Full research paper

A JOINT ASSESSMENT OF REASONING ABOUT GENERAL STATEMENTS IN MATHEMATICS AND BIOLOGY

Libuše Samková[1]✉
Lukáš Rokos[2]
Lukáš Vízek[3]

[1]Department of Mathematics, Faculty of Education, University of South Bohemia in České Budějovice, Czech Republic

[2]Department of Biology, Faculty of Education, University of South Bohemia in České Budějovice, Czech Republic

[3]Department of Mathematics, Faculty of Science, University of Hradec Králové, Czech Republic

✉ lsamkova@pf.jcu.cz

ABSTRACT

This contribution belongs to a larger empirical study that focuses on issues related to the implementation of inquiry-based learning and formative assessment in science and mathematics education, while it also refers to the issue of STEM education. Here, we discuss the two topics from the perspective of professional preparation of primary school teachers. We employ an educational tool called Concept Cartoons and perceive it as a common diagnostic tool for investigating modes of reasoning about general statements in arithmetic, geometry and biology. The presented qualitative exploratory empirical study maps and codes various kinds of reasoning that can be identified with the tool and investigates possibilities of a joint coding procedure. As a result, it provides a conversion table between various modes of reasoning in the three subject domains. The arisen code categories cover the field of generic examples, including the initial stages so that they can be used for scaffolding the process of learning the foundations of deductive reasoning. The joint approach to reasoning in mathematics and biology shows how argumentation and formative assessment can be understood equally and developed simultaneously in both school subjects. It helps us to see how the two school subjects can be integrated didactically.

KEYWORDS

Argumentation, biology education, Concept Cartoons, formative assessment, future primary school teachers, mathematics education

HOW TO CITE

Samková L., Rokos L., Vízek L. (2021) 'A Joint Assessment of Reasoning about General Statements in Mathematics and Biology', *Journal on Efficiency and Responsibility in Education and Science*, vol. 14, no. 4, pp. 270-287. http://dx.doi.org/10.7160/eriesj.2021.140406

Article history
Received
June 4, 2021
Received in revised form
August 3, 2021
Accepted
September 1, 2021
Available on-line
December 20, 2021

Highlights

- *Establishing a framework for joint assessment of reasoning about general statements in arithmetic, geometry and biology, with a conversion table.*
- *Providing a formative assessment background for scaffolding the process of learning the foundations of deductive reasoning.*
- *Promoting opportunities for integrating mathematics and biology education within professional preparation of primary school teachers.*

INTRODUCTION

In recent decades, inquiry-based teaching and formative assessment have belonged among educational frameworks that have been discussed in relation to possible enhancement of the form and outcomes of teaching practices. Inquiry-based teaching has been considered as a means of providing a sustainable connection between up-to-date scientific research and everyday school practice (McComas, 2002; Minner, Levy and Century, 2010), and formative assessment has proved its ability to provide a suitable evaluative approach to student outcomes and development of student learning during inquiry-based school activities (Dolin and Evans, 2018). Due to some particular timing of the initial recent interest in inquiry-based pedagogy (the first US and European curricular documents involving inquiry referred only to the school subject of science: National Research Council, 1996; Rocard et al., 2007), the inquiry has often been perceived solely in relation to science teaching and learning. However, the inquiry can be implemented in various school subjects including mathematics (Artigue and Blomhøj, 2013; Dorier and Maass, 2014). In our paper, we present an approach that goes beyond the initial framework of science education and perceives inquiry in a comparable manner both in science and mathematics school subjects.

In recent years, we have been analysing an educational tool called *Concept Cartoons* (Keogh and Naylor, 1999) and its possible use in the professional preparation of future teachers, especially in investigating various aspects of teacher knowledge in mathematics. We focused on Concept Cartoons as a tool for assessing reasoning on arithmetic topics (e.g., Samková and Tichá, 2017), in qualitative as well as mixed research designs (Samková, 2019). To systematically map the terrain, we naturally proceeded to a question whether and how it is possible to use Concept Cartoons to assess reasoning about mathematical statements. Combined with the aim to provide a common approach to professional preparation of future teachers that would address both science and mathematics education, we raised a question whether and how it is possible to do the assessment of reasoning jointly in science and mathematics subjects.

Such an arrangement resulted in an exploratory qualitative empirical study that is presented in this paper. As the main participants in the study, we involved future primary school teachers. We draw on the fact that primary school teachers are generalists in our country (i.e., are intended to teach both science and mathematics). Thus, we can pay comparable attention to their argumentation skills in both school subjects. To obtain results that would be applicable in professional preparation of primary school teachers at our university, we proceeded from the structure of our five-year professional preparation program that in its first three years approaches the study of content and didactics in mathematics and biology simultaneously. Other science subject domains (physics, chemistry) are approached later in the program and to a much lesser extent. Thus, we choose biology as the domain of our interest in science.

The presented research study is of an iterative nature. It started as a study that was conducted by the first author and focused on modes of argumentation related to reasoning about general statements in arithmetic (Samková, 2020); for the rest of the paper, we will call this study *an arithmetic study*. Following the findings of the arithmetic study, the first author contacted the second author and together, we started an *arithmetic-biology study*. We conducted explorations covering general statements not only in arithmetic but also in biology, especially in marine zoology. But the coding process did not lead to a conceptually coherent structure of code categories that would be joint for both subjects. We were faced with the decision to either collect additional data or modify the perspective. At the same time, the first author conducted an independent study with the third author, where we explored reasoning about general statements in geometry (Vízek and Samková, 2021); we will call this study *a geometry study*. The course of the coding process in the geometry study indicated that the geometric perspective might be crucial for gaining the desired coherence in the arithmetic-biology study. Therefore, we combined the arithmetic-biology study with the geometry study to create a *mathematics-biology study* that is the main study of this paper. The arithmetic study, the arithmetic-biology study and the geometry study serve as sub-studies to the main study.

This paper has been developed as an extension of conference contributions that introduced separately the arithmetic study (Samková, 2020) and the geometry study (Vízek and Samková, 2021). It brings the two mathematical sub-studies into a common context, adds a biological dimension to the focus, and provides a conversion table between modes of reasoning about general statements in arithmetic, geometry and biology.

The text is organised as follows: at the beginning, it presents the background of the presented research, the three sub-studies and the main study. Then it discusses the findings and captures the implications for further research.

Argumentation and general statements in mathematics

When mathematics learners develop, their view of a proper response to questions whether and why a mathematical statement is true develops as well. The first attempts of such a response usually have the form of a *reference to an authority* (e.g., “It is true because you told us yesterday.”), later on to one or more *confirming examples* – particular cases in which the statement holds. The confirming examples are welcome when the statement is existential (because they confirm the desired existence) but not so welcome when the statement is general. In the latter case, the disadvantage of confirming examples is that sometimes they can be easily used to mistakenly show the veracity of a false statement. For instance, the general statement that multiplying one positive number by another positive number always produces a bigger number looks true when you multiply 2 by 3, or 5 by 2.

Quite a different role in argumentation is played by *counter-examples* – particular cases in which the statement does not hold although they meet all the prerequisites of the statement. Such counter-examples are welcome when the statement is general since one counter-example is enough to disprove the statement. For instance, the statement that multiplying one positive number by another positive number always produces a bigger number may be disproved by a counter-example consisting in multiplying 2 by 1.

Various types of argumentation in assessing and proving mathematical statements were systematised by Harel and Sowder (1998) under the name of *proof schemes*. In their interpretation, ‘a person’s proof scheme consists of what constitutes ascertaining and persuading for that person’ (ibid: 244). Their multi-level classification of proof schemes includes:

- *external conviction proof schemes* (doubts are removed by the ritual of the argument presentation, by the word of authority, or by the symbolic form of an argument);
- *empirical proof schemes* (doubts are removed by quantitatively evaluating the conjecture by one or more specific cases – *inductive proof scheme*, or by reasoning from an illustration or a geometric figure, regardless of possible transformations – *perceptual proof scheme*);
- *analytical proof schemes* (doubts are removed by means of logical deductions: *transformational proof schemes* are based on operations on objects and anticipations of results of the operations, *axiomatic proof schemes* are based on some prior results, axioms and definitions).

Electronic ISSN 1803-1617 Printed ISSN 2336-2375

In some later sources (e.g., Harel and Sowder, 2007), the name of *deductive proof schemes* is used instead of *analytical proof schemes*, to highlight the role that deductive reasoning plays in this type of proof schemes: 'Deductive reasoning is a mode of thought commonly characterized as a sequence of propositions where one must accept any of the propositions to be true if he or she has accepted the truth of those that preceded it in the sequence.' (ibid: 811). Harel and Sowder also relate their proof schemes to other taxonomies, e.g., the taxonomy given by Balacheff (1988), and refer to a *generic example* as a sample of a transformational proof scheme. In their interpretation, a generic example is a 'justification by an example representing salient characteristics of a whole class of cases' (Harel and Sowder, 2007: 810).

Primary school teachers and their reasoning about general statements in mathematics

Although Harel and Sowder's research leading to the typology of proof schemes in mathematics was conducted generally, with college students as respondents and with special attention paid to mathematics major students, the typology might be applied to any stage of mathematical education, including the primary school level. Axiomatic proof schemes might be too formal for the primary school level, but all the other types of proof schemes might naturally appear there (Komatsu, 2010).

A typology that might be considered an alternative to the Harel and Sowder's system of proof schemes and is more apposite for the case of future primary school teachers, was introduced by Simon and Blume (1996). According to data collected from future elementary school teachers during a mathematics content course, Simon and Blume propose a list of five levels of responses attempting to justify mathematical statements (ibid: 17):

- 'Level 0 – Responses identifying motivations that do *not* address justification.
- Level 1 – Appeals to external authority.
- Level 2 – Empirical demonstrations.
- Level 3 – Deductive justification that is expressed in terms of a particular instance (generic example).
- Level 4 – Deductive justification that is independent of particular instances.'

