Erscheint in : Zehetmeier, S., Rösken-Winter, B. , Potari, D., Ribeiro, M. (Eds.). (2016). *ERME Topic Conference on Mathematics Teaching, Resources and Teacher Professional Development*, S. 46-55.
https://hal.archives-ouvertes.fr/ETC3/

The design of a high inference rating system for an evaluation of metacognitiv-discursive instructional quality

Edyta Nowińska

University of Osnabrück and Adam Mickiewicz University in Poznań,
enowinska@uni-osnabrueck.de

Due to the lack of observational systems for evaluation of metacognition in mathematics instruction, rarely anything is known about promoting metacognition when teaching and learning mathematics in class. This paper discusses the design of a rating system developed for an evaluation of metacognitive-discursive instructionnal quality (MDQ). It focusses on this system understood as a research tool aiming at a reliable and valid evaluation of MDQ in class, but it also refers to the usefulness of this tool when applying it as an analytical or diagnostic tool in teacher education. The paper explains a two-step procedure for the application of the rating system, and discusses how this procedure enhances the accuracy of evaluation of instructional quality.

Keywords: metacognition, discursivity, rating system, evaluation.

INTRODUCTION

Despite the recognition of the important role of metacognition in student's learning process (cf. Hattie, 2009; Wang, Haertel, & Walberg, 1990), little is known about the implementation of metacognition in classroom instruction and on relations between teacher's and students' metacognition in a class. Assuming that enhancing learners' metacognition is essential for promoting learning, research on the implementation of metacognition into the school practice, and on supporting teachers in establishing a metacognitive-discursive culture in their classes definitely merits future research (cf. Mevarech & Kramarski, 2014; Depaepe et al., 2010). One challenge for this kind of research had been described by Veenman et al. (2006, p. 10): "Teachers are absolutely willing to invest effort in the instruction of metacognition within their lessons, but they need the tools for implementing metacognition as an integral part of their lessons". This statement raises the issue of the kind of tools that would be appropriate as diagnostic tools for analysing, assessing, and improving metacognitive practices of individual teachers and their students in a class.

This paper reports on a research project aimed at developing and evaluating a rating system for analysing and assessing metacognitive-discursive instructional quality (MDQ) of a class discussion (RSMDQ) (Nowińska, 2016). RSMDQ can be used in different settings and for various research and practical aims:

- as a *research tool* to evaluate the metacognitive-discursive instructional quality in a given teaching-learning group (e.g. when a relation between this quality and students' learning gains should be investigated),

- as an *analytical tool* in in-service and pre-service teacher education (e.g. to learn how to analyse a classroom discussion, and to identify variables influencing the precision, coherence and clarity of the mathematical and meta-mathematical content of the discussion),
- as a *diagnostic tool* in in-service teacher professional development programmes (e.g. to analyse and diagnose the metacognitive-discursive quality in a particular class of an individual teacher, and to find strategies for improving its potential for promoting learners' metacognitive behaviour).

This paper explains the design of RSMDQ and exemplifies its use as a research tool but it also discusses the usefulness of RSMDQ in the other two settings.

METACOGNITION IN A MATHEMATICS INSTRUCTION

The origin of research in metacognition in mathematics education lies in learners' difficulties in *solving problems*, and is closely related to the question of how to learn and teach solving non-routine problems. When applying to *learning* mathematics in a class, metacognition refers to a broader spectrum of activities then during problem solving. The groundwork for the operationalisation of the concepts of metacognition with the objective of making it understandable und evaluable in terms of empirical observations of classroom situations has been done by Cohors-Fresenborg and Kaune (2007), as they constructed a category system for an interpretative, transcript-based analysis of metacognitive and discursive activities (CMDA)[1] in class discussions. Their paper presents this category system, explains and exemplifies in detail the use of it. CMDA does not differentiate between metacognitive processes understood as cognition about (one's own or of the others) cognition – in particular when problem solving – and cognition about the results from cognition (calculation, verbal or written information, argumentations, questions). Also in the second case the purposeful application of such cognitive behaviour at the appropriate moment results from metacognitive thoughts, and reflects the intention to control and understand the given calculation, information, argumentation or question. According to this conceptualisation, the objectives of metacognition[2] in learning mathematics are, for example, to *plan* the use of mathematical tools, methods, and representations to justify an argumentation or to explain an idea; to *control and evaluate* the accurateness of argumentations, the adequateness of external (e.g. formal) or internal representations of mathematical concepts, the correctness of the use of tools and procedures; to *reflect* on the ways of reasoning, defining or proving, and on similarities and differences in conceptions and arguments. Since a learning process in a class can only lead to a deep understanding of concepts, representations and tools, if the planning, monitoring and reflection related to them are connected accurately to the matter discussed and take

