ZDM Mathematics Education (2010) 42:231–244
DOI 10.1007/s11858-010-0237-x

ORIGINAL ARTICLE

The role of metacognitive monitoring in explaining differences in mathematics achievement

Elmar Cohors-Fresenborg · Silke Kramer · Frank Pundsack · Johann Sjuts · Norbert Sommer

Accepted: 4 February 2010 / Published online: 12 March 2010
© FIZ Karlsruhe 2010

Abstract The relationship between practised monitoring activities and performance, especially in mathematics was examined within three nested studies. The first study deals with problems of faulty term rewritings submitted to three groups of subjects—10th to 13th graders, differing in their mathematical performance—whose task was to find the mistakes. Moreover, a questionnaire on the practice and appreciation of monitoring activities was developed. The third study, first, repeats the first study with a similar population and secondly adds interviews with some of the subjects while solving additional items concerning faulty term rewritings. Studies 1 and 3 show similar success in finding mistakes and in the replies to the questionnaire within the various groups. Furthermore, the third study points up that the subject's answers do neither predict the practised monitoring nor the success in the test. However, the success correlates significantly with the practised monitoring. For a deeper understanding concerning the role of metacognition in explaining performance, the second study examined two of the groups who had already been involved in the first study. These were assigned some problems of a matrices test as used in cognitive psychology. While trying to solve the problem, their eye movements were recorded by means of an eye-tracker. Afterwards they had to justify their solutions in an interview. The eye movements were analysed, the verbal comments classified. Again, the groups differ in their problem solving success, dependant on the quality of the monitoring practised. Altogether, the results of the three studies elucidate the importance of practised metacognitive monitoring activities not only for success in school algebra, but furthermore the ability and the willingness to do it is deeper anchored in a person than just a trained behaviour for school algebra.

Keywords Metacognition · School algebra · Achievement · Figural matrices tests · Eye-tracker study

E. Cohors-Fresenborg (✉) · S. Kramer · F. Pundsack
Institut für Kognitive Mathematik, Universität Osnabrück,
49069 Osnabrück, Germany
e-mail: cohors@mathematik.uni-osnabrueck.de

J. Sjuts
Studienseminar Leer, Bürgermeister-Ehrlenholtz-Str. 15,
26789 Leer, Germany
e-mail: sjuts-leer@t-online.de

N. Sommer
Niedersächsische Schulinspektion, Schloss,
49186 Bad Iburg, Germany

1 Introduction

In the last 10 years of international discussions about how to improve learning mathematics, one focus is on actions which promote students' metacognitive competencies. These are regarded as important in improving the effectiveness and the sustainability of mathematics teaching and learning. Often the analysis of the components of metacognition is based on situations in which a mathematical problem is to be solved. Therefore, one important component is the *planning* of problem solving steps with suitable mathematical tools. Second, the use of the tools has to be controlled, an analysis of the latest state of what has been achieved is necessary, a comparison with the goals set has to be made. The administration of this controlling and comparison is called *monitoring*. A third component is *reflection*, for example, on the given problem, on the understanding of mathematical concepts and tools, or on the discrepancy between representations and conceptions.

In this paper, we deal with the importance of practised monitoring activities for performance, especially in mathematics. The reason for studying this question arose as follows: on one hand, we were led by our project "Analysis of situations for practising metacognition and reflection in mathematics teaching at secondary level",[1] to turn our attention to the question to what extent mental activities of teachers and learners are precisely monitored: computations or argumentations are controlled, discourses are tied and orientated to the goals or the matter just debated. One outcome of our analysis was that the interplay between metacognitive activities practiced and a discursive[2] teaching culture play an important role (Cohors-Fresenborg & Kaune, 2003; Kaune, 2006). On the other hand, it seemed to us that in the international debate on metacognition—in general (e.g., Wang et al., 1993) or in mathematics education (e.g., Kramarski & Mevarech, 2003)—little importance is attached to the component *monitoring*.

This paper focuses on three nested studies concerning the correlation of practised monitoring activities and performance, especially in mathematics. They differ partly in the methodology and the populations, but they have considerable intersections, study one and two with respect to the population, study one and three as regards the methodology.

The first study (Sjuts, 2003) deals with problems of faulty term rewritings which were submitted to three groups of pupils of different performance level who had to find the mistakes. In addition, a questionnaire had to be answered on the practice and estimation of monitoring activities.

Two of the groups were involved in the second study (Brinkschmidt, 2005), in which students had to solve some problems of the test QuaDiPF (Schwank, 1998). It is a kind of advanced progressive matrices test (Raven, 1965) often used as intelligence test, but here the persons have to invent and then justify a suitable pattern. The QuaDiPF-tasks proved to be useful in other experiments to predict typical functional or predicative behaviour of the subjects (Schwank, 1986, 1993). While the students were trying to solve the problem, their eye movements were recorded by means of an eye-tracker. Afterwards they had to justify their solutions in an interview (Schwank, 2001). The eye movements were analysed and the verbal comments classified. The goal of this study is on one hand to analyse how students of the first study show metacognitive monitoring activities in this test, which is not concerned with mathematics, and on the other hand, how metacognitive activities are mirrored in the different eye movement sequences. As those eye movements are done unconsciously, results could give hints how deep the practise of metacognitive activities is anchored in a person.

The third study (Pundsack, 2009) repeats the first study (Sjuts, 2003) with a similar population but adds interviews with some of the participants while they solve additional items concerning faulty term rewritings. The goal is to analyse, on one hand, the correlation between the declaration of metacognitive monitoring and the practised monitoring activities, and on the other hand, the correlation of both to the success in the written test.

The similar correspondence between (practised) monitoring and performance in these studies gives reason to re-analyse some of PISA-2000E results of German pupils under the perspective of monitoring activities needed (Cohors-Fresenborg, Sjuts, & Sommer, 2004). The hypothesis that missing monitoring activities are responsible for the failure in specific tasks can explain some of the results.

2 Two studies on the correspondence between mathematical performance and monitoring

2.1 Theoretical framework

The concept "metacognition" as such was introduced by Flavell (1976) in connection with the analysis of problem solving processes. But the idea that in mathematical problem solving processes phases of planning, monitoring, and reflection have to be organized and connected is older than the concept of metacognition. Polya (1945) introduced it in his famous book "How to Solve It". Schoenfeld (1985, 1992) elaborated these ideas for research in mathematics education. For an overview about the conceptual and methodological considerations concerning metacognition and learning, see Veenman et al. (2006). The importance of metacognitive activities for the success in mathematical problem solving is demonstrated in several studies (for an overview, see Kramarski & Mevarech, 2003). But the question, whether the monitoring or controlling (as part of metacognition) of mathematical activities functions precisely is not specifically in the focus of these studies. For us this question became crucial, when we started to analyze more deeply the results of the PISA-2000E study. In this German addition to the international PISA-2000 study, there had been introduced a set of so called technical items, in which for example, terms had to be re-written and equations to be solved (Neubrand et al., 2001; Neubrand et al., 2004b). Although the competencies which are necessary to solve these items have to be used nearly every day in school-algebra, some of these items have a very high

[1] The project "Analyse von Unterrichtssituationen zur Einübung von Reflexion und Metakognition im gymnasialen Mathematikunterricht der SI" was supported by Deutsche Forschungsgemeinschaft under Co 96/5-1.

[2] By discursivity, we summarize students' and teachers behaviour to insure that, for example, contributions are tied to the matter discussed or the problem given.