Future primary school teachers have often got deeply rooted misconceptions about argumentation and proving: they tend to rely on an external authority as the basis of their conviction (ibid), believe that it is possible to affirm the validity of a generalisation through a few examples, randomly selected examples or big-number examples (Martin and Harel, 1989; Stylianides and Stylianides, 2009). Some of them do not understand the role of counter-examples in refuting general statements and consider one counter-example not being enough (Zazkis and Chernoff, 2008). Similar findings were also reported for secondary school students (Galbraith, 1981) and college students (Selden, 2012). Such a state may be persistent and, later in school practice, may affect how primary school teachers notice essential mathematical reasoning forms when justification and generalisation appear in the classroom (Melhuish, Thanheiser and Guyot, 2020).

Argumentation in science

In science education, argumentation is considered a core skill for the development of students' scientific literacy, critical thinking, and reasoning (Berland and Reiser, 2009; Lazarou, Sutherland and Erduran, 2016). The ability of students to form and understand arguments related to scientific phenomena and the processes behind them is the basis of scientific literacy. The quality of students' argumentation allows the teacher to recognise how the students understand the issue (Cullen et al., 2018).

The use of reasoning in science subjects is a complex and systematic approach that can include a lot of activities (Erduran and Jiménez-Aleixandre, 2007). One of the approaches used in science subjects is the process of evaluating and justifying claims mediated by Concept Cartoons (Naylor, Keogh and Downing, 2007). The arguments provided by students and teachers in science classrooms can support previously discussed knowledge, put it in a new context, and use the creation of various artefacts through which students could support their arguments (Furtak et al., 2010). Scientific reasoning and argumentation had a positive effect on students' achievement (Dofner et al., 2018) and this approach could be used to improve students' scientific knowledge. Ping, Halim and Osman (2020) found that students were able to provide explanations when using the argument, which included evidence supporting their opinion.

The role of general statements in inquiry-based education and formative assessment

Inquiry-based education has been recently perceived as a means of providing a sustainable connection between up-to-date scientific research and everyday school practice (McComas, 2002; Minner, Levy and Century, 2010). Thus, it consists of classroom activities involving students that observe, pose questions, reason, search for information, collaborate, collect data and interpret them, discuss obtained results (Dorier and Maass, 2014). Such an environment is naturally rich in generalisations (Bulková, Medová and Čeretková, 2020) and general statements of various appearances and validity are frequently voiced in the classroom.

As for the teachers, the inquiry-based environment requires teachers' feedback that would support students' learning (Dolin and Evans, 2018). The feedback typically addresses four different aspects of the learning situation: the task, the process to complete the task, student's self-regulation and student's persona (Hattie and Timperley, 2007). In this paper, we focus mainly on the second aspect and its part that relates to the components of the solution procedure already given by the student (e.g., whether the procedure is correct, suitably expressed, explained and interpreted). Namely, we focus on feedback that the teacher provides to the student during the process of generalisation and its interpretation: we plan to investigate the quality of responses that the teacher provides to assess the validity of a general statement given by the student.

MATERIALS AND METHODS

Our study aims to answer two research questions, the first one serving as preparatory for the second one:

RQ1: What kinds of reasoning about general statements in biology and mathematics can be observed in future primary school teachers when using Concept Cartoons as a diagnostic instrument?

RQ2: What are the possibilities of joint assessment of reasoning about general statements in biology and mathematics?

The study is of an exploratory qualitative design since the phenomenon of a possible joint approach to reasoning about general statements in mathematics and biology through Concept Cartoons has not been studied before. To explore and describe the nature of the phenomenon, collected data are analysed qualitatively, using open coding and constant comparison (Miles, Huberman and Saldaña, 2014).

The research is iterative; it consists of three consecutive sub-studies and the main study. The first and third sub-studies focus just on the first research question narrowed to the context of mathematics. The second sub-study and the main study cover both research questions to their full extent.

Participants

The three sub-studies were conducted with participants studying the second or third year of a five-year full-time master degree program for future primary school teachers. This program is mostly frequented by prospective teachers that came to university directly from the upper-secondary school. In the Czech Republic, where the study takes place, primary school teachers are generalists; they are supposed to teach all school subjects at the primary school level (students from 6 to 11 years of age).

The first sub-study had 28 participants (labelled as *group 1*), the second sub-study 49 participants (*group 2*) and the third sub-study had 29 participants (*group 3*; 26 of them being the same as in group 1). In all three sub-studies, data were collected at a compulsory course focusing on mathematical content preparation of future primary school teachers. The participants from groups 1, 2 and 3 are called *standard participants*.

The main study reprocessed data collected during the three sub-studies and, additionally, included nine participants studying the last year of a full-time master degree program for future lower-secondary school teachers (students from 11 to 15 years of age). In the Czech Republic, these teachers are specialists; each of them specialises in two school subjects of their own choice. We included five future lower-secondary school teachers specialising in mathematics (labelled as *group 4*) and five specialising in biology (*group 5*); one of them specialised in both mathematics and biology (i.e., belonged to both groups). All of them were selected from future lower-secondary school teachers working on a diploma thesis in mathematics or biology specialisation and, for the purpose of data collection, contacted individually. The participants from groups 4 and 5 are called *additional participants*.

As mentioned above, the main study worked with data collected from all standard and additional participants, i.e., from 28 + 49 + 29 – 26 + 5 + 5 – 1 = 89 different participants. To identify individual participants, we used codes consisting of a letter referring to a group of participants (S for groups 1 and 3, R for group 2, D for groups 4 and 5)[1] and a randomly assigned number. For instance, R29 identifies a participant from group 2.

Diagnostic instrument – common aspects

As a diagnostic instrument, we used *Concept Cartoons*. Concept Cartoons had been initially developed by Keogh and Naylor (1999) as a tool to motivate and support the learning of students during elementary science lessons. Later, they were also implemented in other school subjects, including mathematics (Dabell, Keogh and Naylor, 2008). Each Concept Cartoon has a form of a picture with a school or out-of-school everyday situation related to a given curricular content and several children discussing the situation through a bubble dialogue. Our previous research showed how Concept Cartoons might be used as a tool for diagnosing various types of teacher knowledge (e.g., Samková, 2019).

In this study, we employ three Concept Cartoons related to the topic of general statements, one located in arithmetic, one in biology and one in geometry. Although the original sets of Concept Cartoons exist in mathematics (Dabell, Keogh and Naylor, 2008) as well as in science (Naylor and Keogh, 2010), none of the original Concept Cartoons suited precisely our purpose. We were looking for Concept Cartoons based on widely spread misconceptions related to Czech primary school curriculum that could be presented in the form of general statements. Moreover, we needed the statements to allow specifying many confirming examples as well as counter-examples. In the bubbles, we expected other statements that might serve as hints to the presented issue. Among the original Concept Cartoons, we did not find any with these attributes. Therefore, the arithmetic Concept Cartoon for our study was created by a modification of the content of all bubbles in an original Concept Cartoon (Dabell, Keogh and Naylor, 2008: 3.2), and the biology and geometric Concept Cartoons arose independently of the original sets. The three Concept Cartoons and the circumstances of their origin will be introduced in detail within the description of the sub-studies.

THE COURSE OF THE RESEARCH

Data collection and data analysis – common aspects

During all three sub-studies and the main study, data collection always consisted of assigning a worksheet with the diagnostic Concept Cartoon(s). Then we asked the respondents to respond to (each of) the pictures in a written form: to decide which children in the picture are right and which are wrong and justify the decision. The participants worked on the worksheet individually.

During data analysis, we first registered which bubbles were chosen by individual respondents as correct and under which additional conditions. Afterwards, we openly coded all the material (the written responses given by the respondents as well as the diagnostic Concept Cartoons), looking for various

1 The distribution of reference letters follows external rules that the first author has for systematic handling of all her data. This system enables to differentiate between different study groups identified by university, study program, and year of commencement. Groups 1 and 3 in this paper are subsets of the same study group, thus the same letter for identification of their members.

aspects related to reasoning, justification and argumentation in mathematics and science. Then we applied the method of constant comparison – from the overall perspective, from the perspective of individual bubbles across all participants and from the perspective of individual participants across all bubbles. At the end of the process, the aim was to have assigned exactly one code category to each of the participants and each of the bubbles included in the research. For instance, in the first sub-study, we worked with 28 participants and data analysis included four bubbles of the arithmetic Concept Cartoon. Therefore, we intended to finish the analytic process with exactly $28 \cdot 4 = 112$ assignments.[2]

From the perspective of the general scheme of qualitative data analysis shown in Figure 1 (the diagram starts at the top), the situation with assignments of code categories turned out differently in different sub-studies. During the first and third sub-studies, we successfully ran several rounds of the constant comparison process (the middle-central part of the diagram) and finished the analytic process by providing the intended one-to-one assignments that led to establishing a conceptually coherent structure of codes (the lower-right corner of the diagram). During the second sub-study, the constant comparison process was not successful and we ended in the lower-left corner of the diagram. The idea of combining all three sub-studies together to create a basis for the main study enriched the second sub-study by desired additional data. It enabled to continue the analytic process (i.e., to return to the top part of the diagram). The detailed overall scheme of data handling and types of assigned code categories during individual stages of the research is shown in Figure 2.