---

[1] The complete German version of CMDA is presented in Cohors-Fresenborg, Kaune, & Zülsdorf-Kersting (2014).

[2] Metacognition in this context is decomposed into *planning*, *monitoring* and *reflexion*. The colors used here refer to the colors used in CMDA. In the same way two colors are used for *discursivitiy* and *negative discursivity*.

students' ways of thinking into consideration, *discursivity* is needed to facilitate the productive use of metacognition in classroom discussions (Cohors-Fresenborg et al., 2014, p. 7). Discursivity means activities carried out to improve the coherence and precision of a discussion: orchestrating single utterances to a comprehensible discourse unit and orientation on students' ways of thinking. Examples of discursive activities are precise (re-)for-mulating and comparing of learners' ideas, strategies, conceptions and misconceptions, connecting them with precisely represented mathematical concepts or argumentations, and confronting the learners with problems regarding the intended precision. On the contrary, *negative discursivity* means activities with a negative influence on the coherence and precision of a discussion. Examples are: the use of inadequate vocabulary or superficially clear sentences with an unclear sense, incorrect logic structure of an argumentation, and bringing into the discourse an alternative idea without references to what has been said and discussed before.

This broad perspective on metacognition in mathematical instruction was used to design RSMDQ, because it provides a wide range of concepts for a detailed analysis and of metacognition in everyday teaching and learning situations, and not when solving mathematical problems.

RATING SYSTEM FOR AN EVALUATION OF MDQ IN A DISCOURSE

In the ongoing research project[3], the Group Cognitive Mathematics at Osnabrueck University (in cooperation with the University of Kassel and The German Institute for International Educational Research - DIPF) works on the development and evaluation of a rating system for a *video-based* analysis (by a *category system*) of metacognitive and discursive students' and teacher's activities in classroom discussion, and for the evaluation (by *rating scales*) of the MDQ of the analysed discussion. By means of a generalizability study (Cronbach et al., 1972; Praetorius et al., 2012) the reliability of the designed rating system (RSMDQ) will be evaluated, and decision studies will be conducted to determine how many lessons from a given teaching-learning group, and how many raters would be needed to get reliable (generalizable) statements about the MDQ in this group.

The design of the rating system

The design of RSMDQ and of its application is influenced by the fact that metacognition and discursivity are intertwined and can be carried out with a different local quality (e.g. elaboration, precision, relevance for the discussed question). Both constructs have to be analysed in teacher-student and student-student interactions, and their potential to foster understanding of mathematical issues, tools and methods discussed in class have to be analysed. Consequently, when assessing metacognitive

[3] The project is supported by Deutsche Forschungsgemeinschaft under reference Co96/8-1.

practices in class, it is necessary to look at several aspects of these practices, and to integrate them into a global complex assessment. This can only be achieved by a *high inference* rating system, but the use of such a system demands complex qualitative decisions and a high degree of the necessary conclusions. This leads, in general, to a reduced reliability of the assessment (Praetorius et al., 2012). An improvement of the reliability through reducing the focus of the evaluation to *observable* or *countable* aspects of metacognitive behaviour does not make sense. Right from the design stage for such a rating system, it has to be prevented that, as a consequence of the pressure to get a satisfactory reliability, only the surface structure of the lesson is evaluated and the deep structure neglected. This could provide incorrect predictions as to what extent a 'good' surface structure of a lesson can promote an understanding learning and lead to sustainable results of this process (cf. Nowinska, 2011).