PISA-item-difficulty (Neubrand et al., 2004a). From this observation, the hypothesis arose that the failure of pupils in school-algebra is not only a question of specific mathematical content-orientated competences, but also that it comes along with a lack of monitoring and controlling together with a low appreciation of these activities.

2.2 Methodology

The first study (Sjuts, 2003) consisted of two parts: a test on controlling given term-rewritings and a questionnaire on monitoring activities. The test consisted of two sets of six items, which had two similar formats. We give an example of an item in Fig. 1.

The practice and appreciation of monitoring activities was collected by a questionnaire (Fig. 2).

As it has been shown in studies concerning metacognition (Artelt, 2000) and concerning the use of learning strategies (Lind & Sandmann, 2003) that in many cases the self-assessment does not correspond to the learning activities actually practised, a third study has been conducted (Pundsack, 2009). It was a repetition of the first study, because the test and the questionnaire were the same. In addition, clinical interviews were conducted with some of the participants, when they had to solve two items, similar to those in the test. They had to control and argue, whether the given solution of an equation was correct. We give an example of one of the two items in Fig. 3: indicate whether (or not) each response is correct or incorrect.

Some of the transcribed interviews have been classified concerning the metacognitive activities using a classification system which had been developed by Cohors-Fresenborg and Kaune (2007a, 2007b) for analysing discourses in lessons.

2.3 Data collection

2.3.1 First study

The study was carried out with three populations. POP I consisted of the students of two classes of grade 10 from two secondary schools (Gymnasium[3]) (about 16 years old), altogether 82 students. The best 10% of POP I form POP Ia. POP II consisted of 16 students, who belonged to the upper 10% of performers in mathematics in all six classes of grade 10 from another secondary school. POP III consisted of 18 students (16–18 years old), who were all participants in a summer school on cognitive mathematics, to which the University of Osnabrueck had invited top students from all over Germany. These three populations differ in their mathematical performance. Relating to the population in German secondary schools (Gymnasium), they represent an average, an above-average, and a top level.

2.3.2 Third study

The study was again carried out with three populations. This time POP A consisted of 20 students belonging to a preparatory course in mathematics at the University of Osnabrueck. POP B consisted of 31 students, who belonged to the top performers in mathematics in grades 10–13 from two secondary schools (Gymnasium). They were preparing for mathematical competitions. POP C consisted of 11 students (16–18 years old), who were all participants in the summer school on cognitive mathematics in 2008. These three populations differ again in their mathematical performance. Relating to the population in German secondary schools (Gymnasium), they represent a group slightly more than average, above-average, and of top level. Compared

$(-a-b)(-a+b)$		
$=-(a+b)(-a+b)$	correct ○	incorrect ○
$=-(a+b)(b-a)$	correct ○	incorrect ○
$=-(a^2-b^2)$	correct ○	incorrect ○
$=-a^2-b^2$	correct ○	incorrect ○

Fig. 1 Item from the written test

1. I am sure about my answers.
   (1: very unsure, 5: very sure)
   1 ○ 2 ○ 3 ○ 4 ○ 5 ○
2. I have read everything thoroughly.
   (1: not very thoroughly at all, 5: very thoroughly)
   1 ○ 2 ○ 3 ○ 4 ○ 5 ○
3. I have carefully considered possible sources of mistakes.
   (1: not at all, 5: very carefully)
   1 ○ 2 ○ 3 ○ 4 ○ 5 ○
4. I have checked my answers well/ thoroughly.
   (1: not at all, 5: very well/ thoroughly)
   1 ○ 2 ○ 3 ○ 4 ○ 5 ○
5. I have imaginated I would have to justify my answer.
   (1: never, 5: often)
   1 ○ 2 ○ 3 ○ 4 ○ 5 ○

Fig. 2 Metacognition by self-assessment

$$a = b \quad \Leftrightarrow \tag{1}$$
$$a^2 = ba \quad \Leftrightarrow \tag{2}$$
$$a^2 = ab \quad \Leftrightarrow \tag{3}$$
$$a^2 + a^2 - 2ab = ab + a^2 - 2ab \quad \Leftrightarrow \tag{4}$$
$$2\left(a^2 - ab\right) = a^2 - ab \quad \Leftrightarrow \tag{5}$$
$$2 = 1 \tag{6}$$

Fig. 3 Test-item in the interview

[3] In the Germany states there is a divided school system with varying academic orientation, the most demanding of which is the “Gymnasium”. Approx. the upper third of an age cohort attends a “Gymnasium”.

Table 1 Test-results in the first study

	N	Mean of correct answers (%)
POP I	82	40
POP Ia (10% best of POP I)	8	75
POP II	16	75
POP III	18	85

with the first study, POP A is between POP I and II, POP B equals POP II, and POP C equals POP III.

2.4 Results

2.4.1 *First study*

The average percentage of correct answers concerning the different populations is shown in Table 1. As all the populations were familiar with the content of the test the low percentages for POP I and POP II is remarkable. For us, the explanation is that the success in solving the items is merely caused by a lack of monitoring instead a lack of domain-specific knowledge. The difference between the two populations is much higher than expected. In addition, the fact that the students in POP III reach only 85% supports that point of view. If one computes the average of the 10% best of POP I, in Table 1 named as POP Ia, the result of 75% of POP II is reproduced.

Although there are reported doubts about the validity of self-reports concerning "feeling of knowledge" in literature (e.g., Artelt, 2000), for us it is remarkable that the pattern in Fig. 4 fits to the means of correct answers. In all the five answers, the values for POP III lie above those for POP II, and these lie above the values for POP I.

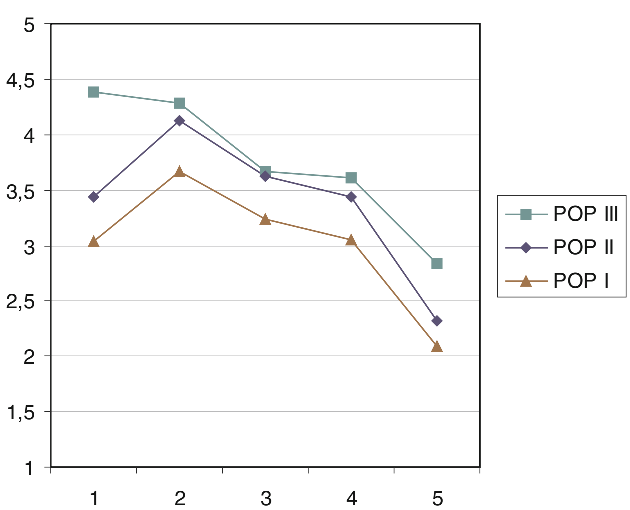

Fig. 4 Mean of answers in the questionnaire

2.4.2 *Third study*

The values in Table 2 for POP B and POP C look nearly like a copy of the values in Table 1. The fact that the percentage of POP A lies between those of POP I and POP B corresponds to the characteristics of the populations: students in a university preparatory course have a higher standard than average students in secondary schools, but they do not belong to the upper 10%.

Concerning the distribution in the answers of the questionnaire (see Fig. 2), it was expected that the results of POP C lie above the results of POP B. Figure 5 shows that this is true with the exception of item 2. But in all three populations, the deviation is quite high for each of the five questions (Table 3).

Analysing the interviews there are large differences in the practised monitoring activities among the participants. Interviews have been transcribed for seven participants. Figures 6 and 7 show excerpts from these interviews, in which the students demonstrate use of control (term rewritings in lines 4–6 of the item shown in Fig. 3). The two excerpts end after the first statement in which the participant indicates that she/he has overlooked a mistake and therefore starts again.

