal, 2019). ET SRI is a research method to investigate cognitive processes by asking probands to retrospectively describe their own thoughts using a video sequence of their eye movements. In their systematic review on teacher's professional vision, Grub et al. (2020) report that numerous eye-tracking studies determined significant differences in gaze behavior between experts and novices. For example, experts were found to have shorter fixation durations, which were interpreted as an indicator of faster information processing. However, in most of these studies, experts were distinguished from novices only by the number of years of job experience, and other knowledge components were not considered. It therefore remains unclear how and whether teachers' knowledge – regardless of job experience – influences their gaze behavior during diagnostic activities.

This Study

This ongoing study aims to examine pre-service teachers' diagnostic judgment process when diagnosing the difficulty of mathematics tasks in the domain of fractions and angles. In this regard, it is intended to examine which task characteristics (mathematical vs. instructional) pre-service teachers identify and evaluate when judging task difficulties. In addition, a particular research interest is to investigate the potential influence of specific knowledge about difficulty-generating task characteristics on identification and evaluation processes. Please note that this contribution constitutes a pre-report of an ongoing study and only covers a sample of the data and results of a larger data set. Based on the above-mentioned findings of Schreiter et al. (2021), we expect for both content areas (fractions and angles) that a) mathematical task characteristics are identified and correctly evaluated more frequently compared to instructional characteristics (H1a) and b) instructional task characteristics are to a large extent not identified and correctly evaluated (H1b). Furthermore, building on the results of existing research on the influence of specific knowledge on judgment accuracy (e.g., Ostermann et al., 2017), we assume that specific knowledge enables pre-service teachers to identify and correctly evaluate more difficulty-generating task characteristics (mathematical and instructional) (H2). As eye-tracking studies focusing on teachers' professional vision revealed significant differences in gaze behavior between experts and novices (here: teachers with and without job experience; Grub et al., 2020), we further aim to exploratively investigate whether specific knowledge influences the gaze behavior of pre-service teachers (who have no job experience) during their diagnostic judgment of task difficulties.

Methods

Participants. $N = 25$ pre-service mathematics teachers were assigned to two conditions: an experimental group ($n = 11$), that received a 90-minute intervention and a control group ($n = 14$) without treatment. The intervention addressed specific mathematical task characteristics in the

domain of fractions and angles that are known to cause difficulties for learners. Furthermore, theoretical foundations of CLT and related instructional task characteristics were addressed.

Material. Eight mathematics tasks were created (four fraction tasks and four angles tasks). Between these tasks, difficulty-generating mathematical and instructional task characteristics were systematically varied. The mathematical difficulty of the fraction tasks was varied by modifying the denominators (like vs. unlike), by mixing natural numbers and fractions, and by using mixed fractions (Padberg & Wartha, 2017). The mathematical difficulty of the angle tasks was varied by the number of argumentation steps needed to solve the task and by the angle values used (tens vs. units) (e.g., Reiss, 2002). The instructional difficulty of both content areas (fraction and angle tasks) was varied according to specific CLT design principles (Sweller et al., 2011): based on the *split-attention effect*, the tasks' relevant information is presented either close or distant from each other. Furthermore, building on the *redundancy effect*, the tasks were created in such a way that a) one and the same information is presented by different information sources or b) additional information irrelevant for the solution is included or c) no redundant information is included. The instructional design of the sample fraction task (figure 1) causes potential difficulties, as different information sources (the graphic and the time information) are presented distant from each other (split-attention effect) and redundant information (the route from Neckarsteinach to Ziegelhausen) is included (redundancy effect). The instructional design of the sample angle task (figure 1) causes potential difficulties, as the same information (values of the given and missing angles, tip) is presented by the text as well as in the graphic which causes unnecessary processing procedures (redundancy effect).

