Teacher noticing in inclusive mathematics education: Analyzing its structure and expert-novice differences using a novel video-based test instrument

Electronic Supplementary Material (ESM)

Measurement instruments

Teacher noticing in inclusive mathematics education

Building on existing widely used and thoroughly validated video-based assessment instruments (Kaiser et al., 2015), a new video-based instrument was developed to capture perception, interpretation, and decision-making in inclusive mathematics education from both mathematical pedagogical and general pedagogical perspectives.

To this end, it was necessary to depict typical situations in inclusive mathematics classes in the form of scripts.

As part of the test instrument development process, scripts were initially developed in collaboration with experienced teachers, showcasing excerpts from such inclusive mathematics classes—focusing on inclusive algebra instruction in secondary education. Within these excerpts, various areas of learning prerequisites were to be diagnostically identifiable, as well as the (differently successful) support measures taken by the teacher. After a revision of the scripts based on expert review, the scripts were filmed with students from a school in Hamburg and processed into short video clips.

The final instrument consists of four video vignettes, each lasting between 2.45 and 4.00 minutes, covering a wide range of content areas in algebra instruction. Before viewing a vignette, participants were given an overview of the subsequent teaching situation, including the grade level, type of school, topic of the lesson unit, lesson objective, current knowledge level of the learning group, and the structure of the presented lesson unit. Participants also received the task that was central to the lesson, as well as the solution to that task, thereby minimizing any effects of differing prior knowledge. Participants then viewed the video vignette. Each vignette features students with varying levels of learning prerequisites. This includes students who demonstrate the necessary learning prerequisites to achieve the lesson objective, such as required prior knowledge, strategic approaches, or motivation. However, each vignette also includes students who lack relevant prior knowledge or who face challenges in areas such as motivation and self-concept, attention and working memory, as well as learning strategies and metacognitive regulation. These students require support to successfully learn in the context of the presented lesson. Additionally, each vignette already showcases specific support offers, such as subject-related feedback or strategy-related support. A description of the four video vignettes can be found in ESM Table 1.

ESM Table 1
Overview of the four video vignettes

Title of the vignette	*Duration*	*Topic / Context*
"Farm"	3.44 min	Formulating and Solving Equations / Grade 8, Secondary Education
Description		
The class is working on a word problem to determine how many sheep and chickens live on a farm, given that there are a total of 162 animals and that there are five times as many sheep as there are chickens. A group out of five students receive strategy-based support through a poster with questions for comprehending texts, as well as cards with strategic approaches (e.g., creating a sketch). Through their statements, differences in prior knowledge become evident; for instance, one student is unaware that letters in algebraic notation represent variables. The students also differ in their strategic approaches and motivation. One student completely refuses to participate. At the same time, the other learners quickly grasp that the problem can be solved using a system of equations. However, the group makes an error by directly translating the language into the equation formulation (5S + C = 162; S = Number of Sheep; C = Number of Chickens), resulting in more chickens than sheep. A drawing created by the student with least prior knowledge helps the group recognize their mistake.		
Title of the vignette	*Duration*	*Topic / Context*
"Image sequence"	3.05 Min	Representing Structures Using Algebraic Expressions / Grade 6, Academic Track
The students work individually or in pairs on a worksheet featuring a sequence of images made up of square arrangements of points. Starting with 8 points (Figure 1), 4 additional points are added with each new figure. The learners are tasked with counting the points in bundles and determining how many points the 100th figure will have. Students with varying levels of prior knowledge engage with the problem. One student shows signs of difficulties with attention as well as with counting in bundles. Other learners find bundled counting easy and discuss different methods of bundling. Another student is able to count, but does not grasp the underlying systematics, which is needed to determine the 100th figure. The teacher provides different feedback to each student regarding their work on the task. While she directly demonstrates the bundling approach for one learner, she also encourages the students to explain their own solution strategies.		
Title of the vignette	*Duration*	*Topic / Context*
"E-Scooter"	2.40 Min	Expressions and Variables in the Context of Functional Relationships
The students are tasked with determining the costs of different E-scooter rides based on a base fee and the cost per minute. Building on this, they are encouraged to formulate a general expression for the costs using x as a variable. They can check their solutions on the worksheet using an interactive learning video via tablet. Both the worksheet and the learning video contain prompts to support metacognition (e.g., prompts for solution verification and formulating questions). The students work in groups of three. One student is able to solve the task confidently and employs strategic methods, meaning she works systematically and checks the results for plausibility. The other group members face both subject-related challenges (e.g., difficulties with multiplying decimal numbers) and issues related to self-regulation and metacognition. For instance, they become distracted by the tablet and experiment with different solutions ("trial-and-error") instead of reflecting on why their chosen solution is incorrect		
Title of the vignette	*Duration*	*Topic / Context*
"Scale model"	3:58 Min	Formulating and Transforming Equations / Grade 7, Secondary School
In the class, the principle of equations was illustrated using a balance scale, where the same number of small counting stones must always be on both sides. Variables were represented as matchboxes in which the stones can be packaged. The students were tasked with determining the value of x for different equations (e.g., x +4 = 12. The video shows a guided classroom discussion aimed at systematizing the results. In the discussion, the teacher prompts the students to verbalize their problem-solving processes. The students employ different approaches, namely physically weighing the stones (enactive), creating a sketch (iconic), and using algebraic expressions (symbolic), and verbal descriptions (verbal). During the discussion, the teacher encourages the students to describe		