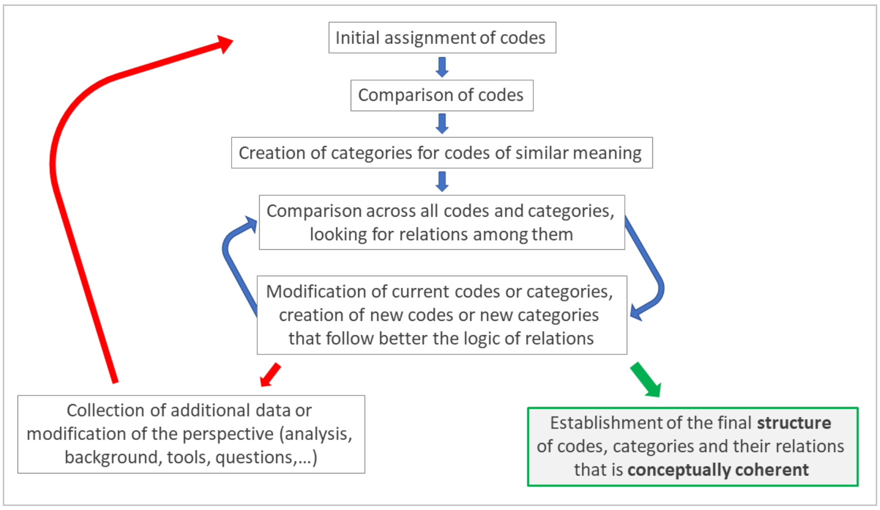

Figure 1: The general scheme of qualitative data analysis (source: own illustration)

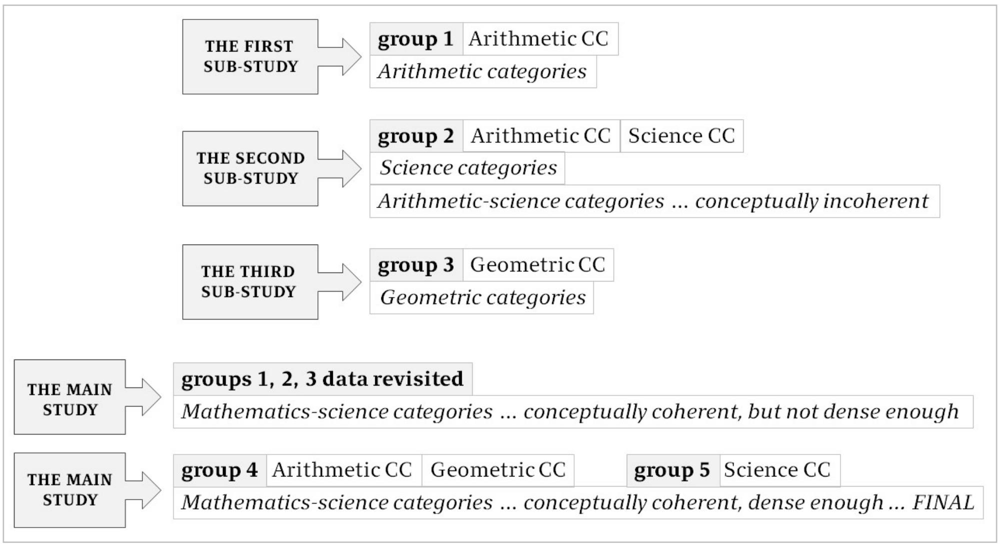

Figure 2: The overall scheme of data handling and types of assigned code categories during individual stages of the research, CC = Concept Cartoon (abbr.)

2 In the Czech Republic, where the research study takes place, the proper symbol for multiplication in school mathematics is given as " · ", i.e., a middle dot. To keep the authenticity of data excerpts presented in this paper and to maintain the coherence of the symbolic notations throughout the paper, we use this symbol also for multiplication that appears in the texts outside data excerpts.

THE ARITHMETIC SUB-STUDY (THE FIRST SUB-STUDY)

The arithmetic sub-study focused on argumentation related to general statements in arithmetic, namely to statements about the properties of results of the multiplication operation. We proceeded from the fact that these properties are different within the domains of natural numbers, rational numbers and integers. Thus, such an environment has great potential for exploring various confirming examples, counter-examples and conditions of validity.

Diagnostic instrument – particularities

In the Concept Cartoon, we presented an informal description of the operation of multiplication and a bubble dialogue comprising of four general statements on multiplication based on four common misconceptions about the operation: “multiplication makes a bigger number”, “multiplication by a fraction makes a smaller number”, “multiplication by a negative number makes a smaller number” and “multiplication by zero makes a smaller number” (Figure 3). The domain set of the numbers in focus is not specified in the picture intentionally; it might equal the set of natural numbers as well as integers, rational numbers (fractions, decimal numbers) or real numbers.

The misconceptions hidden behind individual bubbles are partially related to critical moments in learning when an earlier way of thinking fails to account sufficiently for new ideas (cf. Confrey and Kazak, 2006); they refer respectively to

- multiplication as a repeated addition of a positive integer which gives a result that is bigger than the repeated number;
- multiplication of a positive integer by a positive proper fraction which gives a result that is smaller than the positive integer;
- multiplication of a positive integer by a negative integer which gives a negative result that is smaller than the positive integer;
- multiplication of a positive integer by zero, which gives a zero result that is smaller than the positive integer.

All misconceptions more or less relate to *natural number bias* (Alibali and Sidney, 2015; van Dooren, Lehtinen and Verschaffel, 2015), a tendency to use considerations and procedures learned in the domain of natural numbers and transfer them unjustifiably to the domain of integers or rational numbers. One of the misconceptions combines natural number bias with a similar tendency related to the transfer from the domain of proper fractions to the domain of all fractions (Stevens et al., 2020).

As for the veracity of the statements in bubbles, none of them is correct since for each of them there exist counter-examples (numbers for which the statement is not valid). However, it is possible to specify under what conditions the statements are valid:

- Jan: for all pairs of positive numbers bigger than 1, for all pairs of negative numbers, for positive numbers multiplied by positive numbers bigger than 1, for negative numbers multiplied by zero, for negative numbers multiplied by positive numbers smaller than 1;
- Emil, Bara: for all positive numbers;
- Tom: for positive numbers multiplied by positive fractions smaller than 1, for negative numbers multiplied by positive fractions bigger than 1, for positive numbers multiplied by negative fractions.

The most general statement is the statement in the upper-left bubble (Jan). It allows variations in multiplier as well as multiplicand, and none of them is specified in the bubble. The other three statements might be considered hints to Jan’s bubble, since they point indirectly out some of the types of multipliers for which the most general statement might not be valid: number zero (Emil), negative numbers (Bara), fractions (Tom).

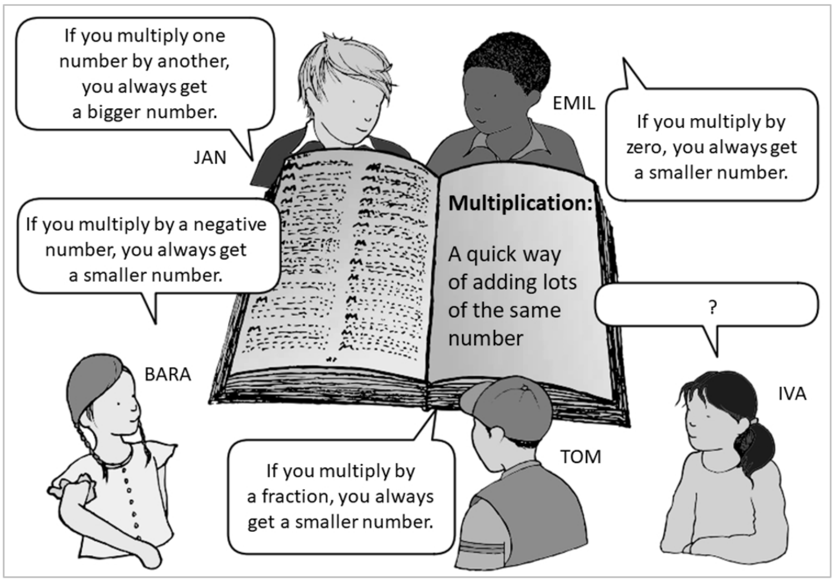

Figure 3: A Concept Cartoon on multiplication; (source of the text in the book, the template of the book and the template of children with empty bubbles: Dabell, Keogh and Naylor, 2008: 3.2)

Results of the arithmetic sub-study

During the data analysis process, $28 \cdot 4 = 112$ assignments were made and eight different code categories appeared as a result of the analytic process. Each of the categories was labelled by an acronym. The list of all relevant acronyms with their short description and frequency in data is shown in Table 1. In the further text, the meaning of the acronyms is explained and accompanied by data excerpts, in the same order as in Table 1.

acronym	description	bubbles				
		Jan	Emil	Bara	Tom	all
CEX	one or more counter-examples	14	5	6	8	33
GE	generic examples, no justification	1	2	6	1	10
GED	generic examples, with justification	0	1	2	2	5
COX	conditions of validity indicated, no justification	7	2	4	4	17
COD	conditions of validity indicated, with justification	3	5	6	3	17
OF	over-fixation to previous learning	0	5	0	3	8
VR	vague response	3	8	4	5	20
XR	no response	0	0	0	2	2
altogether		28	28	28	28	112

Table 1: List of relevant acronyms for arithmetic code categories and their frequency in data, *n* = 28, 2016 (source: own calculation)

According to the results of the analysis, most of the participants based their reasoning just on one or more counter-examples, without any additional reasoning (code category CEX):

S32 Jan: No! $1 \cdot 358 = 358 \rightarrow$ stays the same.

