Deepening prospective mathematics teachers' diagnostic judgments: Interplay of videos, focus questions and didactic categories

Susanne Prediger & Carina Zindel
TU Dortmund University, Dortmund, Germany
For correspondence: prediger@math.uni-dortmund.de

Abstract
This article combines different conceptualizations of teachers' diagnostic competence in listening to students' mathematical thinking processes on the levels of general perspectives, noticed aspects and activated didactic categories. An empirical study of 159 prospective mathematics teachers' diagnostic judgments investigated how these levels are related and how they can be enhanced by focus questions and an intervention introducing relevant categories. For this purpose, 2×159 written diagnostic judgments, before and after a short intervention, were compared with respect to general perspectives, activated categories and aspects noticed in the video. As a means of comparing, a coding scheme was developed with respect to a sound conceptualization of diagnostic competence and to interrater reliability. The comparison shows that video alone cannot enhance prospective teachers' capacities of noticing, hence the activation and reflection of didactic categories is crucial. Possible theoretical consequences for a multi-facetted conceptualization of diagnostic competence and practical consequences for teacher education courses are discussed.

Keywords: Prospective teachers' diagnostic judgments, Videos, focus questions.

Introduction

It is widely accepted that a student-centred mathematics classroom requires teachers' competence innoticing and listening to student thinking in a profound manner (surveyed by Empson & Jacobs, 2008), here called *diagnostic competence*. Diagnostic competence has gained increasing attention in research on professional knowledge and professional development for practicing teachers (e.g., by the use of videos; Sherin & Es, 2009), and prospective teachers (Clarke et al., 2011; Girulat et al., 2013), and this research has resulted in different suggested conceptualizations with varying in scopes and priorities. In this article, we order these conceptualizations on three levels:

- (P) the level of general diagnostic perspectives(initially resulting from general attitudes),
- (A) the level of noticed aspects and
- (C) the level of categories activated for analysing and interpreting.

As many authors implicitly or explicitly assume some connections between these levels, we pursue the research question of how these three levels are connected in prospective teachers' diagnostic judgments and how the diagnostic judgments can be enhanced by minimal interventions.

The aim of this article is to contribute to the empirically based knowledge about conditions of video use and the interplay of general attitudes (expressed in general diagnostic perspectives), noticed aspects and underlying didactic categories. We briefly examinethe theoretical background on diagnostic competence and the use of videos for enhancing it. We analyse the material used for the presented study before formulating refined research questions and presenting the methods of

empirically investigating 2×159 written diagnostic judgments. The main section reports on selected results, showing the necessity of adequate focus questions, which can influence the use of categories andactivation of general diagnostic perspectives. The connection between activated categories and addressed aspects is not straightforward, but the change resulting fromthe intervention shows the crucial importance of didactic categories.

Theoretical background: Enhancing teachers' diagnostic judgments by the use of videos

Different conceptualizations of diagnostic competence between general diagnostic perspectives, noticed aspects and didactic categories.
Teachers' competence for listening to student thinking (referred to as diagnostic competence; see Helmke et al., 2003; Brunner et al., 2013) has gained an increasing attention in mathematics teacher research due to its high relevance for mathematics teaching and learning. Many authors point out that teachers' diagnostic competence

- is an important condition for student-centred teaching approaches in which the teacher is responsive to students' ideas (Ball & Cohen, 1999;Empson& Jacobs, 2008)
- is a necessary condition for adaptive teaching in which a teacher is attentive to students' specific challenges (Even & Tirosh, 2002; Empson & Jacobs, 2008)
- has an empirically proven impact on students' learning success (Wilson &Burne, 1999; Franke et al., 2001; Brunner et al., 2013)
- once acquired as expertise, can provide a long-term opportunity to continuous individual professional development for teachers by learning from their students (Franke et al., 2001).

However, the community deals with various divergent conceptualizations of diagnostic competence. Most quantitative studies have referred to accuracy of predicting student achievement in tests (ranks, success rates or variances for items or students; e.g., Schrader & Helmke, 1987; Brunner et al., 2013) and some to speed of evaluating student utterances (Krauss & Brunner, 2011). Both conceptualizations seem to only partly address what is required for supporting student-centred, adaptive teaching in a profound sense. Therefore, they are not considered in this article.

Most theoretical or qualitative studies on diagnostic competence or teachers' listening to students have suggested more holistic conceptualizations with a stronger emphasis on unfolding students' uttered ideas rather than predicting their success without seeing the student (see Empson & Jacobs, 2008, for an overview and Prediger, 2010a, for a critique). In many of these qualitative studies or theoretical papers, diagnostic competence has been characterized by general attitudes expressed in general diagnostic perspectives, which can be organized in three dichotomies:

(P1) Surface judgment vs. deeper focus on mental aspects. In particular, the *speed of evaluation* construct is contradictory to many conceptualizations of diagnostic competence, which emphasize the need to overcome *surface judgments* of students' utterances and develop a *deeper focus on mental aspects*rather than surface aspects in order to analyse students' thinking systematically (Empson & Jacobs, 2008).

(P2) Deficit orientation vs. resource orientation.An interest in mental aspects of student thinking should not only be encouraged by simple studies of errors and misconceptions in teacher education classrooms, because this risks stabilizing a deficit-oriented attitude towards students. Instead, a *resource-oriented perspective* has been emphasised within constructivist perspectives (Selter, 2001; Smith et al., 1993).

(P3) Directive listening vs. reconstructing inner logic. A third general attitude to be overcome for diagnostic competence has been characterized by *directive listening*, i.e., evaluating whether students utterances match expected responses (Empson & Jacobs, 2008, p. 268). The oppositeis a so-called interpretative attitude (Jungwirth et al., 2001), consisting of being able to *reconstruct the inner logic* of student thinking, even if it does not appear logical in the adults'

view (e.g., Jungwirth et al., 2001; Smith et al., 1993). Arcavi and Schoenfeld (2008) have described this by the fruitful metaphor 'decentering'.

Beyond these general perspectives, an expert diagnostic judgment can be characterized by the richness of the *aspects noticed* by the expert. For example, Morris et al. (2009) showed that unpacking the mathematical sub-goals of a task is an important facet of diagnostic competence when the diagnostic judgment is a base for planning the further learning process.

Sherin et al.'s (2011) construct of *noticing* refers to the complexity of whole-class teaching and emphasises the need for selecting aspects by defining noticing as 'the processes through which teachers manage the blooming, buzzing confusion of sensory data with which they are faced . . . during instruction' (Sherin et al., 2011, p. 5). Hence, noticing selected aspects involves filters of perception and consists of two processes:' attending to particular events in an instructional setting 'and' making sense of events in an instructional setting' (ibid., p. 5). For the second process of making sense, they hint at the relevance of categories: Interpreting means 'relating observed events to abstract categories and characterizing what they see in terms of familiar instructional episodes' (ibid., p. 5).

This conceptualization of noticing as selecting and interpreting by categories resonates with the emphasis on general and topic-specific *categories activated* for perceiving and interpreting phenomena during a diagnostic judgment: 'A description of persons [or their utterances] only becomes a diagnostic analysis when it is based on an explicit theoretical fundament, i.e., when based on a categorical system or sensitizing concepts' (Helmke et al., 2003, p. 19). Hence, Jungwirth et al. (2001), Prediger (2010a) and Schwarz et al. (2008)have included the application of systematic categories into their definition of diagnostic competence, even if they refer noticing to more narrow situations than Sherin et al. (2011). Also Walkoe (2014) and Busch et al. (2015) emphasise the relevance of applying (in these cases topic-specific) categories.

Summing up, existing conceptualizations of diagnostic competence vary in scope (e.g., complexity of whole-class teaching or situations with one or two students) and assume different priorities and relations between general diagnostic attitudes, activated categories and what is noticed. As this paper refers to introductory lectures in initial teacher education rather than continuous professional development of practicing teachers (such as Sherin et al., 2011), the scope of noticing is here restricted to diagnostic situations with one or two students in order to reduce complexity. Notwithstanding these restrictions, we combine the sketched conceptualizations and assume that diagnostic judgments can be characterized by (P) the underlying general diagnostic perspectives, (A) the aspects noticed and (C) the categories applied.