Seven complete transcripts have been classified by the categorizing system (Cohors-Fresenborg & Kaune, 2007b). For the following analysis, only those parts of the transcripts have been classified which deal with the false transformation from line 5 to line 6 in the test-item shown in Fig. 3. The subcategories of "self-monitoring" (**M8**) have been used. In the transcripts, the corresponding text is marked in bold and classified by a specific subcategory in the right column.

Magret is very cautious and precise in her monitoring. Nevertheless she overlooks the mistake. But when she sees the equation $2 = 1$ she indicates that she has overlooked a mistake. This is classified by *r***MS8f** which means a reasonable (*r*) self-monitoring of a metacognitive monitoring activity[4] (**M8f**) done by a student (**S**). Then she decides to start over from the top without any demand of the test-leader (**T**). She detects the mistake without any help.

Ansgar, in contrast, pretends to be monitoring without really being concerned. First, he states that the transformations are correct. Then he withdraws this statement. This is classified as a self-monitoring activity concerning the subject-specific action (**MS8a**). After a break of 8 s, he states again that this is correct. The test-leader (**T**) twice inquires critically, which is classified by a demand (***d***) for a self-monitoring activity concerning the subject-specific

[4] This subcategory has been newly introduced in addition to the categorizing system. It describes a monitoring activity of a person's own monitoring processes from a meta-meta-level.

Table 2 Test-results in the third study

	N	Mean of correct answers (%)	Deviation (%)
POP A	20	67	18
POP B	31	77	19
POP C	11	87	18

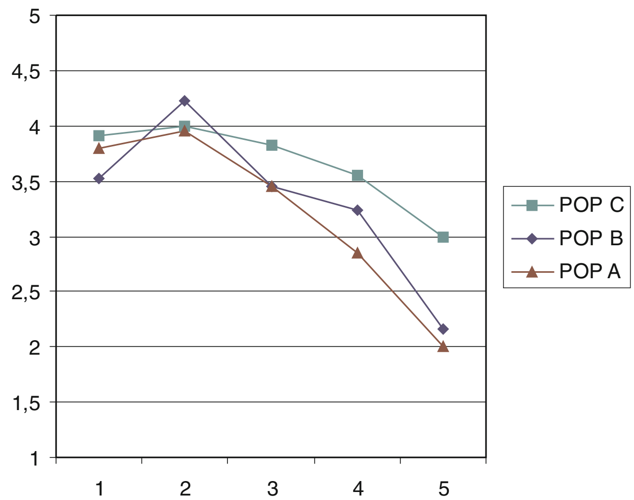

Fig. 5 Mean of answers in the questionnaire

Table 3 Mean (deviation) of answers in the questionnaire

	Question				
	1	2	3	4	5
POP A	3.80	3.95	3.45	2.85	2.00
	(0.83)	(0.83)	(0.89)	(0.88)	(1.45)
POP B	3.52	4.23	3.45	3.23	2.16
	(0.77)	(0.67)	(0.81)	(0.81)	(1.29)
POP C	3.91	4.00	3.82	3.55	3.00
	(0.30)	(0.63)	(0.75)	(0.82)	(1.27)

action (**MT8a**). Then Ansgar changes his opinion ones more and states that it should not be correct. This is again classified by **MS8a**. After this double demand of the test-leader he starts again. He detects the mistake without further help.

As the next step of the evaluation, the transcripts have been grouped in three categories (Table 4): self-monitoring of monitoring (**M8f**) followed by a restart of the monitoring process without demand (2), self-monitoring (but not subcategory **M8f**) followed by a restart of the monitoring process only after demand (1), self-monitoring only after demand, but no restart of the monitoring process (0).

In addition, the percentage of words, which are classified as representing a self-monitoring activity, is computed in relation to the words spoken. By computing a ratio, the

Magret: *ab minus 2 ab*, what remains is *negative ab*, so the (...) from *ab* the *negative 2ab* is subtracted, so that a^2 *minus ab* remains.
[She looks surprised.]
Magret: And then should (...)
(5 sec)
Magret: Then it says here *2 equals 1* which cannot be true. **So, there must be an error somewhere.** *r*MS8f
(6 sec)
Magret: Mmh (...)
(3 sec)
Magret: So the best is if I start over from the top.

Fig. 6 Magret controlling lines 4–6 in item shown in Fig. 3

Ansgar: And on the right side, one has calculated *ab minus 2ab*.
Indeed makes sense.
(4 sec)
Ansgar: And so if one then now divides by the term a^2 *minus ab*, then one gets *2 plus 1*, that means, it is correct.
[He puts a tickmark behind line 6.]
Ansgar: **I think. Cannot be at all. Yes.** MS8a
(8 sec)
Ansgar: Ehm, yes. Actually, I do think that this is correct.
T.: **That this is correct?** *d*MT8a
Ansgar: Ehm, yes.
T.: **Even if it says *Two equals One*?** *d*MT8a
Ansgar: **Yes, actually that should not be correct, I know. But (...)** MS8a
Ehm, I should start afresh.

Fig. 7 Ansgar controlling lines 4–6 in item shown in Fig. 3

differences between very talkative and not very talkative persons are leveled.

In Table 4, it is shown that the self-assessment concerning metacognitive monitoring, represented by the answers in the questionnaire, is quite different from the practised self-monitoring, represented on one hand by the categories 0, 1, 2, and on the other hand by the ratio of classified words to words spoken in the transcripts. This supports the findings of Artelt (2000) and Lind & Sandmann (2003). The self-assessment, in general, cannot be used as indicator for what the respondents to such a questionnaire actually do in practice. The answers of Michaela and Ursula to question 3 and 4 are a good example: Michaela and Ursula both give the same answers, but show very different behaviour.

Concerning the success in the written test, a very good indicator is the category indicating up to which degree self-monitoring is practised and leads to a restart of the monitoring processes: the three persons in the upper half of

Table 4 Answers in the questionnaire in relation to practised self-monitoring and success in the test

Students	Question					Self-monitoring		Success
	1	2	3	4	5	Cat	Words (%)	
Michaela	3	5	4	4	5	0	0	33
Leo	3	4	3	4	5	0	0	42
Lotta	4	5	5	4	3	0	0	42
Ansgar	4	4	4	3	1	1	17	83
Ursula	4	5	4	4	3	1	63	92
Peter	3	3	4	4	3	2	28	100
Magret	4	4	3	3	2	2	61	100

Table 5 Correlations between the success in the test and answers in the questionnaire ($N = 62$)

	Question				
	1	2	3	4	5
Success in the test	0.31	−0.13	0.11	0.15	−0.15
Significance	0.01	0.30	0.40	0.25	0.24

Table 4—belonging to category 0—have, on average, half of the success compared with the other four persons—belonging to category 1 or 2. The correlation between success and this categorization (0 vs. 1 or 2) is 0.976 and statistically significant at the 0.01 level. The only two students who practised self-monitoring of their monitoring (**MS8f**) are the only ones who detected 100% of the mistakes in the test.

But if we look at the answers given in the questionnaire for questions 3 and 4 (see Table 4) in detail, we see a different picture: the values of Lotta do not fit to her success (42%). Even the values of Ansgar and Magret do not indicate the differences in their success. Ansgar's high values, especially for question 3, fit to the impression we got from his behavior in the interview that he only pretends to be monitoring.

It is therefore not astonishing that a statistical analysis for the whole population of 62 participants in the third study shows nearly no correlation between the success in the written test and the answers to the different questions (Table 5).