their methods and provides corrective feedback. The teacher successfully links the different levels of representation by, for example, asking a student to mentally place the drawing on the balance scale. In this way, the students experience the principle of equivalence transformation on various levels to determine the value of the variable x. The students' statements suggest varying levels of prior knowledge. For instance, one student does not understand the variable as a placeholder for a number. Another student shows signs of a significantly low self-concept regarding mathematics.

To closely simulate real classroom conditions for ad-hoc diagnosis and support during instruction, each vignette can only be viewed once. After each video, participants are presented with test items to assess their noticing skills. Each item focuses on (1) one of the three facets, that is, perception, interpretation, and decision-making, (2) either mathematics pedagogy or general pedagogy, and (3) either diagnosis or support. For each of the three noticing facets, both closed response format (i.e., rating scales and dichotomous items) and open response formats are used.

Test items focusing on the cognitive process of perception exclusively address the domain of diagnosis, assessing the extent to which participants perceive directly observable indicators of necessary learning prerequisites or the absence of such prerequisites. This conceptualization of perception differs from the conceptualization of attending to events, used as first facet in other conceptualizations of noticing (such as by van Es & Sherin, 2002). Open response formats require descriptive statements about events shown in the video. Items with closed response formats contain descriptive statements about the observed instruction, which participants have to evaluate as true or false (dichotomous response format). In the shown example (ESM Figure 1), participants are asked to indicate that the student, when counting a square arrangement of points, first counts the four corners (4 x 1) and then bundles the four spaces in between (4 x 4) to arrive at a total of 20 points. For a correct answer, this procedure merely has to be described. Interpretive conclusions about learning prerequisites are not required.

Items focusing on the cognitive process of interpretation, in contrast to perception items, require interpretive conclusions about classroom observations, for example, drawing on professional knowledge. The items address both the area of diagnosis, which involves inferring the (non-)existence of relevant learning prerequisites, and the area of intervention, such as conclusions about the purpose of support measures or specific feedback from the teacher. Open items either provide an interpretation and ask for indicators that supported this interpretation, or conversely, present specific events from the video and require an interpretation. Closed items contain interpretive statements about the teaching events, for which participants could indicate their level of agreement on a four-point scale (strongly agree, agree, disagree, strongly disagree). As the sample item shows (see ESM Figure 2), participants are required to recognize two of the displayed indicators for a possible attention difficulty. That is, the student does not focus on the task, interrupts the counting process multiple times, becomes discouraged, abandons the task after a few unsuccessful attempts, and finally behaves in a restless manner. These indicators are based on the literature concerning attention as a learning prerequisite but are not exclusively signs of attention problems, which is why the wording of the task was formulated carefully.

ESM Figure 1

Test Task for Assessing Perception with a Focus on Mathematics Education and Diagnosis

	Summarize in bullet points how Noah sums up the number of points in the point pattern (bundling).

ESM Figure 2

Test Task Focusing on Assessing Interpretation with a Focus on Pedagogy and Diagnosis

	Name two indicators that suggest Mica may have attention difficulties.