Several of the counter-examples were accompanied by a commentary that linked them to a wider group of numbers with the same behaviour within the given context, i.e., these counter-examples might be considered as generic examples (code category GE):

S33 Bara: She considered neither zero nor negative numbers
$(-8) \cdot 0 = 0$ is not smaller
$(-3) \cdot (-3) = 9$ is not smaller

Other responses consisted in counter-examples that were generic examples accompanied by other deductive arguments (code category GED), more or less correct:

S29 Bara: When I multiply a negative number by a negative number, I get a number that is positive, so bigger, $-3 \cdot (-3) = 9$. When I multiply zero by a negative number, I get zero – the same number.

Some participants did not apply counter-examples, instead they just indicated conditions under which the given statement holds, with a proper deductive justification (code category COD) or without (code category COX). Some of them were able to cover properly the conditions for Emil and Bara, but none of them covered completely the conditions for Jan or Tom:

S26 Jan: It would be true if we were working within the set of natural numbers.

Eight of the 28 participants displayed erroneous considerations due to their over-fixation to a particular learning context (natural numbers) or a particular didactic model (e.g., a fraction as a part of a whole); they were assigned the code category OF:

S10 Tom: Yes – a fraction is a part of a whole → the result will be always smaller.

Twelve of the 28 participants provided vague or unclear responses to some or all of the bubbles (code category VR). These responses could not be considered as properly justified arguments:

S33 Emil: If you multiply by 0, you always get 0.

Among the 112 responses, 2 were blank and 32 belonged to the categories GE, GED, COD that refer to generic examples and other deductive reasoning.

Emerging concerns

After the arithmetic sub-study, the above-mentioned findings emerged a concern whether similar code categories would also appear when letting the participants reason about a Concept Cartoon with general statements in a different school subject, namely in biology. We addressed this concern in the arithmetic-biology sub-study.

THE ARITHMETIC-BIOLOGY SUB-STUDY (THE SECOND SUB-STUDY)

The arithmetic-biology sub-study intended to focus on argumentation related to general statements together in arithmetic and biology. For this purpose, we created a new Concept Cartoon situated in the context of biology, added it to the Concept Cartoon from the arithmetic sub-study and conducted a new sub-study with both of them.

Diagnostic instrument – biology – particularities

For the biology Concept Cartoon, we chose a zoology topic, namely the topic of large marine animals and their classification. To ensure a similar environment for argumentation as in the arithmetic case, we focused again on possible misconceptions related to critical moments in learning when an earlier way of thinking fails to account sufficiently for new ideas. As pointed out by the research about student thinking in such a context, students often form their knowledge and ideas according to their own experience with the environment and on the basis of intuition (Lazarowitz and Lieb, 2006). A common misconception consists in sorting organisms into incorrect taxonomic classes according to their habitat and/or similarity in a body structure (Kattmann, 2001), for instance, in determining large marine mammals as fish based on the fact that they live in the water (Trowbridge and Mintzes, 1988; Berthelsen, 1999). A similar misconception consists in the inclusion of turtles and reptiles among amphibians (Yen, Yao and Chiu, 2004) or invertebrates (Braund, 1997).

Kubiatko and Prokop (2007) performed a study with 468

Slovak lower-secondary school students in which they monitored various misconceptions related to mammals. An often-occurring misconception was the claim that the penguin is a mammal, justified by the fact that penguins live in the sea just like large marine mammals (cf. Prokop, Prokop and Tunnicliffe, 2007). Only half of the respondents reported that baby whales feed on milk, even though most of them correctly classified whales as mammals. More than two thirds of the respondents displayed a misconception about dolphin breathing, as they reported that dolphins breathe through the gills because they live in an aquatic environment. In connection with the students' difficulties in the classification of whales, Kubiatko and Prokop also provide a possible linguistic explanation for this issue: the Slovak translation of the word *whale* has a meaning of *big fish*. The same situation is in the Czech language.

Based on the above-mentioned findings, we created a Concept Cartoon displaying various improper ways of affiliating animals according to their habitat and/or similarity in body structure or function (Figure 4):

- classifying all big animals living in the water as fish, since they inhabit the same environment as fish;
- including an animal among fish based on the fact that it has gills;
- including an animal among fish because it reproduces by laying eggs.

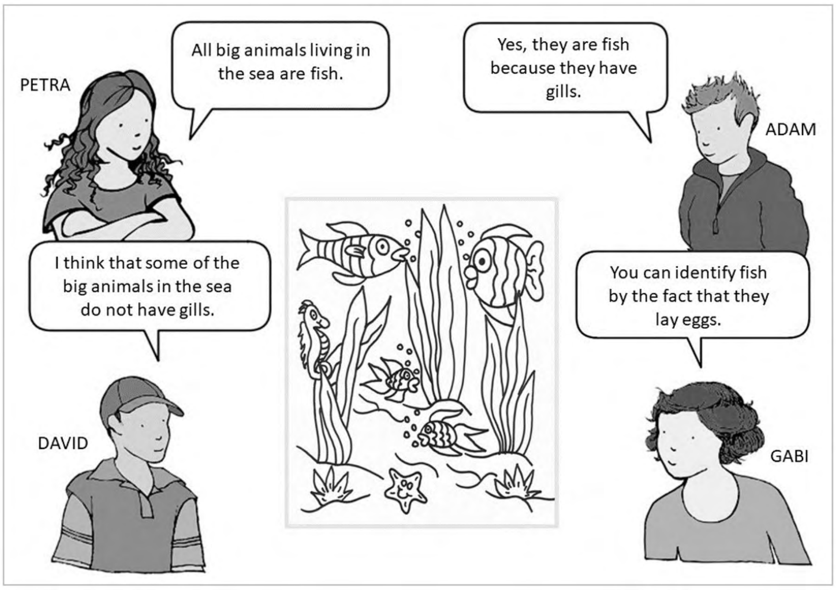

Figure 4: A Concept Cartoon on large marine animals; (source of the template of children with empty bubbles: Dabell, Keogh and Naylor, 2008: 2.10; source of the central picture: Clipart Library, 2019)

Three of the bubbles are not correct, and it is possible to specify many confirming examples as well as counter-examples for their statements:

- Petra: there are large fish living in the sea (e.g., giant oarfish – *Regalecus glesne*) but the largest marine animals are mainly mammals (e.g., whales, dolphins), cartilaginous fishes (e.g., sharks) or cephalopods (e.g., giant squid); with this bubble, the question also emerges of how big must an animal be to be considered a large marine animal;
- Adam: we can identify fish by having gills but most large marine animals are representatives of mammals and thus breathe through the lungs;
- Gabi: most fish lay their eggs, and it is their main way of reproduction; however, there are exceptions, e.g., the family of breeding fish – *Poeciliidae*, who give birth to live babies.

The fourth bubble (David) contains the only statement in the Concept Cartoon that is not a general statement. It is an existential statement that is correct.

Similarly to the arithmetic case, the top-left bubble (Petra) shows the most general statement. The other three bubbles can be considered hints to Petra's bubble, since they indirectly indicate some partial characteristics of the counter-examples related to Petra's bubble. Again, none of the presented characteristics is complete.

Data analysis – particularities

For data analysis, to obtain as similar environments as possible for arithmetic and biology, we decided to analyse just two bubbles, one from each of the Concept Cartoons. We chose the most similar bubbles from the perspective of general statements: Jan's bubble in the arithmetic Concept Cartoon (Figure 3) and Petra's bubble in the biology Concept Cartoon (Figure 4). Both are the most general bubbles in the picture, and the other bubbles serve as their support.

Results of the arithmetic-biology sub-study – the arithmetic part

During the data analysis process, 49 assignments were made for the arithmetic Concept Cartoon (one assignment for each participant – we analysed just Jan's bubble). These assignments covered six of the eight code categories from the arithmetic sub-study (Table 1, excluding GED, COD), plus three new categories for erroneous responses (Table 2).

acronym	description	Jan
DC	decimal numbers error	4
ON	number one error	1
DN	different numbers error	1
altogether		6

Table 2: List of acronyms for new arithmetic code categories and their frequency in data, *n* = 49, 2020 (source: own calculation)

acronym	description	single	multiple	altogether
NC	names of classes	9	1	10
NO	names of orders	1	0	1
NS	names of species	2	0	2
NCS	names of classes and their species	15	11	26
NOS	names of orders and their species	0	1	1
NCO	names of classes and their orders	0	1	1
NCS&NOS	names of orders and their species, names of classes and their species	0	1	1
NCOS	names of classes and their orders and their species	1	1	2
altogether		28	16	44

Table 3: List of acronyms for biology code categories for counter-examples and their frequency in data, *n* = 49, 2020 (source: own calculation)

acronym	description	
BA	big-animals error	2
DM	dolphins not being mammals	1
altogether		3

Table 4: List of acronyms for new biology code categories for errors and their frequency in data, *n* = 49, 2020 (source: own calculation)

Code category DC is for the opinion that decimal numbers are numbers between 0 and 1:

R43 Jan is not right, for instance, when I multiply by a decimal number ($6 \cdot 0.5 = 3$), a smaller number comes out.

Code category ON for the opinion that multiplying by 1 gives a smaller number:

R34 Jan is not correct because when we multiply by 1 or 0, the result is smaller.

Code category DN for the opinion that a product of two different numbers gives a bigger number:

R22 $2 \cdot 5 = 10$ when multiplying two different numbers, we get a bigger number.

Results of the arithmetic-biology sub-study – the biology part

For the biology Concept Cartoon, the analytic process was not successful. For 44 of the 49 respondents, we came in data across argumentation aspects that had not appeared in the arithmetic context: counter-examples that consisted of various combinations of names of classes, names of orders and names of species. In these cases, we were not able to eliminate the assignments to one per respondent. Instead, for these 44 respondents, we concluded the analytic process with two assignments per respondent: one assignment for the type of counter-example(s) provided by the respondent (namely, for the particular combination of classes, orders and species), and the other for the number of counter-examples provided by the respondent (single vs multiple), see Table 3.