Investigating the *relation between perspectives*, noticed aspects and knowledge about categories in teachers' diagnostic judgments is still an important issue of research, as Schoenfeld (2011) and Dreher and Kuntze (2014) outline. As a reaction to this research need, different empirical investigations have been conducted in order to unfold diagnostic competence manifestedin diagnostic judgments: Schwarz et al. (2008) investigated the diagnostic judgments of 491 prospective secondary teachers on students' solutions for a modelling problem. The diagnostic judgments were analysed with respect to the number of noticed aspects and how these aspects were related to categories. Prediger (2010a) showed how categories for interpreting the equal sign were necessary for diagnosing students' process of meaning construction and how these categories were connected to general attitudes and perspectives. Walkoe (2014) showed that priming teachers with topic-specific categories can enhance what they notice and in particulardiscussed student algebraic thinking as they receive vocabulary and sensitizing lenses. When Busch et al. (2015) evaluated a continuous professional development course with respect to the development of 26 practicing teachers' diagnostic judgments on students' conceptions of functions, they focused on the general diagnostic perspectives (deficitorientation and

focus on deep mental aspects) and systematic use of categories, namely conceptions of functions; their professional development contributed to deepening teachers' judgments and to achieve more fruitful perspectives. Philipp and Leuders (2014) compared three teachers' and three teacher educators' diagnostic judgments with respect to the relation of perspectives, applied knowledge and noticed aspects. They confirmed the expected tendency that the teacher educators noticed a greater number of different aspects, had a stronger resource orientation and drew more explicitly on their knowledge of conceptions of fractions. Dreher and Kuntze (2014) investigated the theme-specific noticing of 67 pre-service and 77 in-service teachers on chances and difficulties of using multi-representations to understand fractions. They systematically showed that the adequacy of theme-specific noticing is directly connected to underlying knowledge of categories and possible chances and risks, since aspects and categories are related statistically, by analysing the teachers' written judgments of videovignettes.

Based on these existing insights into the connection between perspectives, aspects noticed and categories applied, the current study investigates these connections in prospective teachers' diagnostic judgments on students' modelling processes and examines how they can be deepened by means ofa short intervention. In this way, we intend to contribute to the important emphasis on topic-specific categories for profound diagnostic judgments.

It has often been emphasized that developing profound diagnostic competence is a matter of several years of in-service professional development, especially when all relevant aspects in the complex reality of classroom instruction are in view (Franke et al., 2001, Empson & Jacobs, 2008, p. 267; Sherin & van Es, 2009). Acknowledging this complexity and the long-term scale involved are crucial to starting the development of diagnostic competence asearly as possible in pre-service teacher education and investigating the relevance of providing a categorial framework for pre-service teachers' diagnostic judgments.

Although pre-service education can hardly enhance teachers' noticing of whole-classroom complexity, it can start with written diagnostic judgments of less complex situations, as different intervention studies have shown (e.g., Heinrichs, 2015;Girulat et al., 2013).

Use of videos for deepening diagnostic judgments in pre-service courses.
The importance of video as a tool for pre-service and in-service mathematics teacher education and many different purposes has often been emphasized (Brophy, 2004; Seago, 2000;Sherin, 2004). Using videos has specifically proven effectiveness for developing teachers' noticing in long-term courses for in-service teachers (e.g., Sherin & van Es, 2009; Jungwirth et al., 2001; Scherer & Steinbring, 2006). These studies also provide important criteria for the effective use of video: It is not the video itself, but the systematic and category-led communicative reflection of videos in repeated sessions that leads to deepening or widening teachers' noticed aspects (Sherin & van Es, 2009), changing their general perspectives (Jungwirth et al., 2001; Scherer & Steinbring, 2006; Sherin & van Es, 2009) and increasing self-initiated reference to categories (Busch et al., 2015). In all these studies, the researchers mention more or less explicitly the relevance of focus questions posed by the participants or the teacher educator for shifting the attention.

In initial teacher education programs, videos are also increasingly used for various purposes (Blomberg et al., 2013) such as showing best practices or analysing teachers' own teaching experiments. Videos designed to develop diagnostic competence have often been applied for diagnostic interviews in small tutor groups (e.g., Clarke et al., 2011), but also for enhancing the variety of aspects teachers notice in classroom observations (Star & Strickland, 2008).

In order to initiate the first steps towards diagnostic competence as early as the students' first year, video can also be used in big introductory lectures (in Germany as well as elsewhere, these usually address 100 or more students). Little is known so far about conditions of video use in this specific teaching learning arrangement, which makes a communicative reflection more complicated. From the existing literature, we draw on several conditions that might influence learning from video:

- **video** has a higher authenticity compared to students' written products,
- adequate **focus questions** by which teachers' noticing is partly guided and
- it offers of **systematic categories** that deepen the analysis (Helmke et al., 2003).

Hence, our research interest refers not only to the relation between general perspectives, noticed aspects and activated categories, but also to the possibility of using video to deepen the judgments and focusing the attention by offering categories and focus questions. As diagnostic judgments depend strongly on the situation to be analysed, the next section thoroughly presents a video, the possible aspects to be noticed and the underlying categories.

Anton's video: A priori analysis of the material and its underlying categories and aspects

The word problem and first categories for analysing it.
In order to investigate prospective teachers' diagnostic judgments, we refer to a video that the first author has often used in pre-service and in-service teacher courses. It shows a student's modelling and solution process for the following task (raw data from the master thesis by Renk, 2009; analysed in detail in Prediger, 2010b):

> **Word problem.** 'The African grey parrot can grow up to 40 cm long. A flamingo measures about 2 m, thus 200 cm. How many times is the flamingo bigger than the grey parrot?'

One typical tool for analysing the necessary and observable cognitive activities for these kinds of word problems is the modelling cycle (Pollack, 1979; here by vomHofe et al., 2006). The activity categories of the modelling cycle allow description of the hypothetically necessary cognitive activities for modelling and solving the word problem, although not necessarily chronologically: The starting point is a (more or less authentic) situation from the non-mathematical 'world' (i.e., a situation that is to be mathematised). Once the situation model is structured, it has to be *mathematised* into a mathematical model, here the arithmetic operation '40 x ? = 200' or '200 :40 = ?'. Inner-mathematical *calculating* results in 5, and then this result is to be *interpreted* with respect to the situation, e.g., the flamingo is 5 times bigger than the parrot. *Validating* here means asking, is that a plausible result? If not, the cycle begins again. The core activities written in italics will later serve as categories and be marked as categories, e.g., *<interpreting>*.

This classical cycle can be complemented by the step of structuring the original situation into a situation model; here that means identifying that the question corresponds to the question, 'How often does the parrot fit into the flamingo?' (see Reusser, 1990, and Borromeo Ferri, 2006, for the relevance of the situation model). When specifying which mathematical knowledge is required for this elementary modelling process, the category of *basic mental models* (BM in Figure 1;vomHofe et al., 2006; similarly, Usiskin, 1991) offers topic-specific keys: In order to *<mathematise>* the question 'How often does 40 fit into 200?'using division, the *quotitive model* (the topic-specific instance of basic mental models *BM*) is required, i.e., the knowledge that the question of fit into can always be mathematised using division, and, vice versa, the result of division can be *interpreted* as the factor between both numbers.
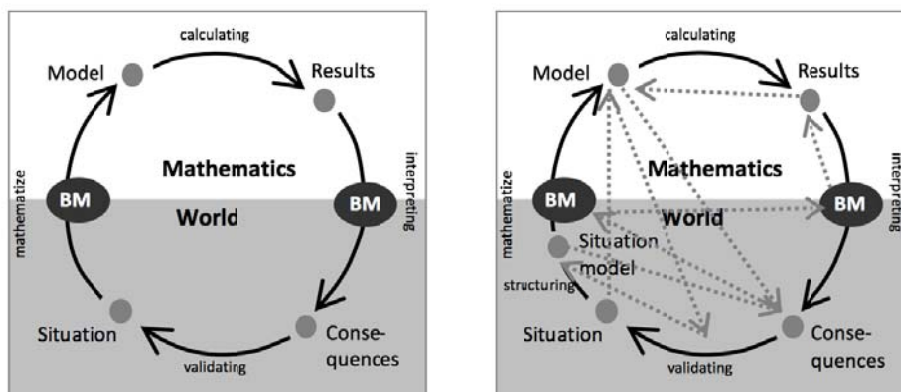
Figure 1. Left: Modelling cycle with basic mental models (BM),
Right: Anton's individual pathway through the modelling cycle

The cognitive activities in the modelling cycles have been often used for analytic purposes in a priori analyses of tasks, but also for a posteriori analyses of students' modelling processes (e.g., Borromeo Ferri, 2006; Galbraith & Stillman, 2006) and for investigating teachers' diagnostic competence (Schwarz et al., 2008). Similarly, the basic mental models serve as important categories for students' difficulties (e.g., vomHofe et al., 2006; Prediger, 2008)and as categories in teachers' diagnostic competence (Davis & Simmt, 2006; Prediger, 2010a).

The case of Anton's video: Reconstructable aspects and further categories.
The above categories were applied to analyse the videotaped multi-facetted modelling and solution process of a10-year-old boy named Anton (A in the transcript), who worked on the parrot-flamingo word problem. The interviewer (I) continuously pushed him to think aloud and selectively helped when the process stopped.

Analysing the video using the five introduced categories (indicated using '<. . .>')allowed the identification of several typical aspects in students' processes, but also required the addition of two further categories (see below).

0A	[*reads aloud the text of the word problem*]
1A	There. . . [*5 sec break*] 40 times 200 . . .
2I	How did you find that?
3A	[*4 sec break*] Oh, no. Or, there I would, or there, or 40 divided by 200.