2.5 Discussion

If we consider studies one and three together, we obtain the following picture: the success in solving term-rewriting items is in all the three levels of mathematical competence (POP I+A, POP II+B, POP III+C) lower than expected. The results correspond to the monitoring activities. But a questionnaire is not a suitable instrument to indicate this.

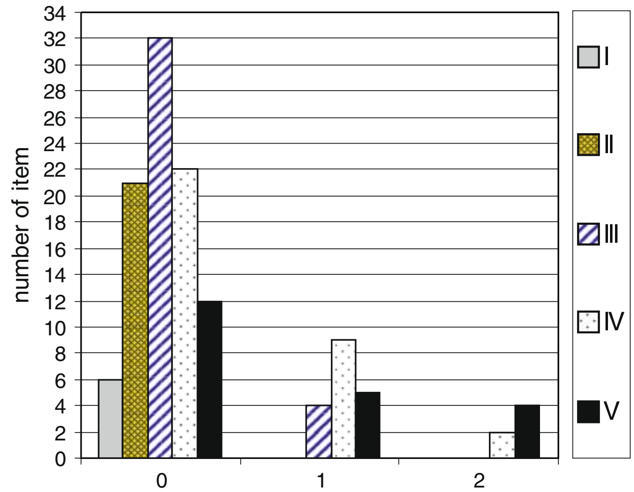

Fig. 8 competence grades and handling formulae

If one has an indicator for the monitoring activities actually practised, this is a predictor for students' success in those parts of school algebra in which the correctness of computations and term rewriting is crucial. It may be useful to distinguish between three levels: on the first level, we find low practising of monitoring, even if there occur "strange results", which should lead to doubts. On the second level, we find a high level of monitoring, but if these processes do not lead to the detection of a hidden mistake, there is no further effort to find it, unless there is a demand from "outside". This could be a teacher or a classmate in a group. The benefit of collaborative work in the studies of Kramarski & Mevarech (2003) could partly be caused by this effect. On the third level, we find a high level of monitoring combined with a self-monitoring of a person's own monitoring. This makes the person independent from an outside interventor.

As term-rewriting items have also been part of the German PISA-2000E study (Neubrand et al., 2001; Neubrand et al., 2004a; Neubrand et al. 2004b), the results of Sjuts (2003) gave reason to re-analyse some of PISA-2000E results of German pupils under the perspective of monitoring activities needed (Cohors-Fresenborg, Sjuts & Sommer, 2004). The handling of formulae has been detected as one of the characteristics which generate difficulties. The feature "Handling of Formulae" especially covers monitoring skills to reliably handle formal mathematical expressions.

The "Handling of Formulae" has been rated at three levels: 0 means not existing, 1 means low competencies needed, the level 2 has been rated, when the competencies needed are above those needed for level 1. That means, level 2 covers a wide range. The items presented here in Figs. 1 and 3 would have been classified with level 2.

The Fig. 8 shows how the items in PISA-2000E are classified concerning "Handling of Formulae" and how often those items belong to the five PISA-2000 competence grades I–V. If an item is classified as level 2, it belongs to the two highest competence grades. With respect to our studies on the correspondence between mathematical competence and actual monitoring activities in practice, especially self-monitoring of a person's own monitoring, this result should mainly not be interpretated as a matter of content-oriented mathematical difficulty of those items, but more as a matter of deficits in actual metacognitive monitoring.

There exists an old study on first year university mathematics students (Gundlach, 1968), in which a correlation was found between their performance in simple school algebra (e.g., computing fractions, solving equations) before university study and their success in a course on abstract analysis. According to our knowledge today about the importance of metacognitive monitoring activities, we would no longer hold to Gundlach's argument concerning mathematical content that such arithmetical skills are obviously necessary for abstract mathematics, but that the ability to practise metacognitive monitoring when term rewriting or concept formatting plays a decisive role.

3 An eye-tracker study on the correlation between behaviour in problem solving and practised monitoring

3.1 Theoretical framework

The distinction between static and dynamic mental modelling as a characteristic of the individual cognitive structure—and not for instance as a characteristic of the task—was introduced by Schwank (1986, 1993). She distinguishes between predicative and functional cognitive structures. Predicative thinking emphasizes the preference of thinking in terms of relations and judgements, functional thinking in terms of processes. This theory was once developed to describe differences in students' behaviour while solving programming tasks. But later it came up to be useful for explaining individual differences in cognitive behaviour outside mathematics (e.g. Cohors-Fresenborg & Schwank 1997). Therefore, the QuaDiPF- test had been developed (Schwank, 1998) to predict typical functional or predicative behaviour of persons in other experiments. It is a advanced progressive matrices test (Raven, 1965). As the interest lies on the decomposition of thinking processes, in QuaDiPF the possibility to start visually matching procedures by not offering a sample choice of eight solutions as in the original APM-test is eliminated. Instead of such a choice, the subjects are asked to draw their solution figure and to argue why their figure fits the pattern. The sessions were videotaped and the reasons given by the subjects were qualitatively analysed.

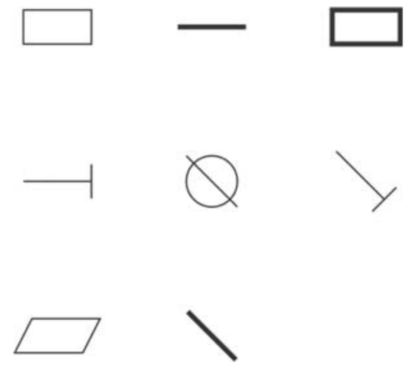

Fig. 9 QuaDiPF: operator item A03

The distinction between predicative and functional thinking is reflected by the different ways of solving the problems. Predicative thinking requires, in order to meaningfully complete the pattern, the subject to get involved with the logic of the static structure of the pattern. Functional thinking requires becoming involved in a dynamic reading of the logic of the pattern; it means a procedure has to be invented which produces the new figure (Schwank, 2003).

Carpenter et al. (1990) attempted to analyse the manner in which participants solve the Raven Matrices using eye-tracking methods. In summary, they stated: "The processes that distinguish among individuals are primarily the ability to induce abstract relations" (p. 404). Schwank (2001) describes this as predicative thinking. But in the Raven test, there is one item for which a solution using predicative thinking is not known. Carpenter et al. (1990) excluded this item from further analysis. They stated the "problem was not classifiable by our taxonomy" (p. 431). For solving this item, functional thinking is needed (Schwank, 2001). For her QuaDiPF test, Schwank invented more of these so-called operator items. An example is shown in Fig. 9.

The second figure of a line and a column has to be understood as a symbol for an operator, which indicates, what has to be done to with the first figure to produce the third figure as a result. Several studies have shown that for persons who prefer predicative thinking these operator items are hard to solve. In an EEG-study using the QuaDiPF-test, it has been shown that for predicative versus functional thinking different brain areas are active (Mölle et al., 2000).

There is a long tradition in using eye-tracker studies to learn more about cognitive processes during problem solving (e.g., De Corte & Verschaffel, 1986). Several studies (e.g., Carpenter & Just, 1976) support the assumption that in general there is a coincidence between the locus of eye fixation and the focus of visual attention. Therefore, from one person's eye movement, it can be derived to her/his focus of interest. Following this idea, Schwank (2001) designed an eye-tracker study using QuaDiPF to look for differences between persons who prefer predicative versus functional thinking. Brinkschmidt (2005) and Armbrust (2006) made detailed studies to elaborate the connection between individual preferences for predicative versus functional thinking and behaviour in different experimental settings, analysing the eye movements of persons while solving QuaDiPF items is one of them.