Among the 44 respondents with counter-examples, 28 presented single counter-examples of various types:

R13 Petra is not right, because mammals can also live in the sea. (code category NC)

R37 Petra – not all large animals in the sea are fish, they are cetaceans. (cat. NO)

R8 Other big animals live in the sea, not just fish, e.g., a whale. (cat. NS)

R25 Not all large animals are fish, for instance, a dolphin is a mammal. (cat. NCS)

R39 Petra is not right – there are also mammals in the sea (cetacean – whale). (cat. NCOS)

The remaining 16 participants provided multiple counter-examples:

R29 Petra is not right, we have also cartilaginous fish and mammals. (cat. NC)

R49 There are not only fish in the sea but also, for example, mammals (dolphin), octopus – cephalopods, turtle – kind of reptiles. (cat. NCO&NCS)

R20 Petra is not right, because cetaceans, e.g., blue whale, dolphin, orca, are big marine animals but mammals. (cat. NCOS)

The rest of the respondents (5 of 49) gave either no response (1 case; code category XR, the same category as in arithmetic), vague response (1 case; code category VR, the same category as in arithmetic) or erroneous answers based on conceptual mistakes. The erroneous answers resulted in establishing two new code categories (Table 4).

Code category BA is for exemplifying marine animals that are not big (e.g., seahorses, shrimps):

R31 Jellyfish and seahorses also live there.

Code category DM for not including dolphins among mammals:

R35 Some animals in the sea do not have gills, such as the dolphin, and mammals also live in the sea.

Due to the inability to eliminate the assignments related to biology counter-examples to one per respondent and relate them to the existing arithmetic code categories, we were faced with the necessity to either change the perspective from which we observed our data or collect additional data to help clarifying the situation. That means, we ended in the situation from the lower-left corner of the diagram in Figure 1. For some time, the analytics process stayed open, unfinished. The resolution came with the geometry sub-study.

THE GEOMETRY SUB-STUDY (THE THIRD SUB-STUDY)

The geometry sub-study also proceeded from the arithmetic sub-study but was conducted independently of the arithmetic-biology sub-study. We got the advantage of the newly established cooperation between the first and third authors and took the opportunity to work on the topic of geometry in which Concept Cartoons have not been properly investigated yet. We employed a Concept Cartoon with data collected many years ago but never researched.

Diagnostic instrument – particularities

The geometric Concept Cartoon discusses the topic of recognising a rectangle. In this context, we deal with two possible ways of comprehending quadrilaterals. The first one considers each quadrilateral as a unique object. For instance, a square is not understood as a special case of rectangle, i.e., rectangles are considered having different length of adjacent sides. De Villiers (1994) calls this approach a *partition classification*. Such an approach is typical for geometric concepts at the primary school level but is not suitable for higher levels of schooling since it makes it difficult to study common properties and relationships between various objects. The secondary school curriculum thus employs a *hierarchical classification* (ibid) that understands specific concepts as subsets of the more general ones. In that sense, squares are considered special cases of rectangles; rectangles are considered special cases of parallelograms, etc. The transfer between partition and hierarchical classifications is a rich source of critical moments in learning when earlier ways of thinking fail to account sufficiently for new ideas.

According to educational research, students often face difficulties with the classification of quadrilaterals: they formulate statements with superfluous (Miler, 2019) or insufficient (Fujita and Jones, 2007) information. Similar difficulties are also observed in future teachers (Tuset, 2019).

The Concept Cartoon in Figure 5 covers various properties of rectangles: “having opposite sides equal in length”, “having all angles right”, “having diagonals that bisect each other” and “having diagonals equal in length”. To be able to identify the rectangle properly, we need a precise identification of the rectangle by assigning properties that are necessary and sufficient for determining the object. Within the framework of hierarchical classification of rectangles, the definition may consist of e.g., in the combination of properties “having diagonals that bisect each other” and “having diagonals equal in length”. Within the framework of partition classification, the two properties would be accompanied by the property “having different lengths of adjacent sides”.

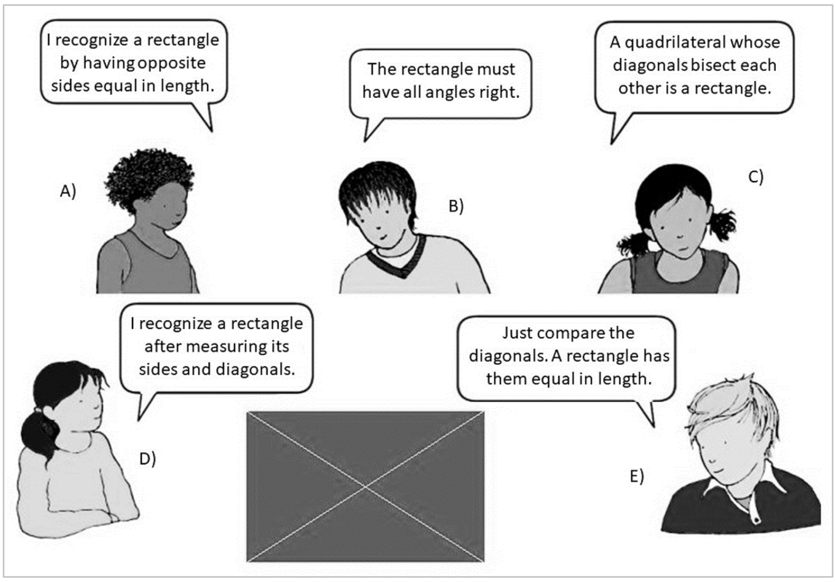

Figure 5: A Concept Cartoon on recognising a rectangle; (source of the template of the children with bubbles: Dabell, Keogh and Naylor, 2008: 1.3; source of the texts in bubbles: Roubíček, 2014)

For this sub-study, we concentrated on bubbles A, C and E of the Concept Cartoon. These three bubbles contain statements that can be all reformulated as implications:

(A) If a quadrilateral has opposite sides equal in length, then it is a rectangle.

(C) If a quadrilateral has diagonals that bisect each other, then it is a rectangle.

(E) If a quadrilateral has diagonals equal in length, then it is a rectangle.

None of the above statements is true since the assumptions do not contain enough information to identify a rectangle unequivocally. Specifically, there exist counter-examples for which the three statements are not valid: A, C and E are not valid for all parallelograms outside of rectangles (e.g., a rhomboid), E is not valid also for some quadrilaterals outside of parallelograms (e.g., an isosceles trapezoid). However, in all three cases, the converse implication is true, i.e., all properties mentioned in the assumptions are actual properties of rectangles. Moreover, it is possible to add additional conditions to the assumptions that would make the statements true; the additional conditions are hinted in the other bubbles.

From the perspective of argumentation, such an arrangement of the content of bubbles can challenge respondents' skills in conditional reasoning, namely their ability to distinguish properly between assumptions and conclusions of a conditional statement that is informally worded. The informal type of conditional statements often appears in primary school classrooms (Hadar, 1977).

Results of the geometry sub-study

During the analytic process, 29 · 3 = 87 assignments were made (3 bubbles for each participant) and eleven relevant code categories were revealed in data (see Table 5). This time, while creating acronyms for code categories, we newly distinguished between three-letter and two-letter ones. The three-letter acronyms were intended for code categories referring to responses that might be considered a constructive basis for further learning of proper argumentation. The two-letter acronyms were intended for code categories referring to blank responses or responses that were incompatible with proper argumentation. The meaning of the individual acronyms is explained in detail in the further text, in the same order as in Table 5.

acronym	description	bubbles			
		A	C	E	A+C+E
GSX	name of a set of counter-examples, no particular example, no justification	6	10	11	27
GEX	name of a set of counter-examples, a particular example given	0	1	0	1
GSD	name of a set of counter-examples, with justification	0	1	1	2
AWX	awareness of the existence of counter-examples, no particular example, no justification	0	4	5	9
AWC	awareness of the existence of counter-examples, conditions of validity indicated but insufficiently	3	0	0	3
COX	conditions of validity indicated, no justification	2	0	0	2
XR	no response	4	7	8	19
AG	agreement with a bubble, no other comments	2	0	3	5
VR	vague response	1	0	0	1
EK	erroneous knowledge of concepts and their properties	0	3	1	4
CI	converse handling of the implication	11	3	0	14
altogether		29	29	29	87

Table 5: List of relevant acronyms for geometry code categories and their frequency in data, *n* = 29, 2014 (source: own calculation)

Data excerpts

According to data analysis, 44 of the assignments of code categories belonged to three-letter acronyms. Among them, the most frequent mode of reasoning consisted of naming a set of counter-examples but not providing a particular counter-example nor justification (code category GSX):

S9 (E) No, it could be, for example, a trapezoid.

Only one of the responses provided a name of a set of counter-examples accompanied by a particular counter-example that was in the form of an illustrative picture (code category GEX):

S9 (C) No, they can bisect in various quadrilaterals (for example, in a square: *a picture of a square with diagonals*).

One of the participants provided the name of a set of counter-examples accompanied by an attempt of a deductive justification (code category GSD):

S31 (C) If we draw both diagonals, 4 isosceles triangles will be formed. But watch out: the diagonals also bisect in a square, a rhombus and a rhomboid.

Some of the participants were aware that the statements in bubbles might not be always true but did not provide any (sets of) counter-examples or additional conditions of validity (code category AWX):

S13 (C) Not only a rectangle.