Having read the word problem, Anton immediately <*mathematises*> without <*structuring*> the situation by combining the two numbers arbitrarily (Lines1 and 3).Aspect A1 (aspects are indicated using'||. . . ||')

   ||A1: Anton mathematises hastily using40 × 200 and 40:200||

reflects a typical pattern in students' modelling processes which has often been found in empirical research: Students combine the numbers in the word problem arbitrarily without paying attention to the choice of operations and without really structuring a situation model (Verschaffel et al., 2000; Prediger, 2010c). Whereas other students stop their process at this point, Anton continues without the interviewer's prompt:

4A	[*without break*] There I have to watch, wait . . .
5I	Yes, of course, just think about it
6 A	[*reflects silently for 16 sec*] Well 160, well 1.60 m it is bigger, the flamingo. That is what I know. I can remember that [*write 160 on his sheet*].
7I	Exactly, write that down.

8 A   And that I have calculated in . . . Well, I have complemented.
9 I   Ehm.
10 A   How do youmake the complementsign?

The reconstructive aspect, A2,in Lines 3 and 4
    ||A2: Anton validates the hasty mathematisation||

is important for the diagnostic judgmentbecause Anton here does not follow the typical pattern of not validating (Verschaffel et al., 2000).Teachers who are acquainted with research results sometimes overlook that not every student fits into the well-known scheme of Aspect A1. Line 6 contains three important aspects to be observed:
    ||A3: Anton searches for distance instead of fit into||
    ||A4: Anton's result 160 is coherent with his situation model||
    ||A5: Anton finds the result before the formal notation||

Aspect A5 is remarkable: Anton finds the result before being able to formalise his way of thinking. This shows that the modelling cycle must not be misunderstood as being a chronological model. As the deeper analysis of the transcript unfolds, this pattern persists. Anton's individual way through the modelling cycle is quite vivid (seeFigure 1 right).

By also considering Line 8, Aspects A4 and A5 are strengthened: Anton has built his individual situation model and structures the situation as if the distance is to be determined, not the quotient. His result 160 is coherent with this additive instead of multiplicative situation model. Again, Anton reproduces a typical pattern of students' modelling processes in which structuring the adequate situation model is a typical obstacle (e.g., Galbraith & Stillman, 2006). Teachers who want to reconstruct students' processes within their inherent logic hence need to apply the additional category *<distinction correctness and match to situation model>*: Anton's result is not correct with respect to the original expectations but matches to his individual situation model. In Line 10('How do you make the complement-sign?'), AspectA5 is strengthened by another observation:
    ||A6: Anton cannot formalise his intuitive completion of 160||

The combination of AspectsA5 and A6 also represents a typical, far-reaching pattern: Informal thinking is more accessible for many students than *<mathematizing>* it formally, an observation made e.g., by van den Heuvel-Panhuizen (2005) and Nunes et al. (1993). Noticing theseaspects is supported by the last category that we introduce:*<distinction between formal model and (informally) finding result>*. In the concrete case, finding the formal model is difficult because Anton does not seem to have the *<basic mental model>* of subtracting as complementing or determining the distance (Selter et al., 2012).

10 A   How do youmake the complement-sign?
11 I   Either you can write it anyway by plus . . .
12 A   Ok, then I can . . . [*writes 40+60+100=160 cm*]then I can 40 plus 60 plus 100 equals 160centimetre
13 I   No, that is not correct [*hints to the written notation*]
14 A   40? plus 60 are one hundred, and then 200—ah, now I have [*starts to cross outthe 160*] now I have . . .there it results in 200 [*writes 200 instead of 160*]
15 A   And then, and then, what do I have calculated? Then I know that the grey parrot [*break 3sec*] it is 40 cm, then I must . . . calculate 200 – 40.
16 I   Yeah, what do you get then?
17 A   Moment [*writes 200 – 40 = 160*], 200 minus 40 equals 160.
18 I   Ehm, and what have you found, now?
19 A   That the grey parrot is 160 cm smaller than the flamingo.
20 I   Okay, and the question was, though, 'How many times is the flamingo bigger than the grey parrot?' How could you find that?

An easy observation in Line 12 is that Anton again chooses a partly wrong notation:
    ||A8: Anton wrongly notes his completion as40+60+100=160||

Line 11 provides a possible reason for Anton's focus which is more difficult to find for the preservice teachers:

||A7: Anton is led to addition by interviewer's hint||.

Another reason seems to be that Anton records his calculation strategy, starting with 40, then adding 60, then adding 100, with the result being that 160 is complemented. In order to reconstruct this aspect, the category *<distinction between formal model and (informally) finding the result>* can be applied again. When the interviewer hints at the mistake in the notation, Anton corrects it to 40+60+100=200, but then searches for another notation:

||A9: Anton corrects notation to 200-160 = 40||.

We speculate that he searches for a new notation due to his *<basic mental model>* of equal signs: Like many students, Anton believes that the equal sign prompts the result rather than symbolises a relation (Kieran, 1981; Prediger, 2010a). After correcting 40+60+100=160 to 40+60+100=200, he is unsatisfied because his result of160 from his completion strategy is no longer visible. Thus he continues searching for a notation emphasizing the 160 and satisfying the mathematical rules of having equal expressions on both sides, which results in 200-160 = 40. However, this interpretative analysis of Anton's unexpressed thoughts is out of the scope of our paper. In Line 19, Anton *<interprets>* and *<validates>*:

|||A10: Anton interprets the result with respect to his situation model||
||A11: Anton validates without reference to the original text||

The original video continues after Line 20 with Anton searching for a new situation model and finally arriving at 40 · 5 = 200 and 200: 40 = 5, although with many further deviations on the way (see Prediger, 2010b). For our research on teachers' diagnostic judgments, we stop the video at this point.

To sum up: The analysis resulted in specifying seven topic-specific categories and 11reconstructive aspects as possibly relevant (see Table 1 below). These categories and the three general perspectives were assumed to be useful for developing a comprehensive diagnostic judgment of Anton's modelling process and his difficulties with his choice of operations. Within the systematic analysis of the transcript from a researchers' point of view, all seven categories were applied to find and formulate the 11reconstructive aspects. It is the core of our empirical study to investigate how prospective teachers analyse Anton's process with respect to these perspectives, categories and noticed aspects, without expecting them to find all aspects.

Refined research questions and hypotheses

On the basisof this a priori analysis, the research questions were specified for the specific case of prospective teachers' noticing Anton's process:

(Q1) What do prospective teachers notice when they write a diagnostic judgment on the video of Anton's thinking process? Which general perspectives become visible, and which categories do they activate for addressing which aspects?

(Q2) To what extent can the diagnostic judgments be enriched by an intervention on the selected categories?

(Q3) To what extent can the diagnostic judgments be enriched by focus questions?

We specifically investigate the following two hypotheses:

(H1) *Relevance of intervention.* The taught categories and aspects grounded in the taught categories are more often addressed in the second judgment than in the first judgment;the untaught categories and aspects basedon them increase less strongly.

(H2) *Relevance of focus questions.* Second judgments are significantly different in perspectives, categories and aspects with or without explicit focus on basic mental models.

Research design and methods

Design for data gathering.
In order to pursue the research questions, we asked 159 prospective mathematics teachers to write diagnostic judgments on the video of Anton's thinking process (to make the situation as real as in a classroom, they received no transcript). In order to test whether prospective teachers' written diagnostic judgments could be enhanced by a short intervention on another example, the data gathering was organised using a pre- and post-intervention design (seeFigure2).
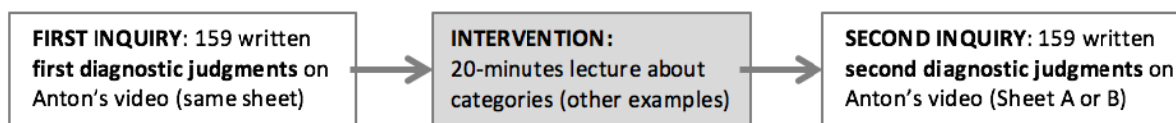

Figure2. Research design: Written diagnostic judgments before and after a short intervention

In the first inquiry, the prospective teachers received a sheet as shown in Figure3. After the first inquiry, a lecture of 20 minutes was given by the first author. With other situations and concrete cases presented using transcripts, Categories C1 and C2 on basic mental models and C3 to C5 of the modelling cycle were introduced, but Categories C6 and C7 were not. No aspects about Anton's process were treated in the intervention so that a transfer from the presented (much less complex) example to Anton was demanded in order to unpack the process and decentre the judgment by means of the offered categories. The second inquiry was conducted immediately after the lecture.

In order to investigate the focus questions'relevance instimulating perspectives, the group was split for the second inquiry. One-half received Sheet A, the other half Sheet B. Both sheets for the second judgments contained figures of the modelling cycle (on Sheet A without basic mental models and on Sheet B with them) and focus questions as shown in Figure 3.

Common sheet for first diagnostic judgment:
Word problem. The African grey parrot can grow up to 40 cm long. A flamingo measures about 2 m, or200 cm. How many times is the flamingo bigger than the grey parrot?
Analyse Anton's process ina manner as detailed as possible. How does he proceed in solving the word problem?