3.2 Methodology

Kramer na'ee Brinkschmidt (2005) chose 11 items from the QuaDiPF test (Schwank, 1998) for the study concerning predicative versus functional thinking. The participants first had to think about a suitable solution to each item: they had to find the correct match for the free space on the bottom right (Fig. 9) which logically complements both the lines and the columns well. During that time, their eye movements were recorded. Then they had to draw and explain their solution to the test stimulus in a video recorded interview. Six items were finally chosen for the analysis of the preference for either predicative or functional thinking. For two items, there is a predicative and functional mental model, for one item there is only a predicative mental model and for three items there is only a functional mental model. These last mentioned items were numbered A03 (Fig. 9), A05 and A08 in Brinkschmidt (2005). The particularity about these items is that the figures do not all have the same status: the second figure of a line and a column indicates, which changes have to be made to the first figure so that the third one arises. These so-called operators are symbols which do not stand for themselves, but for something else. In Fig. 9, for example, the thick line in the middle of the first line indicates that the rectangle drawn in fine lines is to be changed in such a way that the lines have to thickened. As the test persons with a preference for predicative thinking find it hard to find a solution to this item, differences in solving these three operator items were considerable.

During the study, it became clear that the weaker test persons easily tend to build a mental model of the problem driven by visually striking impulses and thus only take a few measures to gather information which might question this model. It may seem that their metacognitive system controls their perception in such a way that cognitive conflicts are avoided if possible. Low-achieving participants very often overlook the operators when dealing with operator items, which prevents them from realizing that this figure cannot be explained with the help of a logic which refers to lines and columns. In Figs. 10 and 11 are represented those sequences of gazes. These pictures are produced by the software QuaDiPF-Eye developed by Armbrust (Armbrust et al., 2002; Armbrust, 2006). The circles represent the fixations. The area of the circles are proportional to the duration of fixation. The lines represent the saccades (rapid eye movements). For details concerning the technical aspects of the eye-tracker study see Armbrust (2006, pp. 49). For details concerning the aggregation of gazes to groups see Brinkschmidt (2005, pp. 131).

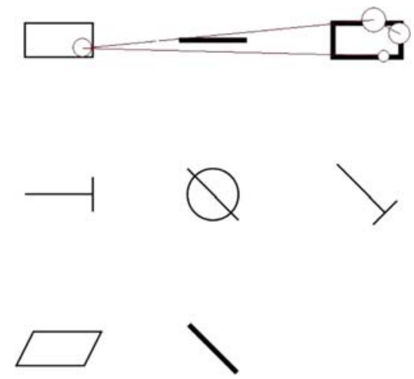

Fig. 10 Sequence of gazes leaving out the operator

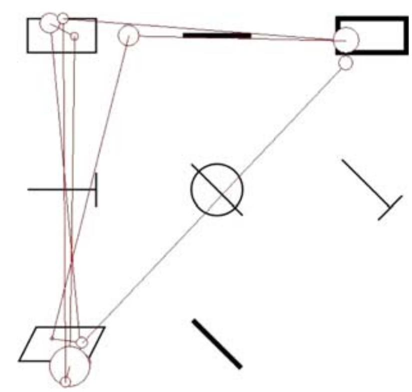

Fig. 11 Sequence of gazes leaving out all the operators

Higher achieving students, however, realize that they only have a chance to come to a conclusion when they intensely consider *all* figures of all lines and columns (see Fig. 12). This observation gave cause to analyse the collected data once more and to check the following:

Hypothesis 1 There are in sum more sequences of gazes in all complete lines and columns and less sequences of gazes which omit the operators with high-achieving students than with lower achieving students.

Another eye movement parameter does not refer to the question which information is to be processed, but how thoroughly in general the information is dealt with. The duration of fixation expresses inter alia the degree of

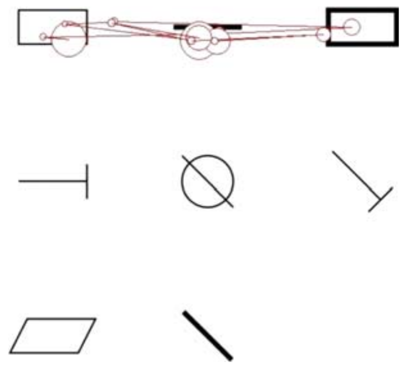

Fig. 12 Sequence of gazes with fixations on all figures

C1: Precision in conception
C2: Abstraction of conception
C3: Awareness of the problem requirements, i.e. to make a statement which logically links all figures of a line and columns.
C4: Checking of consideration with respect to obscurities
C5: Checking of formulations as regards discrepancies.

Fig. 13 Criteria for analysing argumentations

intensity at which the information is processed. Long fixations can be caused by a very thorough and deep processing. Metacognitive processing means extremely long fixation times (Velichkovsky et al., 1997). This leads to the following:

Hypothesis 2 The medium duration of fixation in average is higher with high-achieving students as with lower achieving students.

The analysis of the argumentations suggests that a very differentiated picture of the participants' metacognitive behaviour is achieved, when the participants' comments are classified according to the five criteria in Fig. 13. The following hypothesis results:

Hypothesis 3 On average high-achieving students meet these five criteria more often than lower achieving students.

In analogy to the examination of De Corte & Verschaffel (1986, 1987) mentioned above, the following can be presumed: higher achieving students, if confronted with items, which do not fit their preferred cognitive structure, will detect that they cannot completely argue their solution; therefore, they have some doubts and check their solutions again. As participants are allowed to decide in the session, how much time they will take to work on an item, it can be presumed that the working time of the higher achieving students is longer than that of the other ones. This fact should especially be true for the operator items, because these are quite difficult for persons with a predicative cognitive structure. One indicator that such an item is difficult to solve for a person could be the fact that the person did not solve the item correctly.

Hypothesis 4 If you only consider the working time of students who have not found a correct solution to an item, the working time of higher achieving students lies—on average—above that of lower achieving students.

3.3 Data collection

The study was carried out with two populations. One population consisted of 25 students (16 boys, 9 girls), who belonged to the upper 10% of performers of grades 10 and 11 from two secondary schools (Gymnasium), 13 of them belonged to POP II in the first study. We therefore name this population POP 2. The second population consisted of 32 participants (20 boys, 12 girls) of the summer schools on cognitive mathematics in the years 2002 and 2003. The 16 participants in 2002 belonged to the population III in the first study. We therefore name this population POP 3. In summary, about one half of the population of the second study belonged to the population of the first study, the second half of students had similar prerequisites in each subgroup.

3.4 Results

Hypothesis 1 Compared with lower achieving students, high-achieving students show in sum more sequences of gazes in all lines and columns and less sequences of gazes which leave out the operators. If this is applied to our study, this means, compared to POP 2, POP 3 shows in sum more sequences of gazes in all complete lines and columns and less sequences of gazes which leave out the operators. Figure 14 shows the distribution (in percent) of the sum of sequences of gazes in lines and columns of the three operator items.

The hypothesis that there are less sequences of gazes in the lines and columns with POP 2 is therefore correct. It is, however, striking that the difference between the populations is smaller for the sums of sequences of gazes in lines and columns in the first item of this type than it is for the two following items. A03 does not yet show a significant difference between the populations according to t test, but A05 ($F = 0.046$, $p = 0.05$) and A08 ($F = 1.108$, $p = 0.001$) do. In POP 2, there is a significant number of participants who have decided that the complete lines and columns do not really have to be taken into consideration for the first item of this type, but that it is more promising to leave out the operators. This is also revealed in Fig. 15, in which the percentages of the sum of sequences of gazes leaving out the operators in operator items are compared to one another.