Others, who were aware that the statements in bubbles might not be always true, provided additional conditions of validity but these conditions were not sufficient (code category AWC):

S6 (A) Yes, that is true. However, we must add that the adjacent sides should not be of the same length. Then it would be a square.

Several participants provided sufficient additional conditions but without justification (code category COX):

S9 (A) No, the other condition is that it must have right angles. Otherwise, it could be a parallelogram.

The remaining 43 assignments of code categories belonged to two-letter acronyms, i.e., to code categories referring to blank responses or responses that were incompatible with proper argumentation. 19 of them referred to cases when the participant did not provide any response to a bubble (code category XR), five to cases when the participant provided only incorrect agreement with a bubble without further comments (code category AG).

One assignment referred to a response that was too vague to be clear (code category VR):

S10 (A) Because a = a'; b = b'.

Some of the participants presented erroneous knowledge of geometric concepts and their properties (code category EK):

S15 (C) A rectangle is a regular quadrilateral – its diagonals do not bisect each other.

Twelve of the participants handled the implications conversely, i.e., confused assumptions with conclusions (code category CI):

S25 (C) True, a rectangle has the diagonals of the same length, they bisect each other and mark the centre of the rectangle.

A reflection on the relation between arithmetic and geometry sub-studies

The results of the geometry sub-study complemented the results of the arithmetic sub-study. Having the current study situated to a different field (geometry) and using a different type of statements in bubbles (implications that are not true but the converse implications are true), only four code categories from the arithmetic sub-study reappeared here: XR, VR, GE and COX. However, the category GE has been renamed as GEX due to the newly established three-letter rule. The remaining code categories are quite new: GSX and GSD refer to the initial stages of GEX and GED where the respondent provides a name of a set of counter-examples but does not provide a particular sample; AWX refers to initial stages of COX or GEX; AWC refers to initial stages of COX; EK refers to erroneous responses similar as OF but with a different source of the error (over-fixation to previous concepts in OF vs weak knowledge of new concepts in EK); CI refers to erroneous knowledge of logical aspects; and AG refers to incorrect agreements without further comments (it might refer to initial stages of VR, EK or CI).

Emerging concerns

The above-mentioned reflection on the relation between the arithmetic and geometry sub-studies emerged a hope that the collection of geometric data might become the desired missing piece allowing us to conclude the still unfinished analytic process of the arithmetic-biology sub-study. For this purpose, we combined the three sub-studies together, revisited all their data and this is how the main study came about.

THE MATHEMATICS-BIOLOGY STUDY (THE MAIN STUDY)

Data analysis

During the main study, we revisited data from the arithmetic-biology sub-study and observed them from the perspective of the findings of the geometry sub-study. Primarily, we focused on possible interconnections between biology and geometry code categories. During the constant comparison process, we also revisited data from the arithmetic and geometry sub-studies. Similarly, like in the arithmetic-biology sub-study, we chose one bubble from each of the three Concept Cartoons for data analysis. From the geometric Concept Cartoon, we selected the C bubble. During the analytic process, we followed the convention of distinguishing between three-letter code categories referring to responses that might be considered a constructive basis for further learning of proper argumentation and two-letter code categories referring to blank responses or responses incompatible with proper argumentation.

Results of the main study – three-letter categories

This time, the analytic process was successful and we made $28 + 2 \cdot 49 + 29 = 155$ assignments in data collected during the first, second and third sub-studies, i.e., in all standard participants' data related to the three selected bubbles (Jan, Petra, C). These assignments covered seven existing three-letter categories (AWX, CEX, GSX, GSD, GEX, COX, COD). Since data were revisited during the main study, some of the excerpts were assigned a different category in the main study than in the sub-study. For instance, one of the responses from the arithmetic sub-study and one of the responses from the biology part of the arithmetic-biology sub-study that had been originally assigned the VR category were labelled later as the AWX category:

S18 Jan: You can get also a smaller number, or the same.

R48 Petra is wrong, not all the big animals in the sea are fish.

We considered four of the three-letter categories *relevant* for the joint mathematics-biology approach (i.e., for the second research question): AWX, GSX and GEX since they appeared in all three Concept Cartoons, and CEX that appeared in arithmetic and biology ones.

The other three-letter categories did not appear across the subjects (COX and COD were just in arithmetic, GSD just in geometry). So that we included additional participants (future lower-secondary school teachers who specialize in mathematics and/or biology) to see whether they could enrich the categories also in other subjects. During this additional analytic process, we made $2 \cdot 5 + 5 = 15$ assignments but none such enrichment came into being, so that we labelled the categories COX, COD, GSD as *irrelevant* for the joint approach.

For details on the relevant three-letter code categories, their frequency in data (from standard as well as additional participants = 155 + 15 = 170 assignments) and relations to the code categories from the sub-studies, see Table 6. The code categories in Table 6 are sorted by the quality of the argument, from the weakest (AWX) to the strongest (GEX).

Electronic ISSN 1803-1617 Printed ISSN 2336-2375

final acronym	description	frequency in data	previous acronyms
AWX	awareness of the existence of counter-examples, no particular counter-example given, no justification	7	AWX, VR (some)
CEX	a counter-example (i.e., an example for which the statement is not valid), detached, no justification	30	CEX, NS
GSX	a generic example in the form of a name or characteristics of a set of counter-examples, no particular element of the set given, no justification	32	GSX, NC, NO, NCO
GEX	a generic example in the form of a name or characteristics of a set of counter-examples, a particular element of the set given, no justification	54	GE, GEX, NCS, NOS, NCOS
altogether		123	

Table 6: Final list of relevant three-letter acronyms for mathematics-biology code categories and their previous appearances, frequency in data from all participants (170 assignments), *n* = 89, 2014-2021 (source: own calculation)

final acronym	description	frequency in data	previous acronyms
XR	no response	10	XR
AG	agreement with a bubble, no other comments	1	AG
VR	vague response	3	VR (some)
EC	erroneous knowledge of mathematical or science concepts and their properties	13	EK, OF, DC, ON, DN, BA, DM
EL	erroneous knowledge of logical aspects	4	CI
altogether		31	

Table 7: Final list of two-letter acronyms for mathematics-biology code categories and their previous appearances, frequency in data from all participants (170 assignments), *n* = 89, 2014-2021 (source: own calculation)

acronym	arithmetic	biology	geometry
	If you multiply one number by another, you always get a bigger number.	All large animals living in the sea are fish.	A quadrilateral whose diagonals bisect each other is a rectangle.
AWX	S18: You can get also smaller number, or the same.	R48: Not all big animals in the sea are fish.	S13: Not only a rectangle.
CEX	S31: $7 \cdot (-3) = -21$	R8: There are other big animals in the sea than just fish, e.g., a whale.	S25 (part): 
GSX	S4: Numbers might be smaller (e.g., after multiplying a negative number).	R26: There are also mammals in the sea.	S6: Yes, but also a square.
GEX	R25: That is not true because when multiplying negative numbers, we can get a number that is smaller: $8 \cdot (-8) = -64$.	R43: There are also mammals in the sea (e.g., dolphin).	S9: No, they can bisect in various quadrilaterals, for example, in a square: 

Table 8: A conversion table between data excerpts from arithmetic, biology and geometry across all relevant three-letter mathematics-biology code categories

Results of the main study – two-letter categories

As for the two-letter categories, the assignments from the main study covered three existing two-letter categories (XR, VR, AG). Moreover, two completely new categories were established to generally cover all the previous categories on various conceptual (EC) and logical (EL) errors. For details on two-letter code categories, see Table 7.

The sum of occurrences from Tables 6 and 7 equals 123 + 31 = 154. The remaining 170 – 154 = 16 assignments belong to the irrelevant three-letter code categories (COX, COD, GSD) that are not listed in the tables.

Results of the main study – the conversion table

To illustrate better the interrelations between code categories across all subjects, we prepared a conversion table that compares data excerpts from arithmetic, biology and geometry across all relevant three-letter code categories (Table 8). Since the CEX category was not represented in geometric data, we drew on the fact that GEX = CEX + GSX and, as an example of CEX in geometry, we selected a part of an excerpt labelled as GEX.

Using the final tables of acronyms and the conversion table (Tables 6 to 8), we may study the modes of argumentation in different subject domains (arithmetic, biology, geometry) from the perspective of individual respondents. We applied this approach to data from the arithmetic-biology sub-study, with the following results: 48 of the 49 respondents were assigned relevant code categories both in arithmetic and biology, of them 17 were assigned the same code category in both subjects, 19 were assigned two different three-letter categories, 3 were assigned a two-letter code category in biology but a three-letter code category in arithmetic, and 9 were assigned a three-letter code category in biology but a two-letter code category in arithmetic. Among the 19 respondents with different three-letter code categories, 7 provided stronger arguments in mathematics (e.g., CEX in mathematics and AWX in biology), and 24 provided stronger arguments in biology (e.g., CEX in mathematics and GSX in biology).

DISCUSSION

This study enriched the findings about possible use of Concept Cartoons in the professional preparation of primary school teachers by focusing on Concept Cartoons from the perspective of argumentation in both mathematics and science. In the context of argumentation about general statements, we brought an approach that is joint for elementary mathematics (arithmetic, geometry) and science (biology). The main part of the study arose from three sub-studies located in arithmetic, in arithmetic and biology, and in geometry. Although the arithmetic sub-study was the first one from which the other two gradually devolved, eventually it was the geometry sub-study that turned out to be the key one for understanding the interrelations between reasoning in arithmetic and biology that appeared in our data.