Sheet A for the second diagnostic judgment (without explicitly mentioning basic mental models):
Analyse Anton's process in as detailed a manner as possible by means of the modelling cycle:
• Which cognitive activities does he activate (not necessarily in the order of the cycle)?
• What is plausible in his thinking and a good starting point, even if not yet completely correct?

Sheet B for the second diagnostic judgment (with explicitly asking for basic mental models):
Analyse Anton's process in as detailed a manner as possible by means of the modelling cycle:
• Which cognitive activities does he activate (not necessarily in the order of the cycle)?
• Where does his thinking deviate from the intended ways?
• Which basic mental models does he activate?

Figure3. Sheets for the diagnostic judgments

Sample.

The sample consisted of all participants of two secondary mathematics education courses: One course with 72 prospective grammar school mathematics teachers in their first year of university and one course with 87 prospective middle school mathematics teachers in their third year of university, both with approximately 60% females and 40% males. These groups were selected due to their varying previous experience in mathematics education. Both groups were randomly assigned to Groups A and B, receiving Sheet A (37 from the first-year class and 47 from the third-year class) and Sheet B (35 from the first-year class and 40 from the third-year class).

Methods for data analysis.

The 2 ×159 written diagnostic judgments were analysedusing the coding scheme, with codes for perspectives, categories and aspects (seeTable 1). What teachers noticed might have been more than they explicitly wrote down in the written judgment, which is why the code' addressed aspects' was used rather than 'noticed aspects'. The data analysis proceeded in the following steps:

Step 0: *Deriving a coding scheme* from the theory and the a priori analysis of the material.

Step 1: *Consolidating the coding scheme.* In a team of researchers, the codes for perspectives, categories and aspects found in teachers' diagnostic judgments were specified and operationalised in order to achieve interrater reliability. Two coders were trained, and they coded independently; their incompatible codings led to refining the operationalization of the codes. Whereas interrater agreement was easy to achieve for most aspects, the coding of activated categories had to be refined several times.

Step 2: *Rating with the coding scheme.* A third and fourth rater were trained for coding the critical part again, independently from the discussions. The interrater reliability check resulted in very good values for Cohen's Kappa for the activated categories (average $\kappa = 0.76$) and acceptable values for the noticed aspects (average $\kappa = 0.67$) but unacceptable values for general perspectives (average$\kappa = 0.37$). As a consequence, the codings for general perspectives were discussed again and only partly included in the statistical analysis.

Step 3: *Performing statistical analysis.* The statistical analysis comprised identifying frequencies for each code in each subgroup and for the first and second judgments. For the activated categories and noticed aspects (with good interraterreliability), *F*-tests were conducted to test the significance of the intervention and Chi-squared tests were conducted to prove that some categories and aspects do not appear to be statistically independent.

Coding scheme.

In order to show connections between categories and noticed aspects, the intervention in our study only treated Categories C1 to C5, whereas C6 and C7 were not addressed in the intervention. By omitting these categories, Aspects A5 to A9 were difficult to ground. The distinction between taught categories and untaught categories allowed investigation of the relevance of categories taught in the intervention in order to address selected aspects.

Table 1. General perspectives and diagnostic categories underlying the reconstructive aspects for Anton's process (grey and * marks the categories not taught in the intervention and aspects based on these categories)

General perspectives		Aspects reconstructable for Anton's process	
P1	Surface judgment **vs.** deeper focus on mental aspects	A1	Anton mathematises hastily using 40×200 and $40:200$
P2	Deficit orientation **vs.** resource orientation	A2	Anton validates the hasty mathematisation
P3	Directive listening **vs.** reconstructing inner logic	A3	Anton searches for distance instead of fit into
Categories		A4	Anton's result 160 is coherent with his situation model
C1	<Basic mental model>	A5*	Anton finds result before finding operation
C2	<Quotitive model>	A6*	Anton cannot formalise his intuitive completion of 160
C3	<Mathematizing>	A7*	Anton is led to addition by interviewer's hint
C4	<Interpreting>	A8*	Anton wrongly notates his completion as $40+60+100=160$
C5	<Validating>	A9*	Anton corrects notation to $200-160 = 40$
C6*	<Distinction correctness/match to situation model>	A10	Anton interprets result with respect to his situation model
C7*	<Distinction formal model/finding result>	A11	Anton validates without reference to original text

Results on prospective teachers' diagnostic judgments of a video

Qualitative approach in three cases of prospective teachers' judgments.
In a first qualitative approach, we discuss three cases of prospective teachers' judgments (in Figure 4) to illustrate the functioning of the coding scheme (from Table 1). For each diagnostic judgment, the expressed general perspectives, explicitly addressed aspects and activated categories are coded.

The case of Ann.
Ann's diagnostic judgments show a substantial enhancement on all levels:
- *Perspectives.*Her first judgment was coded as resource oriented (P2) but a surface judgment (¬P1). (Here and in the following, ¬X means that X was not coded.) Her second showed a deeper focus on mental aspects (P1), even though the inner logic was still not reconstructed (¬P3). The development between the first and second judgment was abbreviated as¬P1P2¬P3➔P1P2¬P3. (This denotation means that Ann had perspective P2, but not P1 and P3 in the first judgment, and this turned into P1 and P3 but not P2 in the second judgment.)
- *Categories and aspects.* As one might assume, Annactivated more categories (C3C6➔C3C4C5C6C7) and addressed more aspects (A1A4¬A5A10➔A1A3A5A6A8A9A10). The wrongly judged ¬A5 in the first judgment was correctly addressed in the second judgment. It is interesting to see that the activation of categories such as C6 *<distinction correctness/match to situation model>*led to addressing Aspect A3 ‖Anton searches for distance instead of fit into‖ but not A4 ‖ result 160 is coherent with his situation model‖ since the perspective P2 of reconstructing inner logic was not adopted.

However, Ann's substantial enhancement was not shared by all students, as the other cases show.

Ann's first judgment (GA14-1)		Ann's second judgment (GA14-2)	
	¬ P1 P2 ¬ P3		P1 P2 ¬ P3 C3 A1
• Tries to use the numbers; first: 40 × 200, then 40 : 200 (after teacher's question "How did you find the 40 × 200?") → "dividing" is known	A1	• Mathematising: uses given numbers 40 × 200 "not" 40 : 200 → calculating "There I have to watch" • Calculating: 160 cm bigger → "that is what I know" →"complements"→calculates with the numbers	
• 200 − 40 = 160 respectively 40 + 60 + 100 → he has "complemented"	A4		A5 C4 A6 A8
• Calculates, how much the flamingo is bigger, not the ratio → has wrongly realized the "how many times bigger"	A3, C3	• Interpreting: 40 + 160 + 100 = 160 cm → "ah, it results in 200" → tries to calculate with this result → other way of writing 200 − 40 = 160 result	A9 C7 C5
• Knows subtraction, addition, division and multiplication, cannot assign division to the word problem	¬ A5	• Validating: The grey parrot is 160 cm smaller than the flamingo → related to situation	A10 C6
• He notices that the flamingo is bigger	A10, C6	• → it is correct so far, but wrong interpretation of "how many <u>times</u> bigger" • → ratio is not clear, distance calculated	A3

Ben's first judgment (GB71-1)		Ben's second judgment (GB71-2)	
	P1 P2 P3		P1 ¬ P2 P3
• He immediately starts to calculate, e.g., wants to multiply	A1	Anton does not activate all fitting basic models, e.g., the quotitive model is missing. He argues exclusively about the basic model of complementing.	A1 C1 C2
• Then he realizes that it does not work like that	A2		
• Finally, he solves the task by transforming it into a subtraction problem (which again he transforms into an addition problem by complementing)	A3 A4	Just in the beginning, he even tried the completely inadequate operation of simply multiplying. We can assume that this initial idea arrived accidentally, not based on a basic model.	A3 C1
• That is why he only arrives at a partial solution that is correct and is a good base for answering the original question	A10 ¬ C6		
• For this, knowledge (basic models) of division or multiplication would be useful	C1 C2		

