

THE QUALITY OF MATHEMATICAL PROBLEMS – EVALUATION AND SELF-EVALUATION

Eva Patáková

Charles University in Prague
eva.patakova@email.cz

Abstract

The research presented in the article consists of two parts. Firstly, opinions on mathematical problem quality are explored within four groups of participants (novices, specialists and experts in problem posing; high school students who never posed their own problems). Secondly, self-reflections written by the participants who have some experience in problem posing (novices, specialists and experts) are explored and compared with the general view of problem quality received in the first part of the research.

The more experienced problem posers have more requirements on problem quality (both as general requirements and within their own work on posing problems). There is a slight decrease in ability to notice important features of mathematical problem quality after the first experience in problem posing. Experts lay stress on mathematical features of the problem whilst novices and specialists more on problem – student interaction.

Key Words

Quality of mathematical problems, self-evaluative comments, problem posing, expert – specialist – novice comparison

ARTICLE INFO

Article type

Full research paper

doi: 10.7160/eriesj.2013.060302

Article history

Received: June 8, 2013

Received in revised form: July 7, 2013

Accepted: August 27, 2013

Available on-line: September 30, 2013

Introduction

In this article, the term “mathematical problem” means a word problem that is more difficult and more elaborated than a common exercise. Stehlíková (2000) says that the exercise becomes a problem if the solver does not immediately see the solving strategy and he / she has to search for it.

The presented study is a part of research on the comparison of problem posing process by problem posers on various experience levels. Problem posing is one of the developing fields in mathematics education. In this work, problem posing by various groups of professionals is explored.

A model of problem posing process of “skilled problem posers¹” and novices has been made by Pelczer and Gamboa (2009). It includes the phases of Setup, Transformation, Formulation, Evaluation and Final assessment. The novices usually use a linear problem posing model (e.g. Setup – Formulation – Evaluation). Not all five phases of problem posing process are usually present. Mostly, there are nearly no transformations. (If a novice finds out that his / her suggested problem is not good, he / she does not transform it but drops it completely and starts from the beginning.) On the contrary, the model of problem posing process by skilled problem posers is cyclic. A skilled problem poser moves through various stages, transforms and sometimes performs the Final assessment stage.

The wider research which is the source of presented study (e.g. Patáková, 2013b) deals with three categories of respondents:

1. **Novices:** Participants with nearly no experience in problem posing.

¹ These are called “Experts” in the article by Pelczer and Gamboa. The term is changed in this study because “experts” have a different meaning there.

2. **Specialists:** Participants – mostly lower and upper secondary school teachers – who pose problems but not more than their teaching profession requires.
3. **Experts:** Very skilled problem posers – e.g. authors of problems for mathematical competitions, textbook authors, ...

The three groups were examined according to *idea types* used during their problem posing (Patáková, 2013b). Experts were found to perform most intentional ideas from the three groups. The intentional idea is such that the problem poser is fully aware of its consequences, it is completely goal orientated. Usually it means that some backward computation of conditions is necessary so that the problem completely fulfils the author’s goals. Specialists performed fewer intentional ideas than experts, participants from the novice’s category did not use this idea type at all.

Problem quality is a subjective concept that cannot be measured in an objective way. However, it is possible to look into the opinions on problem quality. For example, Tarhan et al. (2008) investigated opinions of 9th grade students while exploring problem-based learning in chemistry classes – within a questionnaire the students had to answer what quality a good problem should fulfil. Four dominant topics important for students as problem quality occurred: “The problem should be related to our prior knowledge. The problem should have some leading questions. The problem should be related to the daily life. Problem and questions should be clearly stated.” (Tarhan et al., 2008: 296)

The goal of the first part of the research is to explore the understanding of the quality of mathematical problems by four groups of participants – experts, specialists, novices, and secondary school students with no experience of problem

solving. (The first part of the research has already been described in Patáková, 2013a.) The second part of the research looks into the problem posing process of the first three groups; occurrence of application of problem quality criteria is explored. The research questions are namely:

1. What are the most important signs of “mathematical problem quality” for the participants?
2. What differences (described in a qualitative way) are there in “mathematical problem quality” opinions of the four groups (experts, specialists, novices, high school students)?
3. How are the opinions on “mathematical problem quality” applied during the problem posing process of the three groups (experts, specialists, novices)?