In the ongoing research project a new idea to cope with this research problem was developed. To obtain reliable assessments, despite the needed complexity of the interpretative discussion analysis, the rating process is designed as a *two-steps* procedure. The decision was made, not to dramatically reduce the complexity of the category system CMDA but to adapt it for a *video-based* analysis, and to use it as an analytical tool for a detailed interpretation of each student and teacher contribution in the *first step*. The same rater uses his interpretation as a basis for the global evaluation of MDQ of the whole class discussion in the *second step*. The obligation for an elaborated interpretation causes that the rater deals with the videos very intensively, and therefore it can be expected that the evaluation of MDQ will be reliable and accurate.

The idea of using these two steps plays a crucial role contributing to the usefulness and practicability of RSMDQ as a research tool, and also as an analytical and diagnostic tool. The result of the first step provides a detailed 'map' of metacognition and discursivity in class interactions and makes these constructs 'visible' for researchers, raters and teachers. This enhances the accurateness of the subsequent evaluation, and helps an individual rater or teacher to analyse the weak and strong aspects of an individual teacher's efforts in fostering understanding in mathematics. To make this usefulness comprehensible to the reader of this paper, a *reference example* will be used in the following sections. The subsequent analysis of this example shows how the 'map' of the metacognitive and discursive activities identified in the given class discussion can be analysed with RSMDQ to find a global comprehensive evaluation of MDQ.

The reference example

The example presents transcript-excerpts from a discussion in a grade 7 classroom in one German secondary school. In the previous lesson the teacher (T.) introduced two types of equivalent transformations (ET) to solve linear equations by writing them on the board. Up to the end of the lesson the ETs have not been explained and justified.

The first (second) ET regards ‘the addition and subtraction of the same number (term) on both sides of an equation’. Two equations have been solved by a sole application of these transformations. In this lesson the students have to solve the equation 4-x=6.

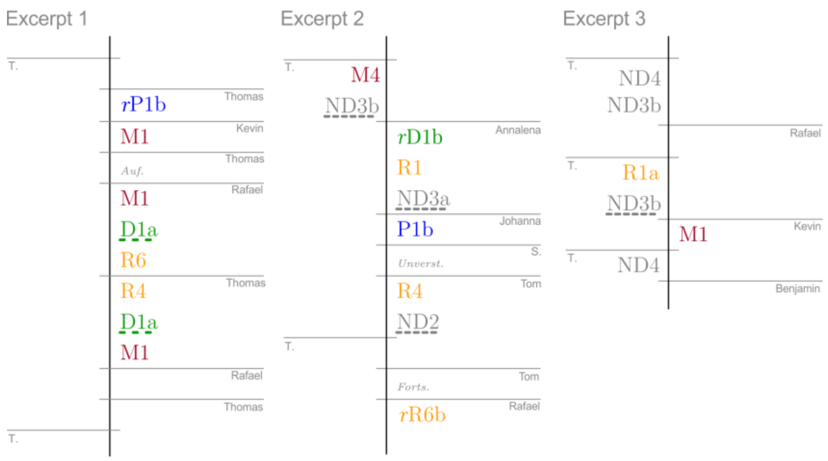
Thomas	The task was 4-x=6. I thought, we could bring x to the other side of the equal sign, and this would make the task easier. Namely, quite simply, plus x, then one has 4=6+x. And now one can see, if one wants to have only x, one has there also the six, hence minus six. And then it is 2=x.
Kevin:	I would say this is correct.
Thomas:	Rafael.
Rafael	I would say, the result, er, I would say it is correct, but I do not know how you got the idea. I did not understand how you, er, how you got 4=6+x. *[a few minutes later:]*
T.	How about the others? This could call in your minds the second type of the equivalence transformations. (8sec) Have a look into your notes. *[one minute later:]*
Johanna	Actually, I only calculated 4-6. I got 2. *[...]* I do not understand why one has to do all these complicated steps with x, if one can just calculate 4-6.
T.	Johanna, we are looking for the number x. We have already said it several times during last lesson. We are trying to change the equation, to transform it so that at the end we get the answer to the question: What is x? On the left side, there is minus x. This does not satisfy Thomas.