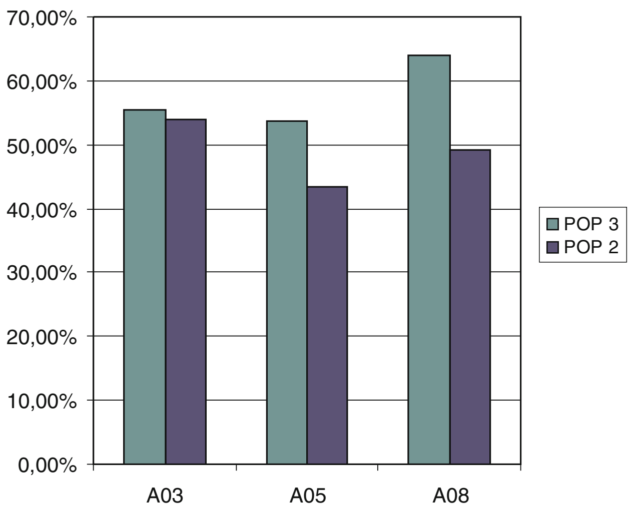

Fig. 14 Parts of sequences of gazes in lines and columns

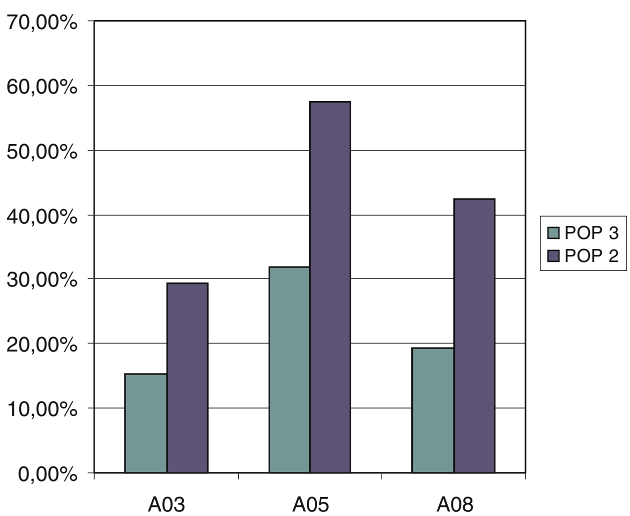

Fig. 15 Parts of sequences of gazes leaving out the operators

From A03 to A05, the percentage of sequences of gazes of this type increases considerably with the two populations and reaches its maximum which coincides with the lowest frequency of problem solving in both populations (only 12% in POP 2 and only 33.3% in POP 3). This emphasizes that these sequences of gazes are indeed not helpful for finding the solution. POP 3, as well, does not show any learning effect from A03 to A05. This is also proved by the low percentage of the sum of sequences of gazes in lines and columns. The percentage of sequences of gazes leaving out the operators in problem A08 decreases considerably in this test group, at the same time, the percentage of the sum of sequences of gazes in lines and in columns increases significantly. This coincides with the highest frequency of correct problem solving (57.6% compared to A03 with only 48.5%). The differences between the two populations are statistically significant with all three items according to the t test (A03: $F = 11.449$, $p = 0.002$; A05: $F = 5.524$, $p = 0.002$; A08: $F = 20.538$, $p = 0.016$).

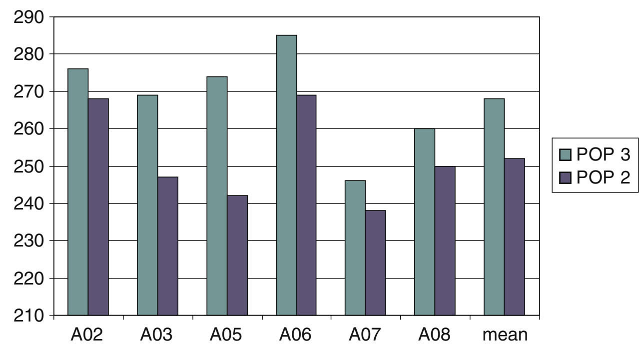

Fig. 16 Medium fixation time in ms

Hypothesis 2 The medium duration of fixation of high-achieving students is higher than that of lower achieving students on average. Applied to our examination this means: the medium duration of fixation of POP 3 is higher than that of POP 2. Figure 16 shows that this hypothesis is correct.

It is to be seen that the medium duration of fixation of POP 3 is higher for each problem than that of POP 2. The difference is extremely high regarding A03 and A05. The t test shows in item A05 that this difference is considerable ($F = 7.456$, $p = 0.016$). The difference is not significant with all the other items. However, the probability that it is random that the medium fixation time is higher for every item in POP 3, only comes up to 3.1%.

Hypothesis 3 On average, high-achieving students meet the five criteria (Fig. 13) more often than lower achieving students. If this is applied to our examination it means: on average POP 3 meets the five criteria more often than POP 2.

With respect to the use of metacognition, it is especially interesting to analyse the argumentations of those persons who do not see the functional coherence between the figures, because then it becomes extremely difficult to generate conclusive rules for the construction of solution figures. A predicative approach to the problem (see Schwank, 2003) oriented to the characteristics of the figures can lead to any complex argumentation. This type of argumentation demands a high degree of preciseness and monitoring in order to avoid mistakes. Examples can be found in Brinkschmidt (2005, pp. 125).

To be able to compare the results of POP 2 better to those of POP 3, the evaluation concerning the five criteria was reproduced to the values −1 (in the case of “does not

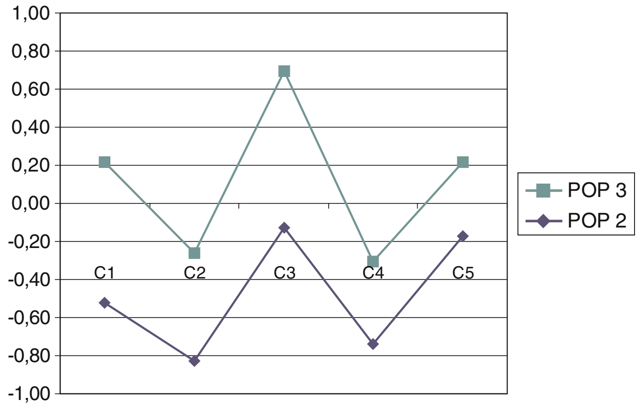

Fig. 17 Mean of meeting the five criteria

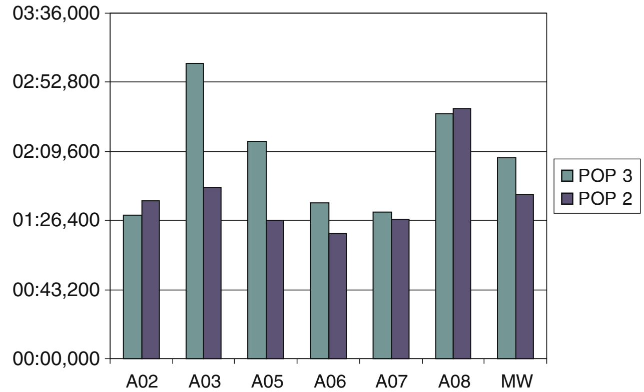

Fig. 18 Working time in m:s:ms

apply at all"), 0 (in the case of "unable to be determined") and 1 (in the case of "applicable"). The graphics (Fig. 17) show the average values gained from these data for item A03.

Point C1 shows the average values for the first criterion, point C2 those for criterion 2, etc. The average values of POP 3 lie above those of POP 2 in all points.