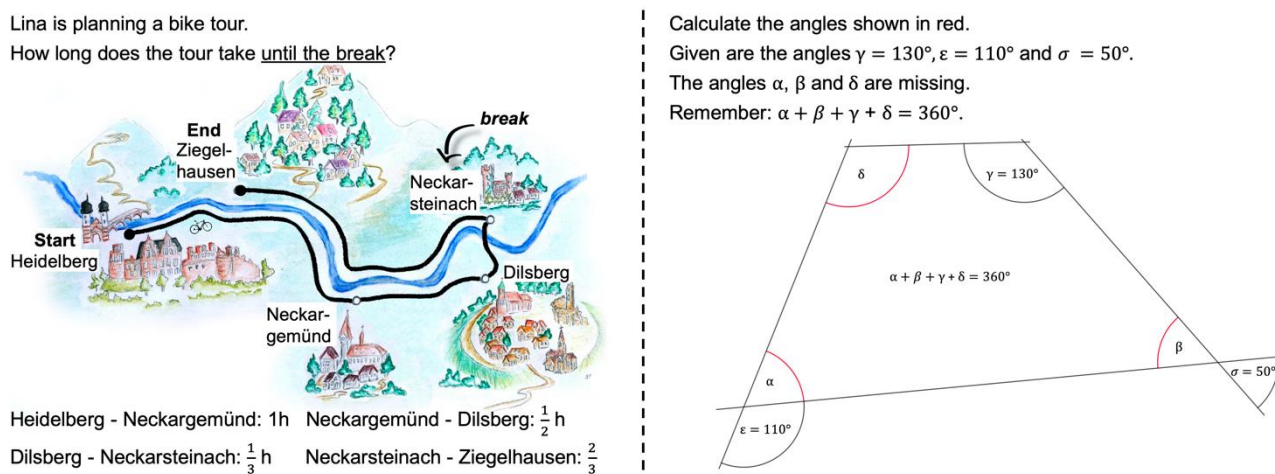


Figure 1: Sample of a fraction task and an angle task with specific difficulty-generating mathematical and instructional task characteristics

Procedure. The participating pre-service teachers were asked to assess eight mathematics tasks regarding the question "What makes the task easy/difficult for students?" The tasks were presented individually and in randomized order on a 24" LCD screen. Eye-tracking data was collected using a monitor-based eye-tracker (Tobii Pro Fusion) that captures binocular eye movements at a sampling rate of 120 Hz. For adjusting the eye-tracker, a 9-point calibration was performed before each task. The time interval between the diagnostic task on the eye-tracker and the subsequent ET SRI was kept

as short as possible to avoid loss of memory (approx. 1-3 min.). During the interview, subjects described what they did and thought during the diagnostic task, based on their shown gaze behavior. For the recording of the ET SRI, the software OBS was used, which records screen contents including sound, so that the videos of the eye movements with the corresponding comments of the subjects were available for the later analysis.

Data analysis. Tobii Pro Lab software was used to analyze the eye-tracking data. In each task, specific Areas of Interest (AOIs) were defined around the varied mathematical and instructional task characteristics. The number of fixations and fixation durations were determined using the Tobii I-VT Fixation Filter. To determine the number of transitions between two AOIs, the videos of eye movements were visually inspected. A mixed-methods approach was used to analyze the ET SRI data: The ET SRI were first transcribed and coded deductively using qualitative content analysis. The following category system was used and binary coded (in parentheses): difficulty-generating task characteristics can be *identified* (1), or *not identified* (0) when diagnosing a task. It turned out that some task characteristics are only identified when reflecting on one's own gaze behavior during the SRI. This resulted in another category *retrospectively identified*, which was evaluated as a subcategory of *not identified*. Identified task characteristics can be *correctly evaluated* in terms of difficulty for students (1) or *incorrectly / not further evaluated* (0). Task characteristics that are only evaluated during the SRI are assigned to the category *retrospectively evaluated* that counted as a subcategory of *not further evaluated*. Transcripts were coded by two raters with high interrater reliability (Cohen's Kappa = .88). The assigned codes were then integrated into a quantitative data set to examine differences across experimental conditions using variance analysis.

Preliminary Results

Identification and evaluation of difficulty-generating task characteristics

Two repeated-measures ANOVAs were calculated with the within subject factors *task characteristics* (mathematical/instructional) and the between subject factor *condition* (experimental/control group), separately for the fraction and angle tasks. Figure 2 gives an overview of the average percentage of difficulty-generating task characteristics that were identified and correctly evaluated in terms of difficulty for students.

Regarding the fraction tasks, the results showed that there is a significant difference with high effect size between the identification and evaluation of mathematical vs. instructional task characteristics ($F(1,23) = 15.01, p < .001, \eta^2 = .40$). This effect is, however, dependent on the experimental condition. Bonferroni-adjusted post-hoc analysis showed that differences between mathematical and instructional task characteristics can only be determined for participants of the control group (cf. figure 2). Here, significantly more mathematical task characteristics were identified and correctly evaluated compared to instructional characteristics (H1a). Instructional task characteristics were to a large extent not identified and adequately evaluated (H1b). Without specific knowledge, less than half of the instructional task characteristics were identified and correctly evaluated on average ($M = .42, SD = .12$). A significant difference with high effect size could be determined between experimental conditions ($F(1,23) = 31.29, p < .001, \eta^2 = .58$). Bonferroni-adjusted post-hoc analysis revealed that participants of the experimental group identified and correctly evaluated a significantly

higher number of both mathematical and instructional task characteristics (cf. figure 2). Specific knowledge about difficulty-generating task characteristics has thus enabled pre-service teachers to identify and adequately evaluate more difficulty-generating task characteristics when judging task difficulties for students (H2).