Thomas suggests (and justifies) a plan to solve the equation and to make it easier first. Kevin and Rafael control the result obtained by Thomas. Rafael gives a critical reflective question concerning Thomas’ plan. Johanna reflects on her difficulties in understanding the sense of Thomas’ idea, and justifies her critical remarks by pointing to an easier way of solving the equation. The transcript shows that the learners are autonomous in planning the way for solving the equation, in critical controlling of the use of mathematical tools, and in reflecting on the sense, usefulness and complexity of these tools. They also try to understand what their classmates think. Their single metacognitive activities indicate a great potential for understanding the idea of solving equations, but – since the different learners’ conceptions of solving an equation have not been elaborated and compared with each other – they do not produce a comprehensible discourse unit. The teacher does not initiate deep reflection on the mathematical activity. Instead of that, he only points to the type of ET (written on the blackboard) that has to be used there.

The first step: local categorizing of metacognitive and discursive activities

In the first (video-based) coding step the rater interprets each students’ and teacher’s contribution. He decides whether a given contribution indicates metacognitive and (negative) discursive activities, he interprets the kind of these activities, and describes them with codes from the category system. This interpretation is based on the category system adapted[4] for this purpose from Cohors-Fresenborg and Kaune (2007). The choice of one category from the category system demands from the rater a careful consideration with regard to alternative categories, and a justification which one of them is the most adequate in the given case. Thus, the rater sets for himself a local interpretation of the class discussion, and gets an overview of the kind and quality (precision and elaboration) of each individual activity on the one side, and of the coherence of the whole discussion on the other side.

One result of the categorizing is presented to the rater in a form called *category line*. It can be considered as a map or an abstract representation of the discussion process and its metacognitive and discursive content. The category line is to be read from top to bottom. For each student's and teacher's contribution there is a short horizontal segment with the name of the speaker. The segments for student's contribution are represented on the right side of the vertical line, and these for the teacher (T.) on the left side. Under each segment, there are codes for metacognitive and discursive activities identified by the rater in the respective contribution. The following figure shows three *excerpts* from the category line generated to for the entire 10-minutes long discussion including the contributions shown in the reference transcript.



The *first* excerpt shows many metacognitive and discursive activities on the students' side; due to the absence of codes on the left side one knows that they are carried out without a demand given by the teacher. This could be interpreted as a great potential for understanding the mathematical issues discussed in this lesson. The *second* piece shows metacognitive activities combined with negative discursive actions. The teacher's monitoring activity M4 is not coherent, and it does not explain the understanding difficulties expressed by the students (ND3b). Two students' activities do not precisely refer to the structure of the given equation (ND3a), and include incorrect vocabulary (ND2) hindering the understanding of what is meant. The learners try to explain there the sense and the validity of the ET suggested by Thomas (R1, R4, *r*R6b). Due to the negative discursivity their efforts do not contribute to clear the problematic issue. The *third* piece shows many codes for negative discursivity indicating that the teacher does not

[4] The further developed version of this category system can be found here: http://www.mathematik.uni-osnabrueck.de/fileadmin/didaktik/Projekte_KM/Kategoriensystem_EN.pdf

consider students' difficulties when explaining the equation (ND4, ND3b). The *missing* reflective and discursive intervention on his side hinders the understanding of the mathematical activities (solving equations) discusses in the class. The local *absence* of metacognition and discursivity has to be considered in the evaluation of MDQ to make this evaluation accurate, and to draw valid conclusions from it. This requires a high degree of the necessary analytical reasoning. Taking into consideration only the observable fact that the learners are quite autonomous in practicing planning, monitoring and reflection would hinder the validity of the intended evaluation and its explanatory value with regard to the expected students' mathematical understanding (like in the study in Depaepe et al., 2010).

The second step: global assessment of MDQ of a classroom discourse

The metacognitive-discursive instructional quality is evaluated by means of seven high inference rating scales. Each of them consists of a guiding question (*GQ*) focusing rater's attention on aspects to be analysed and evaluated complementary, and of several answers to the *GQ* describing in detail how these aspects are reflected in the discussion. Different answers describe qualitatively different situations. *Their order reflects the increasing quality of the discourse with regard to the relevant aspects*. The rater has to choose the one that best describes the given situation, and to justify the choice. In the following, the guiding questions will be explained. To each of them the answer for the entire 10-minutes long discourse including the discussion in the reference transcript will be given in brackets, written in *italics*, e.g.: (*Answer no. 1 out of 4*).