Cathleen's first judgment (GA048-1)		Cathleen's second judgment (GA048-2)	
	P1 P2 P3		P1 P2 P3
1) 40 × 200 Or 2) 40 : 200 → long reflection 3) 160 cm is Flamingo bigger 4) 40 + 60 + 100 = ~~160~~ 200 5) 200 − 40 = 160 → grey parrot is 160 cm smaller than the flamingo		(1) His solving process starts with the situation (2) His mathematization process is problematic in his first two trials (× and :)	A1 C3 A2 A3
Anton first starts without plan and tries to solve the task (possibly without having understood the question) by the standard operations × and : .	A1	(3) In the third trial, he discovers that the difference is 160 cm, i.e., he goes to a result without searching a mathematical model, but does not know how to formulate the calculation (formally)	
After longer reflection and repeated retracing of the situation, he discovers correctly that the flamingo is 160 cm bigger than the grey parrot. But he cannot formally formulate the discovery.	A2 A5, A6 A9, C7 A8	(4) Then he starts with his result (160 cm) and tries to mathematize it, i.e., to show how he found it.	A5 A6 C7
In the following, he tries to note the 160cm difference formally.		(5) 40 + 60 + 100 = 160 is his first model, which fits to his result (for him) → teacher draws attention to the problem	A4 A8
Only in the fifth step (see above), he gets the insight that the grey parrot is 160 cm smaller than the flamingo.	A10 A11 C3 ¬ C6	(6) → 40 + 60 + 100 = ~~160~~ 200 Anton calculates / processes it correctly does not fit result	A9
However, he has not solved the task with respect to the question, but has arrived at an intermediate state in the solving process of the task.		(7) Then, he constructs a new model: 200 − 40 = 160 which fits to his result (8) He interprets his result and draws the consequence that the grey parrot is 160 cm smaller than the flamingo. (9) → But the validation with respect to the task does not take place.	A10 C4 C6 A11 C5

Figure4. Three case studies of changing diagnostic judgments and their codings (¬ C6 symbolises the inadequate activation of C6)

The case of Ben.
- *Categories and aspects.* Ben addressed many more aspects in his first judgment than in his second (A1A2A3A4A10→A1A3) and activated only two categories in the second judgment

(C1C2¬C6→C1C2) while elaborating his text. This can be traced back to the focus question on his Sheet B that primes the focus on basic mental models C1 and C2.

- *Perspectives.* In Ben's case, this focus seemed to narrow his perspective, excluding other aspects and resulting in more restrictive perspectives (P2→¬ P2).

The decrease in addressed aspects also appeared because many teachers (including Ann) primarily wrote additional aspects 30 minutes later in the second judgment as they were considering their two texts together and wanted to avoid repetition (according to their oral utterances after the data gathering). Taking this into consideration, we will also report on cumulatively addressed aspects in the first and second judgment below (Coding 1, if first *or* second judgment is coded by 1).

The case of Cathleen.
Cathleen startedon a very high level:

- *Perspectives.* She showed the intended general perspectives from the beginning (P1P2P3→P1P2P3).
- *Categories and aspects.* She addressed many aspects in the first judgment and included the Aspects A3A4 in the second (A1A2A5A6A8A9A10A11→A1A2A3A4A5A6A8A9A10A11) as well as the taught Categories C3-C5 (C3C7¬C6→C3C4C5C6C7).

As Cathleen was able to use the modelling cycle flexibly, she addressed many more categories and aspects than most of the teachers. However, it is worth investigating what effects appeared for others after teaching Categories C1-C5.

The cases of Ann and Cathleen show how the introduction of categories can lead to an extension of addressed aspects, but also to an increasing precision of addressed aspects. Both were able to describe A1 ‖Anton mathematises hastily by 40×200 and $40:200$‖ more precisely when getting to know C3 *<mathematising>*. This subtle difference is not visible in the quantitative analysis. However, as the analyses of three cases already reveal a huge variety in the individual judgments, the quantitative analysis of all 2×159 diagnostic judgmentsis necessary to find more generalizable pattern.

Quantitative analysis of perspectives, categories and aspects addressed in the first diagnostic judgments.
This section widens the analysis of judgments to all 159 prospective teachers. In order to investigate all uttered general perspectives, categories and addressed aspects (Research Question Q1), Table 2 shows the frequencies in all 2×159 diagnostic judgments. This section discusses the columns on the first judgments and the next section investigates the change between first and second judgments.

- *General perspectives.* Whereas other studies have shown the existence of problematic perspectives in many prospective teachers' thinking, more than half of the prospective teachers already showed fruitful *general perspectives* in the first inquiry. Sixty-four point eight percent adopted a deeper focus on mental aspects (P1) more frequently than a surface judgment and 52.8% evaluated in a more resource-oriented (P2) than deficit-oriented way. Fifty point three percent even made more of an attempt to reconstruct the inner logic of Anton's thoughts (P3) than they did directive listening. Even without comparing to a control group, the existing literature on the use of video permits us to speculate that this majority of fruitful perspectives might have been elicited by showing a video instead of a written product, because video invites consideration of the process rather than solely products of the process.

Table 2. Perspectives, categories and aspects in the first and second judgment (arrows indicate increase or decrease)

		First diagnostic judgment		Second diagnostic judgment		
General diagnostic perspective I vs. II		Perspective I	Perspective II	Perspective I	Perspective II	Direction of change
P1	Surface judgment **vs.** deeper focus on mental aspects	35.2%	64.8%	20.8%	79.2%	↗
P2	Deficit orientation **vs.** resource orientation	47.2%	52.8%	45.3%	54.7%	→
P3	Directive listening **vs.** reconstructing inner logic	49.7%	50.3%	52.8%	47.2%	→
Categories activated explicitly or implicitly		Activated in-adequately	Activated (partly) adequately	Activated in-adequately	Activated (partly) adequately	Direction of change in category activation
C1	\<Basic mental model\>	0.0%	5.7%	1.3%	33.3%	↗
C2	\<Quotitive model\>	0.0%	3.8%	0.6%	26.4%	↗
C3	\<Mathematizing\>	0.0%	37.7%	0.0%	83.0%	↗
C4	\<Interpreting\>	0.6%	9.4%	0.0%	62.9%	↗
C5	\<Validating\>	0.0%	4.4%	0.0%	48.4%	↗
C6	\<Distinction correctness/match to situation model\>	27.7%	57.9%	32.1%	33.3%	↘
C7	\<Distinction formal model/finding result\>	30.8%	5.7%	15.1%	1.9%	↘
Aspects explicitly addressed in the teachers' judgment		Addressed in-adequately	Addressed adequately	Addressed in-adequately	Addressed adequately	Cumulatively addressed adequately in 1st or 2nd judgment
A1	Anton mathematises hastily by 40× 200 and 40:200	0.6 %	74.2%	0.0%	65.4%	↗ 86.8%
A2	Anton validates the hasty mathematisation	0.0 %	35.2%	0.0%	30.2%	↗ 49.7%
A3	Anton searches for distance instead of fit into	2.5%	61.0%	5.7%	55.3%	↗ 71.1%
A4	Anton's result 160 is coherent with his situation model	1.9%	39.0%	0.6%	23.9%	→ 46.5%
A5	Anton finds the result before the formal notation	11.3%	37.1%	11.3%	24.5%	→ 44.0%
A6	Anton cannot formalise his intuitive completion of 160	4.4 %	23.3%	5.7%	5.7%	→ 23.3%
A7	Anton is led to addition by interviewer's hint	0.6 %	5.0%	0.0%	1.3%	→ 5.7%
A8	Anton wrongly notes his completion by 40+60+100=160	20.8 %	20.8%	11.3%	8.2%	→ 23.3%
A9	Anton corrects notation into 200-160 = 40	19.5 %	16.4%	10.7%	6.9%	→ 18.2%
A10	Anton interprets result with respect to his situation model	0.6%	26.4%	3.8%	54.1%	↗ 60.4%
A11	Anton validates without reference to original text	0.0 %	1.9%	5.0%	27.7%	↗ 27.7%

- *Categories activated explicitly or implicitly.* As expected, the prospective teachers only rarely activated didactic categories not yet learnt (C1, C2, C4 andC5: between 3.8% and 9.4%). Only Category C6, the *\<distinction between correctness/match to situation model\>*was (mostly implicitly) already activated by 57.9% of the sample and Category C3 *\<mathematizing\>* by 37.7%.
- *Aspects explicitly addressed in the teachers' judgment.* Even without applying categories, the prospective teachers were able to notice important aspects in the video. As they sawthe video only one time and did not have access to a transcript, it is obvious that they could not address all 11 aspects found in researchers' thorough analysis. On average, the teachers' addressed 3.40 aspects adequately in the first judgment and 0.62 inadequately. The most frequent were AspectsA1 (addressed by 74.2% of the sample) and A3 (61.0%). In contrast, the teachers rarely addressed the role of the interviewer in the child-adult interaction (A7 with 5.0%) and the fact that Anton's validation did not go back to the text (A11 with 1.9%).

Statistical connections.
In order to find statistical connections between activated categories and addressed aspects, the first approach was to conduct chi-square tests for all combinations of aspects and categories in the first

judgments. Only four out of 77 combinations showed a significant statistical connection (on a significance level of $p<0.01$):

- Category C3 <*mathematizing*> with Aspect A2 ($\chi^2 = 9.23$),
- Category C4<*interpreting*> with Aspect A2($\chi^2 = 7.18$),
- Category C4 < *interpreting*> with Aspect A10 ($\chi^2_2 = 9.61$) and
- Category C6 <*correct/match*> with Aspect A3 ($\chi^2_2 = 10.57$).

The connections C3/A2, C4/A10 andC6/A3 were expected since these aspects are directly grounded on the respective categories. In contrast, the fourth statistical connection was more unexpected: Using Category C4 <*interpreting*> and noticing Aspect A2 ‖Anton validates the hasty mathematisation‖ is not immediately linked (although however plausible).The fact that all other aspects and categories do *not* provide a statistically significant connection (not even on a level of I <0.05) confirms that the coding scheme grasps different phenomena.