Material and Methods

Overall 106 participants took part in the study: 21 experts, 17 specialists and 23 novices; and in the follow-up study 45 high school students who never posed their own mathematical problems.

The novices – first year university students of mathematics education – and high school students took part in the research compulsorily; the participation of experts and specialists was voluntary.

In the *first part* of the research the participants were asked to write a short essay on their opinion on mathematical word problem quality.

The written essays were processed in Atlas.ti software and methods derived from the grounded theory were used for their interpretation. The first coding process was made without any abstraction (in vivo coding). There was no grouping of similar statements – the quotations were coded equally in case

they had the same meaning. 45 codes were obtained in this way (e.g. “ ‘Nice’ numbers in problem solution”, “Presence of propedeutics”, “Adequate difficulty”, ...). A first comparison was performed and codes related to each other were merged (these are called “sub-codes” in the following text). There were 16 sub-codes (e.g. “Benefit for the student”, “Adequacy”, “Inventiveness”, ...). After the second comparison and second code merging, 5 codes were obtained for the last analysis: “Problem assignment”, “Mathematical features”, “Motivational strength”, “Student”, “Comfort”:

1. **Problem assignment:** The code covers all assignment requirements. These are e.g. the assignment length, topic originality, intelligibility, unambiguous assignment, ...
2. **Mathematical features:** In most cases quotations coded as “Mathematical features” mean requirements on the process of problem solution. The participants want the problem to require non-standard solving methods, to be solvable by more than one method, ...
3. **Motivational strength:** The quotations coded as “Motivational strength” concern both the effect of the problem on a student and the features of the problem itself. It may concern the attraction of the problem, satisfaction felt by the student who solved the problem successfully, the surprising result, ... This code goes more into the mathematical attractiveness than into the attractiveness of the context.
4. **Student:** All quotations concerning a student, with the exception of motivational strength, are coded as “Student”. It means all requirements on problem difficulty, adequacy

of problem goals and problem purpose, effect on students, development of key competences of students², ...

5. **Comfort:** The code “Comfort” covers statements about practical use of a problem and its context. The problem should be applicable without any adaptations and using the problem should be comfortable. The main sub-code is “Correctness” which regards mistakes in problem assignment, in the author’s solution, in the non-mathematical context, ... The other sub-codes of Comfort concern e.g. the presence of an intermediate result, preferences of concrete topics, “nice” numbers, ...

A small follow-up study was carried out after the main study. When the data for the three groups of participants (experts, specialists, novices) had been processed, the same task – to write a short essay on their view of mathematical word problem quality – was given to 45 high school students. Their essays were processed by the code system gained during the main study phase.

The *second part* of the research was based on analysis of the written self-reflections of the whole problem posing process of the participants. The participants from the groups of experts, specialists and novices were asked to pose a difficult, interesting and original mathematical problem for approximately 15-year old students and to write a detailed self-reflection of their problem posing process. They were not told what exactly would be explored in the self-reflections. The topic of the problem was not given – Stoyanova and Ellerton (1996) call this “free problem posing situation”. (Here is impossible to repeat the follow-up

² The term “key competences of students” is one of the main terms used in recent Czech school reform (Kotásek et. al 2001) so it is no wonder that it appears in the essays frequently.

study with students not posing problems. If the task to pose a problem and write a self-reflection is given to them, they shift to the group of novices in problem posing.)

The opinions on the quality of mathematical problems are thus explored from a different point of view. Self-evaluative comments are searched for – i.e. notes in the self-reflections where the poser evaluates his / her steps and ideas in the problem posing process. So self-evaluative comments express some application of the general opinions on the quality of Mathematical problems – this application is made by the respondents within the context of the own problem posing process.

The example of self-evaluative comments:

*... So the question on the whole amount of divisors can be asked. ... I made a computation and found three solutions. ... **This is lovely but this too difficult for 15-year old students.** What to do about this? If the numbers are 7 and 9, ... ?*

The type and frequency of self-evaluative comments present in the reflections are observed. The comments were coded by 5 codes obtained during the first research phase. The three groups of respondents were compared again. Differences between general opinions on problem quality of the participants and their practical applications were described as well.

Results

Results of the first phase will be first presented in terms of the group of participants and next in terms of individual codes.

General view of the quality of problems by experts, specialists and novices

21 **experts** participated in the study. There are 88 coded quotations about opinions on quality of mathematical problems in their essays, which means 4.2 coded quotations per person, see Fig. 1.