Regarding the first research question

As an answer to the first research question, "*What kinds of reasoning about general statements in biology and mathematics can be observed in future primary school teachers when using Concept Cartoons as a diagnostic instrument?*" we may state that the results themselves are promising from the perspective of participants' knowledge as well as from the perspective of applicability of Concept Cartoons as a diagnostic instrument.

As for participants' knowledge displayed in responses to the three Concept Cartoons, the findings are partially in accordance with previous research studies. Like in (Simon and Blume, 1996), some of the participants provided responses that could not be considered justifications since being vague or unclear (category VR) but these responses accounted for only 3 of the 170 assignments made in the main study. Unlike in (ibid), none of the participants presented a response that could be considered an external conviction proof scheme. Some of the participants showed erroneous knowledge of mathematical or science concepts and their properties (category EC; 13 of 170 assignments in the main study), mainly related to critical moments in learning (Confrey and Kazak, 2006; Lazarowitz and Lieb, 2006). Unlike in (Trowbridge and Mintzes, 1988), none of the misconceptions in biology was based on determining large marine mammals as fish based on the fact that they live in the water. None of the participants confirmed the findings of Zazkis and Chernoff (2008) since they widely used counter-examples but did not hesitate to refute the statements on their basis. Like in (Martin and Harel, 1989), some of the counter-examples in arithmetic were big-number examples (e.g., S32/Jan). In mathematics as well as biology, none of the participants tried to affirm the validity of a general statement through several confirming examples.

From the perspective of the learning process in argumentation, we distinguished between three- and two-letter acronyms for code categories. The three-letter ones (139 of the 170 assignments) referred to responses that might have been considered a constructive basis for further learning of proper argumentation, the two-letter ones (31 of the 170 assignments) referred to blank responses or responses that were not compatible with proper argumentation. Similarly, like in Buchbinder and Cook (2018), some of the responses incompatible with proper argumentation were caused by erroneous knowledge of logical aspects (category EL), namely by confusing assumptions and conclusions in an informally worded implication (Hadar, 1977). Others were caused by erroneous knowledge of concepts and their properties (category EC), for instance by natural number bias (Alibali and Sidney, 2015) or by taxonomy determined according to the habitat (Kubiatko and Prokop, 2007).

As for the Concept Cartoons in the role of a diagnostic instrument, the willingness to respond to the bubbles was high among the participants: only 10 of the 170 responses were blank (category XR). A wide range of proof scheme types and subtypes appeared in data, densely filling the content-related code categories (AWX, CEX, GSX, GEX, GSD, COX, COD, EC; 152 of the 170 assignments). From the perspective of Simon and Blume's (1996) levels of future teachers' argumentation, most of the three-letter assignments could be considered generic examples (level 3; categories GSX, GEX; 86 of the 170 assignments). They were followed by empirical demonstrations (level 2; category CEX; 30 of the 170 assignments), and instance-independent deductive justifications (level 4; categories GSD, COD; 4 of the 170 assignments). Some of the three-letter responses were too incomplete to be clearly related to a particular level (categories AWX, COX; 14 of the 170 assignments).

The three Concept Cartoons differed not only in the subject domains where they were situated but also in other aspects, and these differences influenced collected data. For instance, the geometric Concept Cartoon was the only one presenting informally worded implications that were not true but their converse versions were. Therefore, it provided an opportunity to investigate whether the respondents distinguished properly between assumptions and conclusions; some of the

Electronic ISSN 1803-1617 Printed ISSN 2336-2375

respondents failed in it (category CI; 14 of the 87 assignments in the geometry sub-study). On the other hand, the arithmetic Concept Cartoon was the only one able to stimulate the respondents to indicate conditions of validity and justify them (category COD; 17 of the 112 assignments in the arithmetic sub-study). Such differences in data can direct possible further research concerning argumentation and Concept Cartoons, e.g., at focusing on relations between detailed characteristics of Concept Cartoons and modes of argumentation revealed in data when using the Concept Cartoons for collecting data.

Regarding the second research question

As an answer to the second research question, "*What are the possibilities of joint assessment of reasoning about general statements in biology and mathematics?*" we may also state that the results are promising. The combination of the three Concept Cartoons from arithmetic, geometry and biology showed that there is a possibility to apply a joint approach to the assessment of reasoning in the three subject domains. As illustrated in the conversion table (Table 8), there were four three-letter categories in our data that had a common description for all three subject domains (Table 6). Their respective excerpts appeared in all three domains (AWX, CEX, GSX, GEX; 123 of the 170 assignments). These categories covered the field of instance-dependent deductive justifications (level 3 according to Simon and Blume, 1996), including its initial stages (AWX, CEX and GSX might refer to initial stages of GEX), and were mutually related (GEX = CEX + GSX).

The field of instance-independent deductive justifications (level 4 according to Simon and Blume, 1996) was represented in data only rarely (categories GSD, COD; 4 of the 170 assignments) and separately (GSD just in geometry, COD just in arithmetic). Even the involvement of additional participants who were future lower-secondary school teachers specialising in mathematics and/or biology did not help increase these two categories' occurence. Even if we considered not only the main study but also the sub-studies, there would be only one other category in data that might be considered of the level 4: code category GED with 5 assignments. This category appeared in the part of the arithmetic sub-study that was not common to the main study. However, the GED category also did not appear in other subject domains.

The current study did not solve the issue of a joint assessment of instance-independent deductive reasoning (level 4 according to Simon and Blume, 1996) due to its sparse occurrence in data. The question is whether the next step in a systematic joint approach to the level-4 terrain should consist in approaching domain specialists (mathematicians, biologists) or in changing the Concept Cartoons for data collection.

From the perspective of individual participants, the joint set of code categories provided by this research can be used for investigating modes of reasoning about general statements comparably in mathematics and biology. During the arithmetic-biology sub-study, individual participants provided modes of reasoning that differed in arithmetic and biology. Such a difference might have been caused by different levels of participant's content knowledge in the two subject topics but also by different approaches to argumentation that the particular participant might have applied within the two subject contexts. Additional questions and/or additional diagnostic tasks would be needed to identify the specific cause of the difference. However, the second cause reveals opportunities for a joint development of argumentation skills: stemming from the comparison of arguments provided in the two subjects and building on the approach applied in the subject with stronger arguments.

Regarding the implications for argumentation and formative assessment

Since generic examples can be considered the first step of deductive reasoning (Harel and Sowder, 2007; Simon and Blume, 1996) and data collected within our study covered the field of generic examples, it is possible to use the categories arisen from data analysis for scaffolding the process of learning the concept of generic examples. Consequently, the categories can also be used in learning the foundations of deductive reasoning. The possibility to label individual steps of the learning process by code categories then may allow teachers to orient themselves better in the course of formative assessment and also in actual levels of students' knowledge that is to be addressed by this assessment. In that sense, it may indirectly help the teachers promote their noticing of various reasoning forms (Melhuish, Thanheiser and Guyot, 2020).

More generally, we showed that there is a possibility for argumentation and formative assessment to be understood equally in mathematics and biology. Such a finding complements our recent study that has resulted in a joint communication model for inquiry and formative assessment in mathematics and biology (Rokos and Samková, 2020).

Regarding the implications for professional preparation of primary school teachers

The joint coding procedure for biology and mathematics offers a diagnostic tool for assessing future primary school teachers' modes of reasoning comparably in both school subjects. This arrangement provides a common background for teacher educators in mathematics and teacher educators in biology that allows them to collaborate on the planning as well as assessment of their respective teacher training courses. In that sense, the procedure enriches subject didactic of mathematics as well as subject didactic of biology (Kubiatko, 2021).

The presented research participants were future primary school teachers in a professional preparation program that simultaneously approaches the study of content and didactics in mathematics and biology. Within this environment, it is possible to intensively utilize a joint approach to argumentation and reasoning in the two school subjects. As mentioned above, any differences identified between the modes of future teachers' reasoning in the two subjects might provide an opportunity for the development of future teachers' argumentation skills and thus positively affect their professional knowledge (e.g., by enhancing the quality of their feedback; Hattie and Timperley, 2007).

The joint approach to reasoning about general statements in mathematics and biology is valuable for future primary school teachers not only as an example related to formative assessment within inquiry-based science and mathematics education practices (Dolin and Evans, 2018) but also as an example related to other educational frameworks utilizing the integration of mathematics and biology school subjects. For instance, to the STEM framework that integrates science, technology, engineering and mathematics education (Moore, Johnson and Glancy, 2020). In this sense, our paper contributes to the educational research on STEM by addressing the objection given by English (2016) that the STEM-related research mostly focuses on STEM integration just from the point of view of individual school subjects. We focus jointly on biology and mathematics, and from a general perspective of argumentation, so that we allow to approach the two school subjects commonly despite their different contents and different approaches to didactics (Hallström and Schönborn, 2019). Unlike some of the research on mathematics within STEM education, we do not perceive mathematics just as a computational tool for other STEM subjects (Valovičová et al., 2020) but as an equivalent subject identity. Such an arrangement forms a suitable ground for supporting and developing STEM-oriented primary (or elementary) school teacher education (Corp, Fields and Naizer, 2020).