For this reason, the connections must be searched in a more dynamic approach by investigating the changes after the intervention.

Changes in the diagnostic judgments before and after a short intervention.
Table 2 also contains data for pursuing Research Question Q2 on changes from first to second judgments after an intervention in which the Categories C1 to C5 were introduced in a 20-minute lecture.

- *Change of general perspectives.* The change is not too high for the *general perspectives* since General PerspectivesP2 and P3 did not increase for the complete sample (the next section discusses the contribution of the focus questions to this effect). In contrast, P1 changed substantially: the intervention resulted in stimulating teachers to adopt a deeper focus on mental aspects: 64.8% in the first judgment and 79.2% in the second judgment.
- *Increase of activated categories.* As expected, the highest increase can be found for activating categories. The taught Categories C1 to C5 were activated to a much higher degree in the second judgment (23% to 53% more students activating them). In contrast, Categories C6 and C7 (being neither taught nor mentioned in the focus questions) were used less successfully: For C6, the frequency of adequate activation even decreases from 57.9% in the first judgment to 33.3% in the second judgment. This confirms the phenomenon reconstructed for Ben, that focusing on some categories might imply the neglect of others. The Category C7 <*distinction between formal model and informally finding the result*>was activated less often, but in fewer cases in an inadequate way. Here, the introduction of the modelling cycle might support a transfer to the distinction of formal/informal.
- *Aspects explicitly addressed.* Whereas the frequency of addressing categories increases enormously, most teachers address even fewer aspects in their second judgments, on average 3.03adequately addressed aspects compared to 3.40 in the first judgment. This might be interpreted as hint that a more professional way of noticing might also include a more selective focus.

As several teachers considered their texts as cumulative and mentioned only new aspects, we also counted aspects cumulatively when addressed adequately in the first or the second judgment (and not inadequately in the second judgment). This cumulative average number of adequately aspects amounts to 4.57.Thisnumber shows that many teachers extended their repertory of addressed aspects to a remarkable degree. Within these cumulatively addressed aspects, the priorities changed from first to second judgment. The biggest relative increases were 34%for Aspect A10 ‖Anton interprets result with respect to his situation model‖ and 25.8% for A11 ‖Anton validates without reference to original text‖. This might reflect the relevance of the categories <*interpreting*> and <*validating*> for these aspects. Analogically, the increase of 12.6% for A1 ‖Anton mathematises hastily‖ and 14.5% for A2 ‖Anton validates the hasty mathematisation‖ might be explained by the increase in the category

<mathematizing>. The lowest increases for the Aspects A6 to A9 (between 0% and 2.5%) might be traced back to the fact that the intervention did not offer categories supporting the noticing of these aspects.

Hypothesis H1. For testing the Hypothesis H1, the categories and aspects were grouped into taught/non-taught in Table 3 (as specified in Table 1). An *F*-Test showed significant differences in the increase of taught categories compared to non-taught categories (with F[1, 316] = 157.76, *p*<0.001; η^2 =.67) and for aspects grounded on taught categories compared aspects grounded on non-taught categories (with F[1, 316] = 89.876, *p*<0.001; η^2 =.534).

This shows that unlike Ann (who also addresses untaught categories in her second judgment; see Figure 4), most prospective teachers addressed significantly more aspects and categories for which the short intervention provided a base than those aspects and categories not mentioned in the intervention. This shows that even a 20-minute intervention can have visible effects by sharpening the attention and stimulating what teachers notice and write in their judgments. The result applies for the categories and for the aspects, although aspects of Anton's modelling process were not mentioned in the intervention, hence a first transfer could be found. However, the results must be interpreted with caution because applying categories heard in a 20-minute lecture cannot yet be interpreted as indicator for a learning process in the long run, it might also be simply a priming for the short-term focus of attention or an offer of language (Walkoe, 2015).

Table 3. Significant differences in the changes for taught categories and aspects grounded on taught categories

Aspects explicitly addressed in the teachers' judgment		Mean (&SD) in 1st judgement	Mean (&SD) in 2nd judgement (for aspects cumulatively)	Results of the F-Tests confirming differences
C1-C5	Number of taught categories being addressed	0.6 (0.76)	2.5 (1.34)	F[1, 316] = 157.76, p < 0.001; η^2 =0.668
C6-C7	Number of non-taught categories being addressed	0.6 (0.56)	0.4 (0.49)	
	Number of			
A1-4,10-11	. . .addressed aspects grounded on taught categories	2.4 (1.39)	3.4 (1.58)	F[1, 316] = 89.88, p <0.001; η^2 =0.534
A5-A9	. . .addressed aspects grounded on non-taught categories	1.03 (1.39)	1.1 (1.36)	

Relevance of focus questions.
The possible extent of focusing attention rather than real (longer-term) acquisition of categories can also be estimated by investigating Research Question Q3 on the effects of focus questions. Two sub samples of teachers received sheets with partly a different focus question in the second inquiry: Sheet A contained a focus question on resources and inner logic, Sheet B on basic mental models and deficits. The data in Table 4 allows estimation of possible effects of focus questions on the second diagnostic judgments.

As shownin Table 4, the changes of perspectives are tightly connected to the different focus questions. The sub sample with Sheet A with the focus question on resources and inner logic shows a significant increase in P2 (from 56.0% to 71.4%) and P3 (from 47.6% to 52.4%). In contrast, the sub sample with Sheet B with focus questions on deficits shows a significant decrease of resource orientation P2 (from 49.3% to 36.0%) and focus on reconstructing inner logic P3 (from 53.3% to 41.3%). In both sub samples, the deeper focus on mental aspect increases (P1 from ~65% to ~78%), triggered by the intervention or the common focus question on the modelling activities.

Although both sub samples were informed about basic mental models, only 25% of the subs ample with Sheet A activated this category. In contrast, 45.4% of the sub sample with Sheet B, who were explicitly prompted to reconstruct the basic mental models, activated this category (42.7% adequately

and 2.7% inadequately). This shows that asking for a specific category can be an effective prompt that enhances its activation, but it should not be forced; over 50% do not activate the category in spite of being asked to. Focusing on some aspects can imply the neglect other aspects, as the higher rate of addressing A10 in the sub-sample with Sheet A shows. For other aspects, the difference is not relevant.

Table 4. Relevance of focus questions: Elements with differences between second judgments with and without focus question (bold figures mark interesting differences; arrows indicate increase or decrease)

	Subsample with Sheet A (2nd judgment with focus question on resources, not on mental models)			Sub sample with Sheet B (2nd judgment with focus question on mental models & deficits)		
General perspectives	Persp. II in 1st judgment	→	Persp. II in 2nd judgment	Perspective II in 1st judgment	→	Perspective II in 2nd judgment
P1 Surface judgment **vs.** deeper focus on mental aspects	65.5%	↗	79.8%	64.0%	↗	78.7%
P2 Deficit orientation vs. resource orientation	56.0%	↗	71.4%	49.3%	↘	36.0%
P3 Directive listening vs. reconstructing inner logic	47.6%	↗	52.4%	53.3%	↘	41.3%
Categories activated explicitly or implicitly	Activated (partly) adequately in 1st judgment	Activated (partly) adequately in 2nd judgment	Activated **in-adequately in 2nd judgment**	Activated (partly) adequately in 1st judgment	Activated (partly) adequately in 2nd judgment	Activated **in-adequately in 2nd judgment**
C1 <Basic mental model>	6.0%↗	25.0%	0.0%	5.3%↗	42.7%	2.7%
C2 <Quotitive model>	2.4%↗	20.2%	0.0%	5.3%↗	33.3%	1.3%
C3 <Mathematizing>	47.6%↗	89.3%	0.0%	26.7%↗	76.0%	0.0%
C4 <Interpreting>	14.3%↗	72.6%	0.0%	4.0%↗	52.0%	0.0%
C5 <Validating>	8.3%↗	57.1%	0.0%	0.0%↗	38.7%	0.0%
C6 <Distinction correctness/match to situation model>	65.5%↘	40.5%	31.0%	49.3%↘	25.3%	33.3%
C7 <Distinction formal model/finding result>	9.5%↘	1.2%	16.7%	1.3%→	2.7%	13.3%
Aspects explicitly addressed in the teachers' judgment	Addressed (partly) adequately in 1st judgment	Addressed (partly) adequately in 2nd judgment	Addressed (partly) adequately in 1st or 2nd judgment	Addressed (partly) adequately in 1st judgment	Addressed (partly) adequately in 2nd judgment	Addressed (partly) adequately in 1st or 2nd judgment
Anton . . .						
A1 . . . mathematises hastily by 40× 200 and 40:200	84.5%	72.6%	↗96.4%	62.7%	57.3%	↗76.0%
A2 . . . validates the hasty mathematisation	42.9%	40.5%	↗59.5%	26.7%	18.7%	↗38.7%
A3 . . .searches for distance instead of fit into	64.3%	63.1%	↗77.4%	57.3%	46.7%	→64.0%
A4 . . . produces result coherent to situation model	44.0%	23.8%	→50.0%	33.3%	24.0%	↗42.7%
A5 . . .finds the result before the formal notation	44.0%	27.4%	→46.4%	29.3%	21.3%	↗41.3%
A6 . . .cannot formalise his intuitive completion of 160	23.8%	10.7%	→27.4%	22.7%	0.0%	→18.7%
A7 . . .is led to addition by interviewer's hint	7.1%	2.4%	→8.3%	2.7%	0.0%	→2.7%
A8 . . .wrongly notes his completion by 40+60+100=160	21.4%	10.7%	→ 23.8%	20.0%	5.3%	→22.7%
A9 . . .corrects notation into 200-160 = 40	16.7%	10.7%	→19.0%	16.0%	2.7%	→17.3%
A10 . . .interprets result with respect to his sit. model	25.0%	60.7%	↗ 66.7%	28.0%	46.7%	↗53.3%
A11 . . .validates without reference to original text	2.4%	29.8%	↗29.8%	1.3%	25.3%	↗25.3%