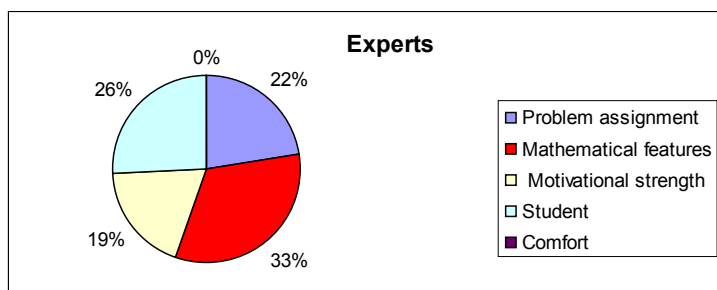


Fig. 1: Requirements on problem quality by experts³

The dominant topic for the experts is “Mathematical features of the problem”. Their problem quality requirements are both general and quite specific. Some examples:

The solution requires non-trivial ideas.

There is no possibility to avoid the supposed solution method by routine and not interesting testing of all possibilities.

The problem enables more than one solving method.

Solving of the problem requires deep understanding of the concepts used.

The story of the word problem does not bring factors enabling us to reject some intermediate results, which can be rejected in a clear

³ The graphs used in the article serve just to describe the research sample, they are not meant to express a general statement.

mathematical way as well. (E.g. ‘I reject the alternative that the width of the allotment is 7 meters because such an allotment would be impractically narrow.’)

The experts’ essays did not contain any quotations regarding “Comfort”. The question is why none of them mentioned correctness of the author’s solution. There are two possible explanations. The experts are likely to take it for granted or they do not perceive the author’s solution to belong to the characteristics of the problem.

17 **specialists** participated in the study. There are 62 coded quotations about opinions on quality of mathematical problems in their essays, which means 3.6 coded quotations per person, see Fig. 2.

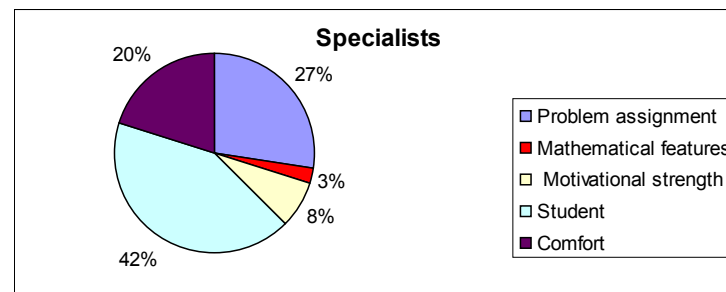


Fig. 2: Requirements on problem quality by specialists

As expected, the dominant feature of problem quality for specialists is “Student”. At least one quotation by a specialist is present in every sub-code forming the whole code “Student”. The quotations concern difficulty of the problem for students, adequacy of the problem for its purpose, adequacy of solving time, development of key competences, benefit of the problem for students. Some examples:

Not entirely easy solution.

The problem motivates a student to search for further knowledge; it cultivates his / her spoken and written language.

Adequacy for the students who should solve it.

To enrich the skills.

Links to further topics that will be taught in the class.

The problem may provide some added value for a student.

On the contrary, “Mathematical features” of the problem are mentioned only by two specialist participants – both of them want the problem to come up with something new.

23 **novices** participated in the research. There are 37 coded quotations about opinions on quality of mathematical problems in their essays, which means 1.6 coded quotations per person, see Fig. 3.

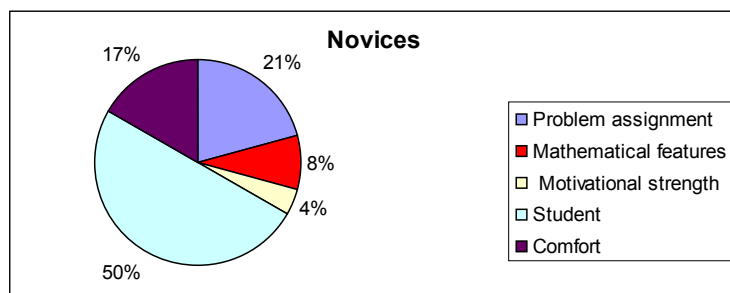


Fig. 3: Requirements on problem quality by novices

The amount of coded quotations per person is worth noticing. If compared with the other groups, this number is very low. The essays of novices were usually very short – mostly containing

only one or two coded quotations. The most frequent code is “Student” – 9 of the total 37 novices’ quotations were marked by “Not too easy” sub-code of the code Student. Some examples:

The problem is not too easy.