Items measuring the cognitive process of decision-making exclusively relate to the professional demand of intervention. Central aspects of the teaching events are already provided in the tasks, so that it was necessary for participants to formulate alternative actions (e.g., specific support measures) or further courses of action. Open test tasks directly focus on generating action options. Closed test tasks already present the possible actions, requiring participants to assess how helpful the proposed approach would be (see example task in ESM Figure 3). In the example task, participants have to recognize that an enactive approach using plastic tiles would be of little help for two students—taking into account their prior knowledge; rather, requiring mathematical justifications could contribute to the further development of the learning process. The development of the test items was guided by theory and occurred in parallel with the development of the video vignettes. In the context of a review by 47 academic experts in mathematics education, pedagogy, and special education, the appropriateness of the answers coded as correct was validated. The experts indicated their agreement with the marked correct answer in the case of closed response formats. For open test tasks, the experts were provided with (theory-driven) solution categories, to which they could report their agreement. In cases of insufficient agreement, as well as based on further comments from the experts, the items were finalized for the instrument used in the main study. A dichotomous coding was used, that is, each answer was either coded as correct or as incorrect. For the open response items, the participants' answers were coded by trained raters based on comprehensive coding manual. 70 cases were coded by two raters independently, resulting in good interrater reliability (Kappa: Mean: 0.77, Min: 0.27, Max: 1.00). A list of the kappa values for each item is provided in ESM Table 2. In a few cases, agreement was initially low. For these items, the coding manual was reviewed and slightly revised, resulting in satisfactory agreement levels for the final 20 double-coded cases. Discrepancies were subsequently discussed within the research team. In cases of coder uncertainty, problematic responses were consistently reviewed and resolved collaboratively within the team.

ESM Table 2

Overview of kappa values for each category

Item (M-PID)	$N_{valid\ responses}$	Cohens κ	Item (P-PID)	$N_{valid\ responses}$	Cohens κ
ba11_imo	58	0.601	ba09_ppo	49	1,000
bi07_pmo	67	0.944	ba10_ipo_tex	54	0.823
bi10_imo_zae	64	0.946	ba10_ipo_hil	54	0.849
bi10_imo_bue	64	0.730	ba12_dpo_fra	55	0.920
bi11_imo_pia	62	0.812	ba12_dpo_kom	55	0.681
bi11_imo_noa	62	0.962	ba12_dpo_zug	55	0.521
bi13_dmo_iko	49	0.913	ba12_dpo_sin	55	0.658
bi13_dmo_str	49	0.616	ba12_dpo_gru	55	0.824
es09_pmo	65	0.923	ba13_dpo_emp	43	0.365
es12_imo_kor	65	0.798	ba13_dpo_erk	43	0.269
es12_imo_abh	65	0.604	bi06_ppo_abr	61	0.816
es12_imo_spa	65	0.838	bi06_ppo_emo	61	0.685
es13_imo_fee	56	0.794	bi08_ipo_bli	63	0.890
es13_imo_hil	56	0.829	bi08_ipo_zae	62	0.835
es13_imo_str	56	1,000	bi08_ipo_fru	62	0.805
es14_dmo_erk	65	0.817	bi08_ipo_mot	62	1,000
es14_dmo_bei	65	0.885	bi09_ipo_lei	46	0.598
wa12_imo_erk	57	0.587	bi09_ipo_ans	46	0.621
wa12_imo_waa	57	0.665	bi12_dpo_abl	58	0.901
wa13_imo_ena	65	0.759	bi12_dpo_rhy	58	0.660
wa13_imo_sym	65	0.885	bi12_dpo_bel	58	0.498
wa13_imo_iko	65	0.877	bi12_dpo_auf	58	0.623
wa15_dmo_ein	42	0.822	bi12_dpo_smo	58	0.659
wa15_dmo_erk	42	0.836	es08_ppo_leo	62	0.816
wa16_dmo_ena	65	0.726	es08_ppo_pau	62	0.751
wa16_dmo_iko	65	0.927	es10_ipo_obe	60	0.472
			es10_ipo_rei	60	0.786
			es10_ipo_tab	60	0.719
			es11_ipo_ver	58	0.937
			es11_ipo_fra	58	0.814
			wa09_ppo	63	0.851
			wa10_ipo_aku	57	1,000
			wa10_ipo_kla	57	0.859
			wa11_ipo_for	51	0.918
			wa14_dpo_kom	62	0.766
			wa14_dpo_unt	62	0.514
	Mean κ	0.811		Mean κ	0.742
	SD κ	0.118		SD κ	0.175
	Min κ	0.587		Min κ	0.269
	Max κ	1.000		Max κ	1.000