Hypothesis H2. The findings confirm Hypothesis H2 that a focus question can trigger the activation of categories and the perspectives expressed in a diagnostic judgment. Interestingly, however, the different activations of categories did not have a relevant effect on the addressed aspects, as they have not changed accordingly. (Because it is difficult to compare values for perspectives, categories and aspects, formal tests were omitted here.)

Discussion and consequences for the use of videos for enhancing teachers' noticing in initial mathematics teacher education

In this paper, we reported on an investigation of 318 written diagnostic judgments on a video of a student's modelling process. The deep-reaching coding scheme allowed us to study the general diagnostic perspectives, activated didactic categories and aspects explicitly addressed by the 159 prospective teachers before and after a short intervention that introduced some of the categories. The study of the statistical connections between perspectives, categories and aspects, the comparison before and after the intervention and with or without certain focus questions allowed generalisation or rejection of patterns found in three brief case studies about the functioning of teachers' diagnostic judgments and the ways in which we can have an impact on them.

Of course, all findings should be discussed with respect to the methodological limitations of the study. This data gathering method allows only a snapshot at a certain moment in the teachers' education. The written diagnostic judgments never reflect everything teachers are able to notice, but only what they address. Additionally, the coding procedure of written diagnostic judgments always leads to simplifications that must be made in order to guarantee interrater reliability without losing validity. For the general diagnostic perspectives, because interraterreliability was not satisfactory, it had to be excluded from some statistical analysis. Furthermore, an intervention consisting of a 20-minute lecture with two inquiries directly before and after this lecture can only give a very first insight into the design possibilities of using video to enhance teachers' noticing, and no claim of validity for teachers' longer-term learning processes can be made. Future research should deepen the results by means of longer interventions, refined coding for perspectives and a scope that is widened by varying the videos in use. However, this study has provided some interesting preliminary empirical findings:

(Q1) The analysis of the diagnostic judgments unpacks a huge variety of perspectives, categories and aspects that prospective teachers can address. This variety itself is an important reason for treating diagnostic issues in teacher education. The multi-facetted insights gained by the very detailed unpacking of a large number of diagnostic judgments strengthen the claims of those researchers who plead for deep-level analysis rather than directive evaluations of teachers' diagnostic judgments (Empson & Jacobs, 2008), but also those who plead for the consideration of something other than only small-scale case studies (Brunner et al., 2013). The specific strength of this empirical study is that it combines deep analysis with a large number of judgments (318). The statistical connection between categories and aspects is not trivially given by the coding scheme, which became evident by the small number of significant chi-squared statistics and also has methodological consequences for further research, as it is crucial to capture both separately.

(Q2) Diagnostic judgments can be enriched by an intervention on selected categories, as they can offer a language and focus the attention to mental aspects not formerly noticed (Walkoe, 2014; Prediger, 2010a). Specifically, the analysis allows confirmation of Hypothesis H1 that the increase for taught categories and aspects basedon taught categories is significantly higher than that for untaught categories and aspects basedon them.This strengthens the plea for using not only videos but also combining the videos with the systematic offer of didactic categories referring to concrete (other) cases (Star & Strickland, 2008; Prediger, 2010a; Walkoe, 2014). Additionally, the qualitative analysis shows how aspects are elaborated by activation of categories, which goes beyond priming preliminary knowledge, as the categories were new for the prospective teachers and really extended their repertoire of explicitly addressed aspects, even with a short-term intervention of only 20 minutes.

(Q3) Both the intervention and the different focus questions have proven to be a strong influence on teachers' diagnostic judgments. However, Hypothesis H2 could only be confirmed for general

perspectives and categories, whereas the effect on the explicitly addressed aspects was only relevant for one aspect. Hence, cueing is possible, but not forceable for extending the repertory of noticed aspects. This again confirms the main assumption of this article that enhancing the quality of diagnostic judgments should mainly by aimed at by offering adequate topic-specific diagnostic categories.

From these main findings and some further side effects, we can draw conclusions for the use of concrete cases in transcripts or videos in initial mathematics teacher education, especially when the courses must be conducted with large group of more than 75, sometimes 300 participants.

- The use of video seems to have triggered very student-centred general perspectives from the beginning; the first-year teachers in particular immediately reached a quite impressive level. This confirms the strength of videos as being very authentic and bringing even first-year teachers near to the children (Brophy, 2004).

- When watching a video for the first time that theoretically contains11 aspects, the prospective teachers explicitly addressed 3.4 aspects on average, hence, missing on average 7.6 aspects. Although of course it could not be expected that 11 aspects would be noticed by novices, work with additional transcripts might increase the number of addressed aspects from the beginning, as it allows students to overcome the problem of speed.

- Focus questions seem to be crucial, especially for the general perspectives adopted by prospective teachers. Therefore, every teacher educator should very carefully prepare adequate focus questions and judge diagnostic outcomes with regard to these questions. However, these cannot guarantee long-term effects, as they mainly prime already existing intuitions.

- Of course, the finding that the intervention could significantly enhance the activated categories (and connected aspects) must be interpreted cautiously, because the intervention was brief and other factors cannot be excluded. However, even tentative indications should already have curricular implications (as claimed by Morris et al., 2009; Heinrichs, 2015): If we want to enhance prospective teachers' diagnostic competence, we should carefully specify all relevant categories that should be acquired and introduce them with carefully chosen cases. Of course, an intervention of 20 minutes is only a beginning, and the post-test directly after it can only examine the short-term effects that have occurred so far, but our (not systematically researched) practical experiences underline the optimism that if a course continues to work with these categories, adequate use by nearly all prospective teachers can be reached. Although details in the interplay of noticed aspects and activated categories must be further investigated, the results are already encouraging for the curricular pathway we took.

Acknowledgements

References

Arcavi, A., and Schoenfeld, A. (2008). Using the unfamiliar to problematize the familiar: The case of mathematics teacher in-service education. *Canadian Journal of Science, Mathematics and Technology Education, 8*(3), 280-295.doi 10.1080/14926150802315122.

Ball, D. L., and Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In G. Sykes, and L. Darling-Hammond (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3-32). San Francisco: Jossey-Bass.

Blomberg, G., Renkl, A., Sherin, M. G., Borko, H., and Seidel, T. (2013). Five research-based heuristics for using video in pre-service teacher education. *Journal for Education Research Online, 5*(1), 90-114.

Borromeo Ferri, R. (2006). Theoretical and empirical differentiations of phases in the modelling process. *ZDM, 38*(2), 86-95.doi 10.1007/BF02655883.

Brophy, J. (Ed.)(2004). *Using video in teacher education*. Amsterdam: Elsevier.

Brunner, M., Anders, Y., Hachfeld, A., and Krauss, S. (2013). The Diagnostic Skills of Mathematics Teachers. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, and M. Neubrand (Eds.), *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers* (pp. 229-248). New York: Springer.

Busch, J., Barzel, B., and Leuders, T. (2015). Promoting secondary teachers' diagnostic competence with respect to functions: development of a scalable unit in Continuous Professional Development. *ZDM Mathematics Education, 47*(1), 53–64.doi10.1007/s11858-014-0647-2.

Clarke, D., Clarke, B., and Roche, A. (2011). Building teachers' expertise in understanding, assessing and developing children's mathematical thinking: the power of task-based, one-to-one interviews. *ZDM Mathematics Education, 43*(6), 901-913.doi 10.1007/s11858-011-0345-2.

Davis, B., andSimmt, E. (2006). Mathematics-for-teaching: an ongoing investigation of the mathematics that teachers (need to) know. *Educational Studies in Mathematics, 61*(3), 293–319. doi 10.1007/s10649-006-2372-4.

Dreher, A., and Kuntze, S. (2015). Teachers' professional knowledge and noticing: The case of multiple representations in the mathematics classroom. *Educational Studies in Mathematics, 88*(1), 89-114. doi 10.1007/s10649-014-9577-8.