Regarding the angle tasks, however, no significant difference could be determined between the identification and evaluation of mathematical vs. instructional task characteristics ($F(1,23) = 0.43, p = .521, \eta^2 = .02$). In the case of the angle tasks, both the mathematical ($M = .34, SD = .24$) and the instructional task characteristics ($M = .29, SD = .30$) were to a large extent not identified and correctly evaluated by the participants of the control group. Thus, only H1b can be confirmed for the angle tasks. Comparing the two study groups, a significant difference with high effect size could be determined ($F(1,23) = 26.29, p < .001, \eta^2 = .53$). As can be seen in figure 2, participants of the experimental group identified and correctly evaluated a significantly higher number of both mathematical and instructional task characteristics. H2 can thus also be confirmed for the angle tasks.

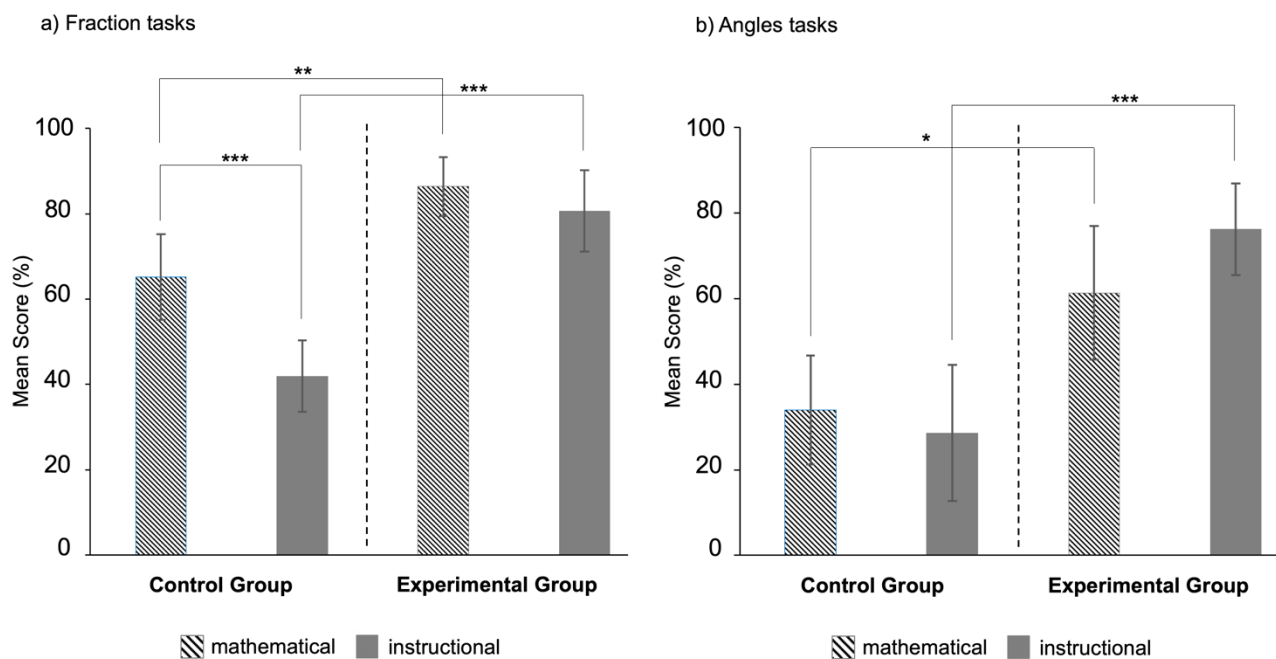


Figure 2: Means and Standard Error for identified and correctly evaluated difficulty-generating task characteristics (mathematical and instructional). * $p < .05$, ** $p < .01$, *** $p < .001$.