The problem forces the solver to think.

The problem requires some input knowledge.

The least mentioned code by the novices is “Motivational strength” – only two novice participants mentioned it.

Codes distribution in essays

The code “**Problem assignment**” contains quotations from all three groups of participants. However, quotations are distributed unevenly within sub-codes “Elegance” and “Intelligibility, Unambiguous assignment”. “Elegance” means clarity, brevity, accuracy of formulations, ... (e.g. *Clarity of problem assessment., The assessment isn't complicated whilst the solution is easy.*) Most of the quotations coded as “Elegance” belong to experts (9 from 12). On contrary the sub-code “Intelligibility, Unambiguous assignment” was mostly related to quotations of specialists (14 from 21).

The majority of quotations coded as “**Mathematical features**” belong to experts (29 from 33). Sub-codes “Non-standard solving methods” and “Types of mathematical imperfections” consist of quotations written by experts only. There are some quotations by specialists and novices within sub-codes “Inventiveness” and “Problem solving process” but they also mostly belong to experts.

The topic “**Motivational strength**” is mentioned mostly by experts as well (17 from 23 quotations). Sub-codes “Attractiveness of problem assessment” and “Pleasure in solving” belong to experts only.

“Student” is a frequent topic for all three groups of participants. Some disproportions can be found within sub-codes “Not too easy” (9 quotations from 15 by novices), “Forces the student to think” (8 quotations from 12 by specialists), “Competences development” (6 quotations from 8 by specialists) and “Added value for the student” (7 quotations from 8 by experts).

As mentioned above, no quotation by experts regards “Comfort”. Some disproportion among specialists and novices can only be found within the sub-code “Correctness” (8 quotations from 9 by specialists).

Follow-up study with high school students

45 high school students (17–19-year old) participated in the study. There are 102 coded quotations about opinions on quality of mathematical problems in their essays, which means 2.3 coded quotations per person, see Fig. 4.

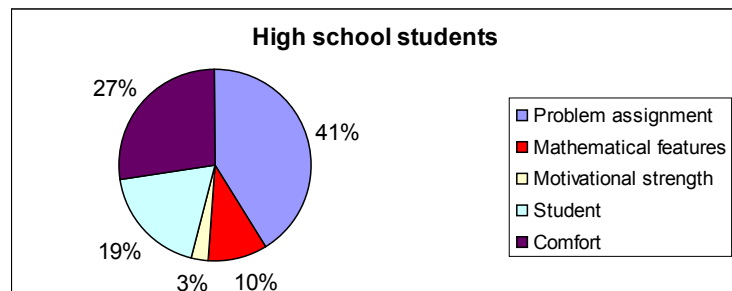


Fig. 4: Requirements on problem quality by high school students not posing problems

The dominant feature of problem quality for high school participants is “Problem assignment” which comes from a single sub-code “Intelligibility, Unambiguous assignment”.

This sub-code is the most frequent one in high school students’ essays (40 from 42 quotations coded as Problem assignment). This was mentioned by 34 from 45 high school participants. (6 of them made requirements on both parts of this sub-code – intelligibility and unambiguous assignment so they contributed twice to this sub-code.)

The second most frequented code was “Comfort” – the students very often required the problem to be pleasantly solvable for them. Whilst the other groups of respondents – in case they mentioned this aspect – want the problem to be original and non-routine, the students often want the opposite. Especially students with lower marks want the problem to be textbook-like, similar to the problem they know how to solve.

Requirements on specific topics which should / should not be included in the problem and requirements on daily-life situations are often present as well. Some examples coded as “Comfort”:

Daily-life situation, real context is more attractive for students.

Not off the topic discussed within the lessons.

It would be the best if there was a solution – both the solution method and the result for my control.

If it is clear what is to be computed.

The problems should ideally be posed by our teacher – such problems are optimal for the class and there are no problems with topics unknown to students.

Self-evaluative comments applied in the self-reflections

In the second part of the research self-evaluative comments were searched for in the self-reflections of the problem posing process. Thus we shift from the level of general requirements to the level of their applications. The respondents evaluated their

steps and intermediate results during problem posing and they used these evaluations to improve their problem. The graphs show the structure of self-evaluative comments in the self-reflections within the three groups of participants.