As part of the item review process, a minimum item-total correlation of .10 was defined as the threshold. This liberal cutoff was chosen due to the broad nature of the construct, which encompasses multiple content domains (diagnosis and intervention; mathematics education and general pedagogy) and facets (perception, interpretation, and decision-making). Three items were excluded from the analysis based on their low discriminatory power. Upon further inspection, these items were found to be somewhat ambiguous or unclear, which further justified their exclusion. After excluding these three items with problematic psychometric properties, the final test instrument included 63 dichotomous items (M_PID-IT: 32; P_PID-IT: 31). Following common practice in performance assessment, also in the context of teacher noticing, item responses were scored dichotomously. This approach helps to prevent issues such as reversed item thresholds in partial credit models.

To reduce testing time and mitigate fatigue effects, a random incomplete test design was implemented, in which participants worked on three out of four video vignettes based on random selection. Participants had exactly 20 minutes to view a video and complete the subsequent test tasks. Thus, the video-based testing lasted a total of 60 minutes.

ESM Figure 3

Task for assessing decision-making ability with a focus on mathematics didactics and intervention

Noah and Pia discuss their structuring of the fourth figure. As the teacher, you want to support them in their learning. Evaluate the extent to which the following approaches are helpful.

Pia and Noah receive plastic tiles with which they should create their patterns.	Very helpful ☐	Somewhat helpful ☐	Somewhat unhelpful ☐	Not helpful at all ☐
Prompt: "Imagine someone in the class does not see that 4 plus 4 times 4 is the same as 5 times 4. Can you, as mathematicians, explain to them why this is the case?"	Very helpful ☐	Somewhat helpful ☐	Somewhat unhelpful ☐	Not helpful at all ☐

Knowledge measures

To assess the participants' MPCK-IT and GPK-IT, instruments were developed based on existing paper-and-pencil tests. Only closed item formats, that is, single-choice items, are used with each item being coded as either correct (1) or incorrect (0). The knowledge assessment lasted approximately 25 minutes.

For MPCK-IT, specific items from previous knowledge tests developed in TEDS-M (Blömeke et al., 2010) and further developed in the TEDS research program published as TEDS-shortM (Buchholtz et al., 2014) related to inclusive teaching were selected and further developed with a focus on understanding-oriented teaching as overall mathematics educational foundation and differentiation of tasks and task complexity, adaptive teaching, scaffolding as specific focus for inclusive mathematics education. The instrument consists of a total of 37 items and covers the fundamentals of algebra didactics, such as understanding variables and corresponding basic concepts, general topics concerning high-quality mathematics instruction like orientation towards understanding, as well as diversity-related topics such as dealing with heterogeneity regarding prior knowledge and language in mathematics instruction (example items in ESM Figure 4). The appropriateness of the answers coded as correct was examined through an expert review by teacher educators at university level, who provided detailed feedback. We focused deliberately on university-based teacher educators as we know from earlier studies that as we know from earlier expert ratings in the frame of the TEDS-Follow-up study that the group of university-based teacher educators evaluate items differently than school-based teacher educators. Due to space restrictions we cannot discuss this important aspect in detail.

As part of the item review process, it was determined that items in the knowledge tests—designed to assess somewhat more homogeneous constructs—should exhibit a minimum item-total correlation of .15. This threshold allowed for the inclusion of as many items as possible in the analysis, thereby ensuring that the constructs could be represented with the necessary breadth. Consequently, 8 items were removed from analysis. Each item exclusion was thoroughly discussed within the research team to ensure that all relevant aspects of the knowledge constructs were adequately represented.