The *first GQ* focuses on teacher-students interactions and their potential to facilitate learners' autonomy in practicing metacognition. There are four answers to it. The answer no. 1 describes the case that metacognitive activities are carried out almost exclusively be the teacher, and it cannot be indicated that the teacher is aimed at fostering metacognitive skills by the learners, or – alternatively – only a few metacognitive activities are practiced by the teacher and by the students, and no effort is made to use these activities accurately to explain the mathematical subject of the discussion. In the case explained in answer no. 4 the learners are autonomous in practicing and regulating metacognition, and they make effort to precisely elaborate the subject of the discussion. (*Answer no. 4 out of 4*)

The *second GQ* focuses on justifications combined with metacognitive activities, on efforts made with them to explain and understand the mathematical subject discussed in class, and on promoting learners' autonomy in justifying. It should be assessed whether justifications are practiced and valued as being important in the culture established in the class. There are four answers, analogue to these to the first *GQ*. To choose the right answer one has to take into consideration the extent to which the teacher and the learners really make efforts to precisely explain and elaborate the mathematical content. Hereby it is important to distinguish between the syntactic form of a justification and the content of it. An utterance with the formal form of a justification does not necessarily have any relevant explanatory content in the given context. Such utterances are called ‘pseudo justifications’. Situations with a high number of such ‘justifications’ left without critical comments and corrections hinder the development of the reasoning skills of the learners. (*Answer no. 2 out of 4*)

An accurate evaluation of MDQ has to differentiate between lots of disconnected teacher’s and students’ metacognitive and discursive activities and an orchestrated discourse producing accurate explanations and justifications for mathematical issues. The *third GQ* focuses on the interplay of the metacognitive and discursive activities carried out, on their potential for understanding the subject-specific issues discussed in the class (questions, tools, methods, argumentations) and for organising and systematising mathematical knowledge in students’ minds. The first answer refers to a class discussion without any productive use of metacognitive and discursive activities. The second describes the case that the understanding can only be indicated by an individual learner. The third refers to the case that the interplay of these activities contributes to a deep understanding in the class. (*Answer no. 2 out of 3*)

Discursivity is in the focus of the *fourth GQ*. It evaluates to what extend the class discussion integrates learners’ ways of thinking, and is aimed at making students’ and teacher’s utterances comprehensible for others and accessible for further analysis regarding individual ways of thinking and reasoning, or differences between what was said or written and what was meant by that. (*Answer no. 1 out of 5*)

The *fifth GQ* deals with negative discursivity and with efforts made to prevent it. The answers to this GQ describe the extent to which negative discursivity hinders the reciprocal understanding in a class and the understanding of the subject-specific issues (tasks, tools, methods or ways of reasoning). (*Answer no. 1 out of 5*)

The *sixth GQ* focuses on stringently guided discourse units called ‘debates’. The answers to this GQ vary between situations without any (even short) debate, and between situations with at least one long debate guided by the learners and characterized by the use of discursive and metacognitive activities with justifications. The other two middle answers refer to situations with only short and not elaborated debates guided by the learners or to situations with a longer debate guided by the teacher. (*Answer no. 1 out of 4*)

The quality of the classroom discussion can change dramatically if a challenging and complex meta-mathematical issue is being discussed. Such discussion requires the use of elaborate metacognitive and discursive activities, and the inclusion of a meta-knowledge with regard to the subject matter. The *seventh GQ* focuses on situations with challenging and complex issues, and on the efforts made by the teacher and by the students to orchestrate their utterances, arguments, ideas and conceptions into a coherent discourse unit. Since such complex issues are rarely discussed in math instruction, this GQ plays an essential role in a long-term evaluation of MDQ in an individual class. The answers vary from the case without complex issues or with an 'intellectual chaos' when discussing such issues, to situations with noticeable efforts made by the teacher or by the learners to find appropriate methods and ways of reasoning to elaborate the given issue. (*Answer no. 2 out of 4*)