21 **experts** participated in the study. There are 118 coded self-evaluative comments in their self-reflections, which means 5.6 coded quotations per person, see Fig. 5.

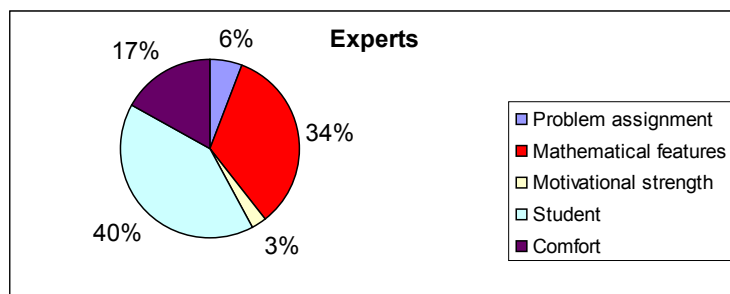


Fig. 5: Self-evaluative comments by experts

The majority of self-evaluative comments by experts regards the code “Student”. The reason is that the evaluation of the difficulty of the problem and its suitability for the student appears frequently. There is a big ratio of quotations coded as “Mathematical features” again. Just as in general requirements, the experts lay stress on existence of different methods to solve the problem, non-routine solving methods etc. in their work. The quotations coded as “Comfort” occur as well – e.g. evaluation of the context of the problem. (I.e. the given numbers should not be in contradiction to reality ...) Some examples:

It is not good, the solution isn't elegant and there is no need to think much to find the idea to solve it. (Mathematical features)

The solution is fine – i.e. the numbers are not small enough to be easy to be guessed. (Mathematical features)

So... Why had the cars gone with the same speed before? Well, this is weird. I'll change it somehow later... (Comfort)

17 **specialists** participated in the study. There are 57 coded self-evaluative comments in their self-reflections, which means 3.4 coded quotations per person, see Fig. 6.

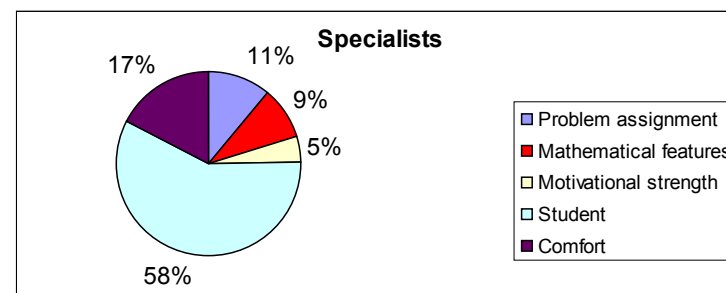


Fig. 6: Self-evaluative comments by specialists

The majority of self-evaluative comments of specialists regard the code “Student” where the dominant topics are the difficulty of the problem for students and the evaluation of the problem with respect to what the student should know (knowledge, abilities, competences). Some examples:

I must say I started to like my problem because of its complexity and simplicity at the same time. (Mathematical features)

It requires the geometrical insight, knowledge of the properties of geometrical shapes and mainly the ability to divide the problem into small tasks leading to the solution of the whole problem. (Student)

As for me it is not too difficult intellectually for children, rather it will demand concentration, patience and accuracy during the work and knowledge of what can be factored. (Student)

23 novices participated in the study. There are 34 coded self-evaluative comments in their self-reflections, which means 1.5 coded quotations per person, see Fig. 7.

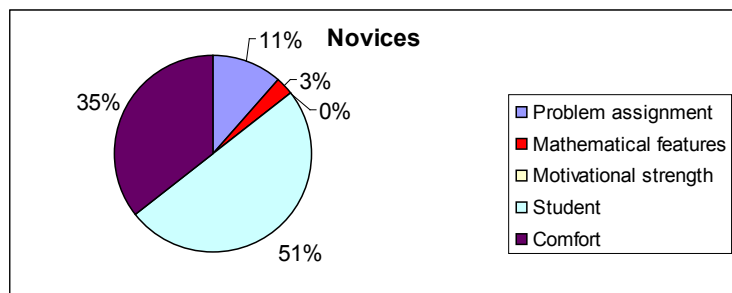


Fig. 6: Self-evaluative comments by novices

The dominant topic for the novices is “Student” where the main sub-topic is the evaluation of the problem difficulty again. Some examples:

But this is not enough to form a difficult problem. (Student)

But still I considered my problem to be too easy. (Student)

... 5 and 7 are better numbers – the computation will be more pleasant. (Comfort)

Discussion

The results – not surprisingly – show that the experts who participated in the study have the most complex set of general requirements on problem quality. Their essays were the longest and contained most coded quotations when compared with the other groups (4.2 coded quotations per person on average). Specialists have also quite a complex set of requirements on problem quality. The amount of their coded quotations is a bit lower (3.6 coded quotations per person). Their essays are shorter than the experts’ ones. Experts usually convey their statements in detail whilst specialists are often more concise. (Sometimes there are one-word quotations only – e.g. “Correctness”.) The least requirements on problem quality are observed by novices (1.6 coded quotations per person). This result comes from the fact that within creative work on problem posing it is necessary to think over problems (and their quality) in detail. As Zhouf (2010) states, an expert usually establishes his / her own requirements on problem quality and on this basis he / she decides about the future use of his / her posed problem. Specialists sometimes pose problems themselves and moreover they work with problems actively by selecting problems for students and working with them during the lessons. Novices usually meet the problems “only” as problem solvers which is a less active role in comparison with the other groups.

The same trend can be found within self-evaluative comments. Experts show the biggest amount of self-evaluative comments as well (5.6 coded quotations per person), lesser amount can be observed by specialists (3.4 coded quotations per person), the least amount was found in the self-reflections of novices (1.5 coded quotations per person). Again the participants show the trend that the more experienced problem poser uses more criteria on problem quality which can be found both in the

general essays and in self-evaluative comments of their own work. The last two mentioned results correspond with each other.

One fourth of novices' coded quotations in the general essays regard the opinion that the problem should not be too easy. The reason is probably that the participants wrote their essays immediately after they tried to pose their own problem. Here they found that if they try to formulate a problem which just crossed their mind, the problem is often trivial. Thus they consider posing a non-trivial problem to be "the art" and their statements about problem quality are the natural reaction on their own problems which they still remembered and which they rejected as poor-quality ones. This explanation is based on several interviews with the participants from the group of novices.

In the general essays of high school students not usually posing problems, the quotations about topics described by Tarhan et al. (2008, see Introduction) are quite frequent. If we consider the findings by Tarhan et al. within the codes introduced in this study, all of them would be coded as "Comfort" or "Problem assignment" (sub-code "Intelligibility, Unambiguous assignment") which corresponds with the findings of the presented study completely. Requirements namely on daily-life situations (coded as "Comfort" in this study) are frequent in another literature as well – e.g. Zhou (2012) states that students prefer real life engineering problems compared to hypothetical, academic problems.

An interesting result is that high school students not posing problems introduce more requirements on problem quality than novices in problem posing (high school students have 2.3 whilst novices only 1.6 coded quotations per person). A similar trend is showed in a different context in Vondrová and Žalská (2012).

They explored the ability of students to notice mathematics specific phenomena when observing mathematics teaching on video. Their interest lies in the amount of mathematics specific phenomena noticed and described by the students in their written reflections. Two groups of university students were explored: students before and after their compulsory pedagogical practice and mathematics education courses. Quite surprisingly a slight decrease in the "ability to notice" after completing pedagogical practice was found. The presented study shows the same phenomenon: the students who completed "their practice" – which means problem posing experience here – showed lower "ability to notice" important features of high-quality problems. The reason seems to be similar as Vondrová and Žalská state. One starts to be less critical when he / she finds what the activity – whether teaching practice or problem posing – really involves. Another reason could come from the differences between high school and university classes. (Participants not posing problems were from high schools, novices in problem posing were first-year university students.) The problems and exercises usually solved at the university are more difficult than high school ones but on the other hand they are rarely word problems.

Looking in self-evaluative comments within the self-reflections, all the three groups lay most stress on "Student". It is obvious that within their work problem posers frequently watch the difficulty of the posed problem. Though the ratio of "Mathematical features" quotations by experts remains high as well, the other groups do not accent it much. Within the code "Comfort" all the three groups watch the adequacy of the context and the "nice numbers" more than they did in general essays. This comes from the situation. They just describe the process of problem posing – it means also posing the context so they

are forced to work with it and it often inspires them to make a self-evaluative comment. The ratio of “Motivational strength” is lower within all the three groups than it was in general essays⁴. The reason is probably that the big part of attractiveness of the problem’s mathematical content lays in the good choice of the topic which can be influenced only a little later.