In line with an existing test of GPK-IT (König et al., 2017), this area of knowledge was conceptualized for the present study based on the two professional demands of diagnosis and intervention. Specifically, topics that play a role in the underlying concept of inclusive teaching in the vignette testing were taken into account, including diagnosis and support regarding heterogeneous prior knowledge, learning strategies and metacognitive regulation, motivation and self-concept, as well as working memory and attention (example items in ESM Figure 5). In addition to ten items from the existing GPK-IT test, three items on pedagogical knowledge (Blömeke et al., 2010) were included, and six items were newly developed with regard to the specific project focus. The newly developed items were reviewed by three authors of the original GPK-IT test. Again, low item discrimination (item-total correlation < .15) lead to excluding three items.

ESM Figure 4

Example item for MPCK-IT with a focus on understanding vs. orientation towards the calculation.

Distinguish between the following algebraic activities in teaching as being either more focused on understanding or more calculation-oriented.

Check <u>one</u> box for each.

	Understanding-oriented	Calculation-oriented
Transform a term into another term according to fixed transformation rules (e.g., simplifying)	☐	☑
Describe a situation, a pattern, or a functional relationship using two different terms.	☑	☐

ESM Figure 5

Example item GPK-IT with a focus on intervention (A) and diagnosis (B)

(A) Metacognitive strategies show high relevance for learning success in studies. Which statement is correct?

Please check one box.

a.	Metacognitive strategies include the planning, monitoring, evaluation, and regulation of the learning process.	☑1
b.	Metacognitive strategies include the repetition, organization, and elaboration of the learning material.	☐2
c.	Metacognitive strategies primarily concern the motivational and volitional regulation of the learning process.	☐3
d.	Metacognitive strategies are mainly helpful for students with learning disabilities, but not for high-achieving students.	☐4

(B) Which teaching technique can serve as a support system for language-sensitive subject teaching across school subjects?

Please check one box.

a.	Scaffolding	☑1
b.	Group work	☐2
c.	Corrective feedback	☐3
d.	Consideration of learning styles	☐4

References:

Blömeke, S., Kaiser, G. & Lehmann, R. (2010) (Eds). *TEDS-M 2008 – Professionelle Kompetenz und Lerngelegenheiten angehender Mathematik-Lehrkräfte für die Sekundarstufe I im internationalen Vergleich*. Münster: Waxmann.

Buchholtz, N., Scheiner, T., Döhrmann, M., Suhl, U., Kaiser, G. & Blömeke, S. (2012). *TEDS-shortM. Teacher Education and Development Study – Short Test on Mathematics Content Knowledge (MCK) and Mathematics Pedagogical Content Knowledge (MPCK). Kurzfassung der mathematischen und mathematikdidaktischen Testinstrumente aus TEDS-M, TEDS-LT und TEDS-Telekom*. University of Hamburg.

König, J., Gerhard, K., Melzer, C., Rühl, A.-M., Zenner, J., & Kaspar, K. (2017). Erfassung von pädagogischem Wissen für inklusiven Unterricht bei angehenden Lehrkräften: Testkonstruktion und Validierung. *Unterrichtswissenschaft*, *45*, 223–242.

Van Es, E. & Sherin, M. G. (2002). Learning to Notice: Scaffolding New Teachers' Interpretations of Classroom Interactions. *Journal of Technology and Teacher Education*, 10(4), 571-596.

Additional Statistics

ESM Table 3

Model comparison for each experience group

	Model		Deviance	BIC	AICc	Model comparison using χ^2-tests against the less complex model					
						1)		2)		3)	
						χ^2 (*df*)	*p*	χ^2 (*df*)	*p*	χ^2 (*df*)	*p*
Master students	1)	1D	15701.2	16254.5	16502.6						
	2)	2D-MP	15692.4	16255.8	16549.2	8.76 (2)	< .05				
	3)	2D-KS	15669.1	16232.6	16525.9	32.04 (2)	< .001				
	4)	4D	15644.2	16242.8	16747.6	57.00 (9)	< .001	48.24 (7)	< .001	24.96 (7)	< .001
Teachers in preparatory service	1)	1D	14729.2	15277.3	15646.9						
	2)	2D-MP	14701.7	15259.9	15692.7	24.77 (2)	< .001				
	3)	2D-KS	14672.0	15230.2	15663.0	57.15 (2)	< .001				
	4)	4D	14632.4	15225.5	15968.9	96.75 (9)	< .001	69.30 (7)	< .001	39.60 (7)	< .001
In-service teachers	1)	1D	15781.0	16337.2	16531.9						
	2)	2D-MP	15748.7	16315.0	16548.0	32.32 (2)	< .001				
	3)	2D-KS	15708.4	16274.7	16507.7	72.64 (2)	< .001				
	4)	4D	15669.8	16271.5	16679.7	111.23 (9)	< .001	78.91 (7)	< .001	38.59 (7)	< .001