DISCUSSION

The evaluation of the 10-minutes long discussion including the reference transcript leads to the following assessment. For the first *GQ* the highest answer (no. 4 out of 4) has to be chosen because the learners are autonomous in practicing metacognition, and make efforts to understand the mathematical activity (solving equations) being the core issue of the lesson. The category line showing many codes for students' metacognitive activities supports this assessment. Nonetheless, this remarkable *observable* characteristic of the class discussion does not automatically lead to a high quality of other aspects of MDQ being substantially relevant for understanding the discussed mathematical tools and formal representations. For the second *GQ* answer no. 2 (out of 4) has to be chosen: there are only a few mathematical justifications and they play no relevant role in the discussion. For the fourth and fifth *GQ* answers no. 1 have to be chosen: the discussion does not respect students' difficulties, questions and ways of thinking. Consequently, answer no. 2 (out of 3) for the third *GQ* states that no relevant understanding processes have been initiated in the class. Such processes can only be indicated in the case of one single student. And, furthermore, due to the intellectual chaos and the absence of any attempts to clarify the reasons of the fundamental understanding difficulties externalised in students' critical remarks and questions, the rater has to choose the answer no. 2 (out of 4) for the seventh *GQ*.

This evaluation leads to the following conclusion. A complex analysis was needed to describe and evaluate the metacognitive and discursive activities in the discussion including the reference example. The question whether the learners are autonomous in practicing metacognition had to be split from the question whether their activities promote an understanding learning process. Also the question whether the learners take the responsibility for managing the discussion and the use of mathematical tools had to be split from the question whether the discussion produces a coherent mathematical argumentation. Furthermore, it was substantially crucial to split the question whether the teacher allows the students to be responsible for solving the equation from the question whether he acts co-responsible for the quality and results of students' activities. None of the separated questions can give an accurate and valid evaluation of the MDQ. They have to be analysed and evaluated complementary.

The implementation of the two-step rating procedure – in which the rater first locally and precisely analyses metacognitive and discursive activities, and in which he gives the global evaluation of MDQ immediately after a sophisticated interpretation of the results from the first step – is a promising design to fulfil the requirements of reliable and valid assessments. It also proved as an effective method in educating raters and pre-service teachers in conducting a detailed discourse analysis, and in analysing strong and weak aspects of observable sight structures and of deep structures of teaching-learning situations. This opens new possibilities for research aiming at implementing metacognition in classes, and in enhancing its effectiveness in promoting students' understanding in learning mathematics.

REFERENCES

Cohors-Fresenborg, E., & Kaune, C. (2007). Modelling Classroom Discussions and Categorising Discursive and Metacognitive Activities. In D. Pitta-Pantazi, & G. Pilippou (Eds.), *Proc. of CERME 5* (pp. 1180–1189). Cyprus: University of Cyprus.

Cohors-Fresenborg, E., Kaune, C., & Zülsdorf-Kersting, M. (2014). *Klassifikation von metakognitiven und diskursiven Aktivitäten im Mathematik- und Geschichts-unterricht mit einem gemeinsamen Kategoriensystem*. Osnabrück: FMD.

Cronbach, L. J., et al. (1972). *The dependability of behavioral measurements: Theory of generalizability scores and profiles*. New York: Wiley.

Depaepe, F., De Corte, E., & Verschaffel, L. (2010). Teachers' metacognitive and heuristic approaches to word problem solving: analysis and impact on students'' beliefs and performance. *ZDM – The International Journal on Mathematics Education*, *42*(2), 205–218.

Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Mevarech, Z., & Kramarski, B. (2014). *Critical Maths for Innovative Societies: The Role of Metacognitive Pedagogies*. OECD Publishing.

Nowińska, E. (2011). A study of the differences between the surface and the deep structures of math lessons. In M. Pytlak, T.Rowland, & E. Swoboda (Eds.), *Proc. of CERME 7*. Rzeszów: University of Rzeszów.

Nowińska, E. (2016). *Leitfragen zur Beurteilung metakognitiv-diskursiver Unterrichtsqualität*. Osnabrück: FMD.

Praetorius, A.-K., Lenske, G., & Helmke, A. (2012). Observer ratings of instructional quality: Do they fulfill what they promise? *Learning and Instruction*, 22(6), 387-400.

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*(1), 3–14.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What influences learning? A content analysis of review literature. *Journal of Educational Research*, 84, 30–43.