The "Mann–Whitney *U* Test" shows that all differences are significant, apart from point C4 where the difference only lies at the limit of significance with a *Z* value of -1.709.

All average values of POP 2 lie below zero. This means, the majority of participants achieve a negative classification with all five criteria. The average value of POP 3 concerning the question regarding precision of conception, the knowledge concerning the demands of a problem and the checking of the formulations lies above zero. The majority of the participants have these aspects under control. This goes especially for the knowledge concerning the demands of an item, where the average value is highest. Here, the difference to POP 2 is most significant: the argumentations of the majority of these participants show that they have not fully understood the setting of the item. As the participants, who have not adhered to the setting of an item, have always been made aware of it by the investigator, it can only be presumed that they did not think that the exact adherence to the setting of the item was important. The lowest value for POP 2 is point C2 (the question if the conception is abstract/general). Only 2 of 25 participants use an abstract description. The analyses of argumentations with respect to the five criteria concerning the other two operator items lead to similar results (Brinkschmidt, 2005, pp. 310).

Hypothesis 4 If you only consider the working time of students who have not found a correct solution to an item, the working time of higher achieving students lies—on average—above that of lower achieving students. If we apply this to our study this means: if you only consider the working time of students who have not found a correct solution to an item, the working time of POP 3 lies—on average—above that of POP 2.

The diagram (Fig. 18) compares the working time of the participants who have not found a correct solution:

For participants from POP 3, the working time for all problems with incorrect solutions lies above that for participants from POP 2. According to the *t* test, these differences are significant with A03 ($F = 4.184$; $p = 0.025$) and A05 ($F = 11.632$; $p = 0.020$), two of the operator items.

3.5 Discussion

3.5.1 Sequences of gazes

The evaluation of the sequences of gazes has shown that the visual perception of low-achieving students is often operated in such a way that possible cognitive conflicts are avoided, which might come up when a spontaneously developed idea to solve an item proves to be unsuitable after checking one more time. As regards the operator problems, it is striking that, on average, POP 2 shows considerably more sequences of gazes where operators are left out. De Corte & Verschaffel (1986, 1987) noticed the same in examinations where pupils of the first year at school had to solve word problems while their eye movements were measured. Low-achieving pupils stand out because of the fact that they looked at certain pieces of information in isolation. They often only concentrated on the numbers given in the text without reading the context in which the numbers appeared. High-achieving pupils, however, also read the words and whole sections of the sentence in a context. Thus, the phenomenon is exactly the same as with problems dealing with the completion of patterns. Low-achieving participants stand out by the way they consider pieces of information in isolation (e.g., only the first and third figure of a line or column), whereas high-achieving students consider the information given in a context, i.e. they read “whole sections of sentences” (all figures of a line or column).

3.5.2 Medium duration of fixation

The visual perception of the higher achieving students is more efficient as they process the information achieved more deeply. This is proved by the, on average, higher medium fixation time. According to Velichkovsky et al. (1997), this is an indicator for more metacognitive activities in POP 3 than in POP 2.

3.5.3 Argumentations

If argumentations of students are analysed the participants of POP 3 show metacognitive knowledge and activities to a higher extend than those of POP 2. This is another analogy to the examinations in De Corte & Verschaffel (1986, 1987). Pupils of their first year at school often ignored the question of the word problem and did not even read it. This corresponds to the observations made in this evaluation, for example, low-achieving students often ignore the setting of the problem.

3.5.4 Working time

The fact that—among those persons, who could not find a correct solution—the participants of POP 3 significantly took more time for the operator items A03 and A05 than those of POP 2, proves that the metacognitive component “monitoring” is clearly more distinctive as regards high-achieving students. They also make mistakes, but not so carelessly as low-achieving students.

4 Discussion

All over the world, there is a debate concerning low competencies of students in mathematics, especially in school algebra. During the last decades, there have been a lot of attempts to change this situation. These concern especially a change in curriculum. As international studies like TIMSS and PISA show, in many countries, the success of these attempts is not convincing.

Several studies have shown that there is a correlation between mathematical performance and metacognitive skilfulness. The influence of metacognition is partly independent of intellectual ability (Veenman et al., 2005). Studies have pointed out that it is possible to train metacognitive behaviour and to increase mathematical performance by this means (Kramarski & Mevarech, 2003; Mevarech & Amrany, 2008).

But, especially in school algebra, it is important to organize the precise use of mathematical tools. The metacognitive monitoring of their use is therefore crucial. Our studies have shown that the practising of monitoring is indeed an indicator of success. But the declaration of monitoring neither correlates with the practising nor with the success. Therefore, the effort in math teaching has to be laid on changing the students’ monitoring behaviour and not only their metacognitive knowledge.

Our studies indicate that the lack of use of metacognitive monitoring activities in practice can be regarded as an important factor to explain failure of German students in the PISA-2000E study (Cohors-Fresenborg et al., 2004). Results of a Chinese-German comparative study with pupils between 10 and 12 years old (Sjuts & Xu, 2007) point to the same conclusion.

As pointed out in Sect. 2.5, a reanalysis of a 40-year-old study on first-year university mathematics students (Gundlach, 1968) leads to the explanation that the lack of practised metacognitive monitoring activities is the cause for the correlation, which had been found between the students’ performance in simple school algebra before the university study and their success in a course on abstract analysis.

Our eye-tracker study concerning monitoring during non-mathematical problem solving indicates that the willingness for monitoring is deeply anchored in a person. Parts of metacognitive monitoring process unconsciously. The debate on aspects of unconscious metacognitive monitoring (see Veenman et al., 2006) should therefore be continued.

As a consequence, the efforts on improving the success of mathematics teaching and learning should focus on the classroom culture: precise monitoring of the intellectual activities of teachers and learners, the effort for anchoring the discourses on the matter just debated and orientation to the goals set should belong to teachers’ and students’ intellectual habit, not only in mathematics. That means, metacognitive and discursive activities, both of teachers and students, play an important role for the quality of teaching learning processes. We have developed a categorizing system for those activities (Cohors-Fresenborg & Kaune, 2007a, 2007b). This allows to represent the appropriate classroom culture. On this bases, profiles can be created concerning the metacognitive and discursive activities in a classroom (or peer group) discussion. These can be used, on one hand, in research and, on the other hand, for teacher education and in service teacher training. The categorizing system is applicable for analysing lessons in which stepwise controllable argumentations take place, that means not only math lessons. Further research is, therefore, needed concerning the question, whether there exists a kind of intellectual behaviour commonly shared in different school subjects.

References

Armbrust, S. (2006). *Die Werkzeuge "CoDyLa" und "QuaDiPF-Eye" zur Untersuchung funktionalen/prädikativen Denkens sowie ihre empirische Erprobung*. Osnabrück: Forschungsinstitut für Mathematikdidaktik.

Armbrust, S., Schwank, I., & Libertus, M. (2002). Augenblickbewegungen beim Lösen von Matrizenaufgaben (QuaDiPF). In E. van der Meer, H. Hagendorf, R. Beyer, F. Krüger, A. Nuthmann, & S. Schulz (Eds.), *43. Kongress der Deutschen Gesellschaft für Psychologie* (p. 378). Lengerich: Pabst Science Publishers.

Artelt, C. (2000). *Strategisches Lernen*. Münster: Waxmann.

Brinkschmidt, S. (2005). *Über die Unterschiedlichkeit kognitiver sowie metakognitiver Prozesse beim Bearbeiten von QuaDiPF-Aufgaben—Empirische Untersuchungen mit Blickbewegungsanalysen*. Osnabrück: Forschungsinstitut für Mathematikdidaktik.