Note. BIC - Bayesian information criterion. AICc - Sample-size corrected Akaike information criterion. 1D - one-dimensional holistic model of competence. 2D-MP - two-dimensional model of mathematics and general pedagogy. 2D-KS - two-dimensional model of knowledge and situation-specific skills. 4D - four-dimensional model of two-dimensional knowledge (MPCK-IT and GPK-IT) and two-dimensional situation-specific skills (M_PID-IT and P_PID-IT).

ESM Table 4
Reliabilities and latent Correlations for each experience group in the four-factor model

		Dimension	Variance	EAP	WLE	Latent correlations			
Master students						M_PID-IT	P_PID-IT	MPCK-IT	GPK-IT
	M4	M_PID-IT	0.30	.74	.54	1.00			
		P_PID-IT	0.19	.65	.44	.88	1.00		
		MPCK-IT	0.39	.72	.64	.71	.50	1.00	
		GPK-IT	0.52	.71	.58	.69	.61	.65	1.00
		Dimension	Variance	EAP	WLE	Latent correlations			
Teachers in preparatory service						M_PID-IT	P_PID-IT	MPCK-IT	GPK-IT
	M4	M_PID-IT	0.45	.77	.64	1.00			
		P_PID-IT	0.44	.71	.63	.69	1.00		
		MPCK-IT	0.47	.76	.69	.68	.35	1.00	
		GPK-IT	0.55	.76	.55	.79	.57	.78	1.00
		Dimension	Variance	EAP	WLE	Latent correlations			
In-service teachers						M_PID-IT	P_PID-IT	MPCK-IT	GPK-IT
	M4	M_PID-IT	0.44	.78	.61	1.00			
		P_PID-IT	0.38	.70	.61	.79	1.00		
		MPCK-IT	0.59	.81	.74	.78	.47	1.00	
		GPK-IT	0.55	.70	.58	.53	.39	.74	1.00

Note. EAP - Expected a posteriori estimator. WLE - weighted likelihood estimator. M_PID-IT - mathematics instruction for inclusive teaching: perception, interpretation, decision-making. P_PID-IT - pedagogy for inclusive teaching: perception, interpretation, decision-making. MPCK-IT - mathematics pedagogical content knowledge for inclusive teaching. GPK-IT - general pedagogical knowledge for inclusive teaching.

ESM Table 5

Mean scores of percentage solution frequency for all competence facets by experience group

Variable			Experience groups						Group differences		
	Total (n = 628/456)[1]		MS (n = 233/153)[1]		TPS (n = 163/146)[1]		IST (n = 232/157)[1]				
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*	*p*	η^2
M_PID-IT	42.90	14.73	43.28[a,b]	13.29	45.35[a]	15.53	40.79[b]	15.30	4.77	.009	.015
P_PID-IT	47.03	15.23	51.49[a]	13.60	46.32[b]	15.84	43.06[b]	15.21	19.08	< .001	.058
MPCK-IT	63.40	15.68	64.43[a]	14.03	64.82[a]	15.32	61.07[a]	17.28	2.69	.069	.012
GPK-IT	60.51	17.78	61.35[a]	17.91	61.77[a]	17.08	58.54[a]	18.25	1.50	.224	.007

Note. M_PID-IT - mathematics instruction for inclusive teaching: perception, interpretation, decision-making. P_PID-IT - pedagogy for inclusive teaching: perception, interpretation, decision-making. MPCK-IT - mathematics pedagogical content knowledge for inclusive teaching. GPK-IT - general pedagogical knowledge for inclusive teaching. MS - Master students, TPS - Teachers in preparatory service, IST - In-service teachers.

[1]All valid cases were included for both tests with their two dimensions respectively, resulting in two sample sizes. The first sample size belongs to the PID-IT test, the second to the knowledge test.

Means within a row that share the same superscript letter do not differ significantly according to the post-hoc test.