Empson, S. B., and Jacobs, V. J. (2008). Learning to Listen to Children's Mathematics. In T. Wood, and P. Sullivan (Eds.), *International handbook of mathematics teacher education* (Vol. 1, pp. 257-281). Rotterdam: Sense.

Even, R., and Tirosh, D. (2002). Teacher knowledge and understanding of students' mathematical learning. In L. English et al. (Eds.). *Handbook of International Research in Mathematics Education* (pp.219–240). Mahwah, NJ: Lawrence Erlbaum.

Franke, M. L., Carpenter, T. P., Levi, L., and Fennema, E. (2001). Capturing Teachers' Generative Change: A Follow-Up Study of Professional Development in Mathematics. *American Educational Research Journal, 38*(3), 653-689. doi: 10.3102/00028312038003653.

Galbraith, P., andStillman, G. (2006). A framework for identifying student blockages during transitions in the modelling process. *ZDM, 38*(2), 143-162.doi 10.1007/BF02655886.

Girulat, A., Nührenbörger, M., andWember, F. (2013). Fachdidaktischfundierte Reflexion von Diagnose und individuelle Förderungim Unterrichtskontext – am Beispiel des Faches Mathematikunter Beachtungsonderpädagogischer Förderung. In S. Hußmann, and C. Selter (Eds.), *Diagnose und individuelle Förderung in der MINT-Lehrerbildung* (pp. 150-166). Münster: Waxmann.

Heinrichs, H. (2015). *Diagnostische Kompetenz von Mathematik-Lehramtsstudierenden. Messung und Förderung*. Wiesbaden: Springer Spektrum.

Helmke, A., Hosenfeld, I., and Schrader, F.-W. (2003). Diagnosekompetenz in Ausbildung und Berufentwickeln. *Karlsruher Pädagogische Beiträge, 55*, 15-34.

Jungwirth, H., Steinbring, H., Voigt, J., andWollring, B. (2001). Interpretative classroom research in teacher education. In H.-G. Weigand et al. (eds.). *Developments in mathematics education in Germany. Selected Papers from the Annual Conference on Didactics of Mathematics.* (pp. 46–56). Hildesheim: Franzbecker.

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics, 12*(3), 317–326.doi10.1007/BF00311062.

Krauss, S., and Brunner, M. (2011). Schnelles Beurteilen von Schülerantworten: Ein Reaktionszeittes tfür Mathematiklehrer/innen. *Journal für Mathematik-Didaktik, 32*(2), 233-251.doi 10.1007/s13138-011-0029-z.

Morris, A. K., Hiebert, J., and Spitzer, S. M. (2009). Mathematical knowledge for teaching in planning and evaluating instruction: What can pre-service teachers learn? *Journal for Research in Mathematics Education, 40*(5), 491-529.doi www.jstor.org/stable/40539354.

Nunes, T., Schliemann, A.D., and Carraher, D.W. (1993). *Street mathematics and school mathematics*. Cambridge: Cambridge University Press.

Philipp, K., and Leuders. T. (2014). Diagnostic competences of mathematics teachers – processes and resources. In P. Liljedahl, S. Oesterle, C. Nicol, and D. Allan (Eds.), *Proceedings of the Joint Meeting of PME 38 and PME-NA 36* (Vol. 4, pp. 426-433). Vancouver: PME.

Pollak, H.O. (1979). The interaction between mathematics and other school subjects. In UNESCO (Eds.), *New trends in mathematics teaching IV* (pp. 232-248), Paris: UNESCO.

Prediger, S. (2008). The relevance of didactic categories for analysing obstacles in conceptual change: Revisiting the case of multiplication of fractions. *Learning and Instruction, 18*(1), 3-17.doi10.1016/j.learninstruc.2006.08.001.

Prediger, S. (2010a). How to Develop Mathematics for Teaching and for Understanding. The Case of Meanings of the Equal Sign. *Journal of Mathematics Teacher Education, 13*(1), 73-93. doi 10.1007/s10857-009-9119-y.

Prediger, S. (2010b). „Aber wie sag ichesmathematisch?" – Empirische Befunde und Konsequenzenzum Lernen von Mathematikals Mittelzur Beschreibung von Welt. In D. Höttecke (Ed.), *EntwicklungnaturwissenschaftlichenDenkenszwischenPhänomen und Systematik* (pp. 6-20). Berlin: LIT-Verlag.

Prediger, S. (2010c). Über das Verhältnis von Theorien und wissenschaftlichenPraktiken – am Beispiel von SchwierigkeitenmitTextaufgaben. *Journal fürMathematik-Didaktik, 31*(2), 167-195.doi10.1007/s13138-010-0011-1.

Renk, N. (2009). *Operationsverständnis von GrundschulkindernbeimMathematisieren von Textaufgaben. Eineempirische Untersuchungzu Strategien des intermodalen Transfers*, Master thesis, supervised by S. Prediger, TU Dortmund.

Reusser K. (1990). From Text to Situation to Equation: Cognitive Simulation of Understanding and Solving Mathematical Word Problems. In H. Mandl, E. De Corte, N. Bennet, and H.F. Friedrich (Eds.), *Learning and Instruction. European Research in an International Context* (Vol. II, pp. 477–498). New York: Pergamon.

Scherer, P., andSteinbring, H. (2006). Noticing children's learning processes – teachers jointly reflect on their own classroom interaction for improving mathematics teaching. *Journal of Mathematics Teacher Education, 9*(2), 157–185. doi 10.1007/s10857-006-0004-7.

Schoenfeld, A. H. (2011). Noticing matters. A lot. Now what? In M. G. Sherin, V. R. Jacobs, and R. A. Philipp (Eds.), *Mathematics teacher noticing* (pp. 223–238). New York: Routledge.

Schrader, F.-W., and Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *EmpirischePädagogik, 1*(1), 27–52.

Schwarz, B., Wissmach, B., and Kaiser, G. (2008). "Last curves not quite correct": diagnostic competences of future teachers with regard to modelling and graphical representations. *ZDM, 40*(5), 777-790.doi 10.1007/s11858-008-0158-0.

Seago, N. M. (2000). Using video of classroom practice as a tool to study and improve teaching. In E. Silver (Ed.), *Mathematics education in the middle grades.* (pp. 63–74). Washington, DC: National Academy Press.

Selter, C. (2001). Understanding - The underlying goal of teacher education. In M. van den Heuvel-Panhuizen (Ed.),*Proceedings of the 25th annual conference of PME*(Vol. 1, pp. 198–202).Utrecht: Freudenthal Institute.

Selter, C., Prediger, S., Nührenbörger, M., and Hußmann, S. (2012). Taking away and determining the difference—a longitudinal perspective on two models of subtraction and the inverse relation to addition. *Educational Studies in Mathematics, 79*(3), 389-408. doi 10.1007/s10649-011-9305-6.

Sherin, M. G. (2004). New perspectives on the role of video in teacher education. In J. Brophy (Ed.), *Using video in teacher education* (pp. 1–28). Oxford: Elsevier.

Sherin, M. G., and van Es, E. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education, 60,* 20–37. doi: 10.1177/0022487108328155.

Sherin, M. G., Jacobs, V.R., and Philipp, R. A. (Eds.) (2011). *Mathematics Teacher Noticing*. New York: Routledge.

Smith, J. P., diSessa, A. A., and Roschelle, J. (1993). Misconceptions Reconceived: A Constructivist Analysis of Knowledge in Transition. *Journal of the Learning Sciences, 3*(2), 115-163. doi10.1207/s15327809jls0302_1.

Star, J. R., and Strickland, S. K. (2008). Learning to observe: using video to improve preservice mathematics teachers' ability to notice. *Journal of Mathematics Teacher Education, 11*(2), 107-125. doi 10.1007/s10857-007-9063-7.

Usiskin, Z. (1991). Building mathematics curricula with applications and modelling. In M. Niss, W. Blum, and I. Huntley (Eds.), *Teaching of mathematical modelling and applications* (pp. 30-45). Chichester: Horwood.

van den Heuvel-Panhuizen, M. (2005). The role of contexts in assessment problems in mathematics. *For the Learning of Mathematics, 25*(2), 2-9.

Verschaffel, L., Greer, B., and De Corte, E. (2000). *Making sense of word problems*. Lisse: Swets & Zeitlinger.

vomHofe, R., Kleine, M., Blum, W., and Pekrun, R. (2006). The effect of mental models ("Grundvorstellungen") for the development of mathematical competencies. First results of the longitudinal study PALMA. In M. Bosch (Eds.), *Proceedings of the CERME 4* (pp. 142-151). Barcelona: Fundemi IQS - Universitat/ERME.

Walkoe, J. (2014). Exploring teacher noticing of student algebraic thinking in a video club. *Journal of Mathematics Teacher Education, 18*(6), 523-550. doi:10.1007/s10857-014-9289-0.

Wilson, S., and Berne, J (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of research in education, 24*(1), 173-209.doi 10.3102/0091732X024001173.