The above findings on problem quality requirements coming from both research phases correspond with the model of Pelzer and Gamboa (2009). In their model, problem posers are grouped as novices and “skilled problem posers”. Using the terminology of this study, novices remain the same and skilled problem posers correspond with specialists and experts together. Novices usually pose problems according to a linear model whilst skilled problem posers follow the cyclic one. The research presented in this study shows that skilled problem posers have a more complex set of requirements on problem quality. If they want to fulfil all of them, it is natural that they have to improve their problem, to enrich it and to adapt it many times – these are typical signs of the cyclic model of problem posing. To fulfil fewer criteria it is enough to elaborate an initial idea directly – without any corrections – just implying one or two criteria. If the initial idea does not prove to be suitable to fulfil these criteria, the poser drops it and looks for another initial idea. These are typical signs of a linear problem posing process described by Pelczer and Gamboa (2009).

The findings correspond with my previous study (Patáková, 2013b) as well. Experts were found to perform most intentional ideas. This is in agreement with the finding that experts have

the most complex set of requirements on problem quality. They know exactly how their problem should look – and intentional ideas help them to fulfil the goal.

Conclusions

The experts are the group who deals most with “Mathematical features” from the explored groups. (They have a lot of requirements on assessment elegance, problem solving process, attractiveness of the problem for students, new pieces of knowledge and new views of the mathematical topics for students, adequate problem difficulty, ...) They have most requirements on problem quality from the explored groups.

Specialists emphasise more the effect of a problem on a student and the practical use of a problem. (The problem should be correct, force students to think, develop important skills and competences.) They have quite a complex set of problem quality requirements as well.

Novices introduce quite a narrow set of requirements on problem quality related mostly with their first experience on posing non-trivial problems with a dominant need to pose a problem which is not too easy for its solver.

High school students not posing problems look at the thing from another point of view. Mostly they do not think about problem quality in general but about himself / herself working with the problem. (How the problem should look to be pleasant for them to solve it.) They have more general requirements on mathematical problem quality than novices in problem posing. The study will be deepened and connected to the findings from the wider research.

⁴ This is not processed statistically because the data from the two phases does not have exactly the same character.

Acknowledgements

The research is supported by grant GAUK 303511.

References

Kotásek, J. et al. (2012) *National programme for the development of education in the Czech Republic: White paper*, Prague: Tauris.

Patáková, E. (2013a) 'The quality of mathematical problems – how we evaluate it'. In Kvasnička, R. (Ed.), *Proceedings of the 10th International Conference Efficiency and Responsibility in Education*, Prague: Czech University of Life Sciences, pp. 488 – 496.

Patáková, E. (2013b) 'Teachers' Problem Posing in Mathematics', *Procedia – Social and Behavioral Sciences*, Elsevier Ltd, Forthcoming.

Pelczer, I. and Gamboa, F. (2009) 'Problem Posing: Comparison between Experts and Novices', *Proceedings of the 18th International Conference of the International Group for the Psychology of Mathematics Education*, Atlanta, pp. 353 - 360.

Stehlíková, N. (2000) 'Analýza písemného řešení žáka, jedna z možných technologií'. In Novotná, J. (Ed.), *Analýza řešení slovních úloh*. Prague: PedF UK, pp. 98 - 117.

Stoyanova, E. and Ellerton, N. F. (1996). 'A framework for research into students' problem posing'. In P. Clarkson (Ed.), *Technology in Mathematics Education*. Melbourne: Mathematics Education Research Group of Australasia, pp. 518–525.

Tarhan, L., Ayar-Kayali, H., Urek, R., and Acar, B. (2008) 'Problem-based learning in 9th grade chemistry class: Intermolecular forces', *Research in Science Education*, vol. 38, no. 3, pp. 285–300.

Vondrová, N. and Žalská, J. (2012) 'Do student teachers attend to mathematics specific phenomena when observing mathematics teaching on video?', *Orbis scholae*, vol. 6, no. 2, pp. 85–101.

Zhou, Ch. (2012) 'Teaching Engineering Students Creativity: A Review of Applied Strategies', *Journal on Efficiency and Responsibility in Education and Science*, vol. 5, no. 2, pp. 99–114.

Zhouf, J. (2010) *Tvorba matematických problémů pro talentované žáky*, Prague: PedF UK.