Analysis of gaze behavior

The collected eye-tracking measures were used as indicators of visual attention in diagnosing task difficulty. Regarding the fraction tasks, a lower number of fixations and lower fixation durations within predefined mathematical AOIs as well as shorter total recording durations could be determined for the participants of the experimental group compared to the control group. These group differences are statistically significant with high effect sizes (cf. table 1). Against the background that participants of the experimental group identified and correctly evaluated significantly more difficulty-generating task characteristics, these eye-tracking measures may indicate a more efficient approach to diagnosing with specific knowledge. For the instructional AOIs, however, no significant differences

in gaze behavior were found between the control and experimental group. Regarding the angle tasks, the eye-tracking data analysis is not yet complete but will be presented at the conference.

Table 1: Eye-Tracking measures (only fraction tasks)

AOI	ET measure	Control Group M (SD)	Experimental Group M (SD)	t	p	Cohens d
Mathematical	Fixation count	28.71 (12.71)	19.14 (7.84)	2.19	.039	0.91
	Fixation duration	8.31 (4.34)	4.89 (2.22)	2.55	.019	1.00
Instructional	Fixation count	34.35 (6.95)	34.21 (11.03)	0.04	.972	0.02
	Fixation duration	6.92 (1.37)	7.68 (2.93)	0.79	.442	0.33
	Transition count	9.04 (3.90)	8.68 (4.24)	0.22	.830	0.09
Overall	Total record duration	118.62 (12.74)	102.07 (12.03)	3.30	.003	1.34

Discussion

The aim of this ongoing study is to investigate which task characteristics (mathematical vs. instructional) pre-service teachers identify and evaluate when judging the difficulty of mathematics tasks in the domain of fractions and angles. Furthermore, a particular research interest is to explore the potential influence of specific knowledge about difficulty-generating task characteristics on identification and evaluation processes during the judgment.

In line with expectations, instructional task characteristics were to a large extent not identified and adequately evaluated, both regarding the fraction as well as the angle tasks. These findings support existing research on diagnostic teacher judgments (Schreiter et al., 2021), which showed that instructional task characteristics are insufficiently considered by teachers. In comparison to the instructional task characteristics, difficulty-generating mathematical task characteristics were identified and adequately evaluated significantly more frequently in the fraction tasks. In the case of the angle tasks, however, both mathematical and instructional task characteristics were to a large extent not identified and adequately evaluated. One explanation for this finding could be differences in prior knowledge. The results of the prior knowledge test of this study (results will be presented at the conference) showed that the participants of both study groups have more prior knowledge regarding difficulty-generating mathematical task characteristics in the domain of fractions compared to the domain of angles. Overall, the results of the study point to deficits with regard to the identification and evaluation of difficulty-generating instructional task characteristics, and, depending on the content area, also regarding mathematical task characteristics. However, for an effective mathematics instruction, teachers should choose, modify, or create tasks in such a way that task difficulties occur in appropriate doses and students are neither over- nor under-challenged (Leuders & Prediger, 2016). To do so, teachers need to be able to identify difficulty-generating task characteristics and to adequately evaluate them in terms of difficulty for students.

In line with expectations, the results showed that specific knowledge enables pre-service teachers to identify and correctly evaluate more difficulty-generating task characteristics. This effect could be found for both mathematical as well as instructional task characteristics. These results are consistent with existing research that highlights the importance of specific knowledge for the accuracy of teachers' diagnostic judgments (e.g., Ostermann et al., 2017). In our study, the collection of direct process indicators allowed to gain insights into the positive influence of specific knowledge on identification and evaluation processes that underlie teachers' judgment. Regarding the fraction tasks, the collected eye-tracking data further suggested a more efficient approach to diagnosing with specific knowledge: pre-service teachers with specific knowledge showed fewer fixations with shorter durations in task areas with difficulty-generating mathematical task characteristics. At the same time, they identified and correctly evaluated more difficulty-generating task characteristics. These results indicate that prospective teachers with specific knowledge are able to process relevant information faster, which is often seen in experts compared to novices (Grub et al., 2020). For instructional task characteristics, however, no significant group differences were found in terms of gaze behavior. It might be possible that there are no observable differences between persons who focus on task characteristics and process them fast and those who only pay little attention to the same characteristics. One possible explanation for our study finding could be that pre-service teachers without specific knowledge may have paid little attention to instructional task characteristics overall.

Overall, impulses for teacher training can be derived: An intervention on specific difficulty-generating task characteristics enables pre-service teachers to identify and adequately evaluate more difficulty-generating task characteristics and allows a more efficient approach to diagnosing. These results point to a need for learning opportunities to build specific knowledge during teacher training. Such learning opportunities should cover a wide range of difficulty-generating task characteristics (Leuders & Prediger, 2016) and typical student difficulties.