Carpenter, P. A., & Just, M. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*, 441–480.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven progressive matrices test. *Psychological Review, 97*(3), 404–431.

Cohors-Fresenborg, E., & Kaune, C. (2003). Unterrichtsqualität: Die Rolle von Diskursivität für "guten" gymnasialen Mathematikunterricht. In *Beiträge zum Mathematikunterricht 2003* (pp. 173–180). Hildesheim: Franzbecker.

Cohors-Fresenborg, E., & Kaune, C. (2007a). Modelling classroom discussions and categorising discursive and metacognitive activities. In D. Pitta-Pantazi, G. Philippou (Ed.), *European research in mathematics education V—Proceedings of the fifth congress of the European Society for Research in Mathematics Education* (pp. 1180–1189). Larnaca: Department of Education, University of Cyprus.

Cohors-Fresenborg, E., & Kaune, C. (2007b). *Kategoriensystem für metakognitive Aktivitäten beim schrittweise kontrollierten Argumentieren im Mathematikunterricht 2. überarbeitete Auflage. Arbeitsbericht Nr. 44*. Osnabrück: Forschungsinstitut für Mathematikdidaktik.

Cohors-Fresenborg, E., & Schwank, I. (1997). Individual differences in the managerial representation of business processes. In R. Pepermans, A. Buelens, C. Vinkenburg, & P. Jansen (Eds.), *Managerial behaviour and practices: European research issues* (pp. 93–106). Leuven: Acco.

Cohors-Fresenborg, E., Sjuts, J., & Sommer, N. (2004). Komplexität von Denkvorgängen und Formalisierung von Wissen. In M. Neubrand (Hrsg.), *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland: Vertiefende Analysen im Rahmen von PISA-2000* (pp. 109–144). Wiesbaden: Verlag für Sozialwissenschaften.

De Corte, E., & Verschaffel, L. (1986). Eye-movements of first graders during word problem solving. In M. Askew (Ed.), *Proceedings of the International Conference for Psychology of Mathematics Education (PME10)* (pp. 421–426). London: University of London, Institute of Education.

De Corte, E., & Verschaffel, L. (1987). First graders eye movements during elementary addition and subtraction word problem solving. In G. Lüer & U. Lass (Eds.), *Fourth European Conference on Eye Movements, Volume 1: Proceedings* (pp. 148–150). Göttingen: Hogrefe.

Flavell, J. (1976). Metacognitive aspects of problem solving. In L. Resnik (Ed.), *The nature of intelligence* (pp. 231–236). Hillsdale, NJ: Lawrence Erlbaum.

Gundlach, K.-B. (1968). Kenntnisse der Abiturienten und Studienerfolg in den Anfängervorlesungen im Fach Mathematik. *Mathematisch-Physikalische Semesterberichte, XV*, 20–31.

Kaune, C. (2006). Reflection and metacognition in mathematics education—Tools for the improvement of teaching quality. *Zentralblatt für Didaktik der Mathematik, 38*(4), 350–360.

Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal, 40*(1), 281–310.

Lind, G., & Sandmann, A. (2003). Lernstrategien und Domänenwissen. *Zeitschrift für Psychologie, 211*(4), 171–192.

Mevarech, Z. R., & Amrany, C. (2008). Immediate and delayed effects of meta-cognitive instruction on regulation of cognition and mathematics achievement. *Metacognition and Learning, 3*, 147–157.

Mölle, M., Schwank, I., Marshall, L., Klöhn, A., & Born, J. (2000). Dimensional complexity and power spectral measures of the EEG during functional versus predicative problem solving. *Brain and Cognition, 22*(3), 547–563.

Neubrand, M., Biehler, R., Blum, W., Cohors-Fresenborg, E., Flade, L., Knoche, N., et al. (2001). Grundlagen der Ergänzung des internationalen PISA-Mathematiktests in der deutschen Zusatzerhebung. *Zentralblatt für Didaktik der Mathematik, 33*(2), 45–59.

Neubrand, M., Biehler, R., Blum, W., Cohors-Fresenborg, E., Flade, L., Knoche, N., et al. (2004a). Eine systematische und kommentierte Auswahl von Beispielaufgaben des Mathematiktests in PISA 2000. In M. Neubrand (Hrsg.), *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland: Vertiefende Analysen im Rahmen von PISA-2000* (pp. 259–270). Wiesbaden: Verlag für Sozialwissenschaften.

Neubrand, M., Biehler, R., Blum, W., Cohors-Fresenborg, E., Flade, L., Knoche, N., et al. (2004b). Der Prozess der Itementwicklung bei der nationalen Ergänzungsuntersuchung von PISA 2000: Vom theoretischen Rahmen zu den konkreten Aufgaben. In M. Neubrand (Hrsg.), *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland: Vertiefende Analysen im Rahmen von PISA-2000* (pp. 31–49). Wiesbaden: Verlag für Sozialwissenschaften.

Polya, G. (1945). *How to solve it*. Princeton: Princeton University Press.

Pundsack, F. (2009). *Zusammenhang von Monitoring und mathematischer Leistung—eine (empirische) Studie und Entwicklung eines Trainingsprogramms*. Unpublished Master Thesis. Osnabrück: Universität Osnabrück.

Raven, J. C. (1965). *Advanced progressive matrices*. Sets I and II. London: Lewis.

Schoenfeld, A. H. (1985). Metacognitive and epistemological issues in mathematical understanding. In E. A. Silver (Ed.), *Teaching and learning mathematical problem solving: Multiple research perspectives* (pp. 361–380). Hillsdale, NJ: Erlbaum.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334–370). New York: Macmillan.

Schwank I. (1986). Cognitive structures of algorithmic thinking. In *Proceedings of the 10th conference for the psychology of mathematics education* (pp. 195–200). University of London, Institute of Education, London

Schwank, I. (1993). On the analysis of cognitive structures in algorithmic thinking. *Journal of mathematical behaviour, 12*(2), 209–231.

Schwank, I. (1998). *QuaDiPF—Qualitative diagnostic instrument for predicative versus functional thinking*. Osnabrück: Forschungsinstitut für Mathematikdidaktik.

Schwank, I. (2001). Analysis of eye-movements during functional versus predicative problem solving. In J. Novotna (Ed.), *European research in mathematics education II* (pp. 489–498). Prague: Charles University.

Schwank, I. (2003). Einführung in prädikatives und funktionales Denken. *Zentralblatt für Didaktik der Mathematik, 38*(4), 350–360.

Sjuts, J. (2003). Metakognition per didaktisch-sozialem Vertrag. *Journal für Mathematik-Didaktik, 24*(1), 18–40.

Sjuts, J., & Xu, B. Y. (2007). Mehr Erfolg mit Metakognition? Ergebnisse einer chinesisch-deutschen Vergleichsuntersuchung. *SEMINAR, Lehrerbildung und Schule, 12*(2), 59–75.

Veenman, M. V. J., Kok, R., & Blöte, A. W. (2005). The relation between intellectual and metacognitive skills in early adolescence. *Instructional Science, 33*, 193–211.

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: conceptual and methodological considerations. *Metacognition and Learning, 1*, 3–14.

Velichkovsky, B. M., Sprenger, A., & Unema, P. J. A. (1997). Towards gaze-mediated interaction: Collecting solutions of the "Midas touch problem". In S. Howard, J. Hammond, & G. Lindgaard (Eds.), *Human-Computer Interaction: INTERACT '97*. London: Chapman & Hall.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63*(3), 249–294.