In addition to the practical relevance, the research strategic approach of this study should be emphasized: In the present study, theoretical predictions were made about how person characteristics (here: specific knowledge of prospective teachers) and situation characteristics (here: difficulty-generating task characteristics, content area of the task) influence the assumed information processing processes (here: identification and evaluation of task characteristics). These hypothesized relationships were then experimentally tested by systematically varying both situation and person characteristics and collecting process indicators (here: Eye-tracking data, and ET SRI) to examine internal judgment processes. This research strategy allows to generate knowledge about the information processing involved in the formation of diagnostic judgments, as has been widely called for (e.g., Herppich et al. 2018).

Acknowledgment

This ongoing study is carried out as part of the graduate school "DiaKom" funded by the Ministry of Science, Research and the Arts in Baden-Wuerttemberg, Germany.

References

- Brunner, K., Obersteiner, A., & Leuders, T. (2021). How prospective teachers detect potential difficulties in mathematical tasks – an eye tracking study [Special Issue]. *RISTAL*, 4, 108–125. <https://doi.org/10.23770/RT1845>
- Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., Behrmann, L., Böhmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A., Böhmer, I., Karing, C., Kaiser, A., & Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*, 76, 181–193. <https://doi.org/10.1016/j.tate.2017.12.001>
- Grub, A.-S., Biermann, A., & Brünken, R. (2020). Process-based measurement of professional vision of (prospective) teachers in the field of classroom management. A systematic review. *Journal for educational research online*, 12(3), 75–102. <https://doi.org/10.25656/01:21187>
- Karing, C., & Artelt, C. (2013). Genauigkeit von Lehrpersonenurteilen und Ansätze ihrer Förderung in der Aus- und Weiterbildung von Lehrkräften [Accuracy of teacher judgments and considerations on their improvement via teacher education]. *Beiträge zur Lehrerbildung*, 31(2), 166–173. <https://doi.org/10.25656/01:13844>
- Leuders, T., & Prediger, S. (2016). *Flexibel differenzieren und fokussiert fördern im Mathematikunterricht* [Flexible differentiation in secondary mathematics]. Cornelsen.
- Loibl, K., Leuders, T., & Dörfler, T. (2020). A framework for explaining teachers' diagnostic judgements by cognitive modeling (DiaCoM). *Teaching and Teacher Education*, 91, 1–10. <https://doi.org/10.1016/j.tate.2020.103059>
- Moreno-Esteva, E. G., White, S., Wood, J., & Black, A. (2017). Identifying key visual-cognitive processes in students interpretation of graph representations using eye-tracking data and machine learning based data analysis. In T. Dooley, & G. Gueudet, (Eds.), *Proceedings of the Tenth Congress of the European Society for Research in Mathematics Education* (pp. 3928–3935).
- Ostermann, A., Leuders, T., & Nückles, M. (2017). Improving the judgment of task difficulties: prospective teachers' diagnostic competence in the area of functions and graphs. *Journal of Mathematics Teacher Education*, 21(6), 579–605. <https://doi.org/10.1007/s10857-017-9369-z>
- Padberg, F., & Wartha, S. (2017). *Didaktik der Bruchrechnung* [Didactics of fractions]. Springer Spektrum. <https://doi.org/10.1007/978-3-662-52969-0>
- Reiss, K. (2002). *Argumentieren, Begründen, Beweisen im Mathematikunterricht* [Arguing, reasoning, proving in mathematics education]. Projektserver SINUS. <http://blk.mat.uni-bayreuth.de/material/db/59/beweis.pdf>.
- Schindler, M., & Lilienthal, A.J. (2019). Domain-specific interpretation of eye tracking data: towards a refined use of the eye-mind hypothesis for the field of geometry. *Educational Studies in Mathematics*, 101(1), 123–139. <https://doi.org/10.1007/s10649-019-9878-z>

Schreiter, S., Vogel, M., Rehm, M., & Dörfler T. (2021). Teachers' diagnostic judgment regarding the difficulty of fraction tasks: A reconstruction of perceived and processed task characteristics [Special Issue]. *RISTAL*, 4, 127–146. <https://doi.org/10.23770/RT1846>

Sweller, J., Ayres, P., & Kalyuga, S. (Eds.). (2011). *Cognitive Load Theory*. Springer. <https://doi.org/10.1007/978-1-4419-8126-4>