CONCLUSION

In this study, we focused on future primary school teachers and their modes of argumentation in mathematics and science, namely in arithmetic, geometry and biology. For each of the subject fields, we prepared a Concept Cartoon focusing on general statements, collected data using them, and then conducted qualitative data analysis looking for displays of various levels of argumentation and for possibilities to assess these displays jointly in mathematics and science. Our effort resulted in a set of code categories for various levels of instance-dependent deductive reasoning (including its initial stages) and in a conversion table that provided the framework for investigating and comparing modes of argumentation of individual respondents across the three subject domains (arithmetic, geometry, biology). Such a joint approach offers the opportunity to understand and develop respondents' knowledge in a deeper way by approaching the same concept (argumentation) jointly within three different subject domains. From a broader perspective, we entered the issue of a joint approach to argumentation in mathematics and biology in relation to inquiry-based education as an attempt to perceive inquiry in a comparable manner in both subjects. Since inquiry-based education is naturally rich in generalisations and in formulating and discussing general statements, our study's findings provide a highly applicable framework in the inquiry-based environment. For future research, it might be valuable to gain a more general perspective by extending the approach to other science subject domains (physics, chemistry) or studying the approach also in relation to other STEM subjects (technology, engineering). These additional subjects and subject domains might be represented in primary school content and/or professional preparation of primary school teachers to a lesser extent than mathematics and biology, however, the joint approach to argumentation might also be investigated in relation to secondary school education.

ACKNOWLEDGEMENT

This article was supported by the Technology Agency of the Czech Republic under Grant 'Learning Hyperspace for Formative Assessment and Inquiry Based Science Teaching', project No. TL02000368.

REFERENCES

Alibali, M. W. and Sidney, P. G. (2015) 'Variability in the natural number bias: who, when, how, and why', *Learning and Instruction*, Vol. 37, pp. 56–61. http://doi.org/10.1016/j.learninstruc.2015.01.003

Artigue, M. and Blomhøj, M. (2013) 'Conceptualizing inquiry-based education in mathematics', *ZDM Mathematics Education*, Vol. 45, No. 6, pp. 797–810. http://doi.org/10.1007/s11858-013-0506-6

Balacheff, N. (1988) 'Aspects of proof in pupils' practice of school mathematics', in Pimm, D. (ed.) *Mathematics, Teachers and Children*, London: Hodder & Stoughton, pp. 216–238.

Berland, L. K. and Reiser, B. J. (2009) 'Making sense of argumentation and explanation', *Science Education*, Vol. 93, No. 1, pp. 26–55. http://doi.org/10.1002/sce.20286

Berthelsen, B. (1999) 'Students' naive conceptions in life sciences', *MSTA Journal*, Vol. 44, No. 1, pp. 13–19.

Braund, M. (1997) 'Primary children's ideas about animals with and without backbones', *Education 3-13*, Vol. 25, No. 2, pp. 19–24. http://doi.org/10.1080/03004279785200191

Buchbinder, O. and Cook, A. (2018) 'Examining the mathematical knowledge for teaching of proving in scenarios written by pre-service teachers', in Buchbinder, O. and Kuntze, S. (ed.) *Mathematics Teachers Engaging with Representations of Practice*, Cham: Springer, pp. 133–154. http://doi.org/10.1007/978-3-319-70594-1_8

Bulková, K., Medová, J. and Čeretková, S. (2020) 'Identification of crucial steps and skills in high-achievers' solving complex mathematical problem within mathematical contest', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 13, No. 2, pp. 67–78. http://doi.org/10.7160/eriesj.2020.130202

Clipart Library (2019) *Ocean floor black and white*, [Online], Available: http://clipart-library.com/coloring/kTMegybTj.gif [14 Nov 2020].

Confrey, J. and Kazak, S. (2006) 'A thirty-year reflection on constructivism in mathematics education in PME', in Gutiérrez A. and Boero P. (ed.) *Handbook of Research on the Psychology of Mathematics Education: Past, Present and Future*, Rotterdam: Sense, pp. 305–345.

Corp, A., Fields, M. and Naizer, G. (2020) 'Elementary STEM teacher education: Recent practices to prepare general elementary teachers for STEM', in Johnson C. C., Mohr-Schroeder, M. J., Moore, T. J. and English, L. D. (ed.) *Handbook of Research on STEM Education*, New York: Routledge, pp. 337–348. http://doi.org/10.4324/9780429021381-32

Cullen, S., Fan, J., van der Brugge, E. and Elga, A. (2018) 'Improving analytical reasoning and argument understanding: a quasi-experimental field study of argument visualization', *npj Science of Learning*, Vol. 3, No. 21. http://doi.org/10.1038/s41539-018-0038-5

Dabell, J., Keogh, B. and Naylor, S. (2008) *Concept Cartoons in Mathematics Education*, Sandbach: Millgate House Education.

Dofner, T., Förtsch, C., Germ, M. and Neuhaus, B. J. (2018) 'Biology instruction using generic framework of scientific reasoning and argumentation', *Teaching and Teacher Education*, Vol. 75, pp. 232–243. http://doi.org/10.1016/j.tate.2018.07.003

Dolin, J. and Evans, R. (ed.) (2018) *Transforming assessment: Through an interplay between practice, research and policy*, Cham: Springer. http://doi.org/10.1007/978-3-319-63248-3

van Dooren, W., Lehtinen, E. and Verschaffel, L. (2015) 'Unraveling the gap between natural and rational numbers', *Learning and Instruction*, Vol. 37, pp. 1–4. http://doi.org/10.1016/j.learninstruc.2015.01.001

Dorier, J.-L. and Maass, K. (2014) 'Inquiry based mathematics education', in Lerman S. (ed.) *Encyclopedia of Mathematics Education*, Dordrecht: Springer, pp. 300–304. http://doi.org/10.1007/978-94-007-4978-8_176

English, L. D. (2016) 'STEM education K-12: perspectives on integration', *International Journal of STEM Education*, Vol. 3, No. 3. http://doi.org/10.1186/s40594-016-0036-1

Erduran, S. and Jiménez-Aleixandre, M. P. (ed.) (2007) *Argumentation in science education: Perspectives from classroom-based research*, Dordrecht: Springer. http://doi.org/10.1007/978-1-4020-6670-2

Fujita, T. and Jones, K. (2007) 'Learners' understanding of the definitions and hierarchical classification of quadrilaterals: Towards a theoretical framing', *Research in Mathematics Education*, Vol. 9, No. 1, pp. 3–20. http://doi.org/10.1080/14794800008520167

Furtak, E. M., Hardy, I., Beinbrech, C., Shavelson, R. J. and Shemwell, J. T. (2010) 'A framework for analyzing evidence-based reasoning in science classroom discourse', *Educational Assessment*, Vol. 15, No. 3–4, pp. 175–196. http://doi.org/10.1080/10627197.2010.530553

Galbraith, P. L. (1981) 'Aspects of proving: A clinical investigation of process', *Educational Studies in Mathematics*, Vol. 12, No. 1, pp. 1–28. http://doi.org/10.1007/BF00386043

Hadar, N. (1977) 'Children's conditional reasoning', *Educational Studies in Mathematics*, Vol. 8, No. 4, pp. 413–438. http://doi.org/10.1007/bf00310946

Hallström, J. and Schönborn, K. J. (2019) 'Models and modelling for authentic STEM education: reinforcing the argument', *International Journal of STEM Education*, Vol. 6, No. 22. http://doi.org/10.1186/s40594-019-0178-z

Harel, G. and Sowder, L. (1998) 'Students' proof schemes: Results from exploratory studies', in Schoenfeld, A. H., Kaput, J., Dubinski, E. and Dick, T. (ed.) *Issues in Mathematics Education (Research in Collegiate Mathematics Education III, Vol. 7)*, Providence: AMS, pp. 234–283. http://doi.org/10.1090/cbmath/007/07

Harel, G. and Sowder, L. (2007) 'Toward comprehensive perspectives on the learning and teaching of proof', in Lester, F. (ed.) *Second Handbook of Research on Mathematics Teaching and Learning*, Charlotte: NCTM, pp. 805–842.

Hattie, J. and Timperley, H. (2007) 'The power of feedback', *Review of Educational Research*, Vol. 77, No. 1, pp. 81–112. http://doi.org/10.3102/003465430298487

Kattmann, U. (2001) 'Aquatics, flyers, creepers and terrestrials – Students' conceptions of animal classification', *Journal of Biological Education*, Vol. 35, No. 3, pp. 141–147. http://doi.org/10.1080/00219266.2001.9655763

Keogh, B. and Naylor, S. (1999) 'Concept Cartoons, teaching and learning in science: an evaluation', *International Journal of Science Education*, Vol. 21, No. 4, pp. 431–446. http://doi.org/10.1080/095006999290642

Komatsu, K. (2010) 'Counter-examples for refinement of conjectures and proofs in primary school mathematics', *Journal of Mathematical Behavior*, Vol. 29, No. 1, pp. 1–10. http://doi.org/10.1016/j.jmathb.2010.01.003

Kubiatko, M. (2021) 'Subject didactics: relevant issues', *Problems of Education in the 21st Century*, Vol. 79, No. 3, pp. 340–342. http://doi.org/10.33225/pec/21.79.340

Kubiatko, M. and Prokop, P. (2007) 'Pupils' misconceptions about mammals', *Journal of Baltic Science Education*, Vol. 6, No. 1, pp. 5–14.

Lazarou, D., Sutherland, R. and Erduran, S. (2016) 'Argumentation in science education as a systemic activity: An activity-theoretical perspective', *International Journal of Educational Research*, Vol. 79, pp. 150–166. http://doi.org/10.1016/j.ijer.2016.07.008

Lazarowitz, R. and Lieb, C. (2006) 'Formative assessment pre-test to identify college students' prior knowledge, misconceptions and learning difficulties in biology', *International Journal of Science and Mathematical Education*, Vol. 4, No. 4, pp. 741–762. http://doi.org/10.1007/s10763-005-9024-5

Martin, W. G. and Harel, G. (1989) 'Proof frames of preservice elementary teachers', *Journal for Research in Mathematics Education*, Vol. 20, No. 1, pp. 41–51. http://doi.org/10.2307/749097

McComas, W. F. (ed.) (2002) *The nature of science in science education*, Boston: Kluwer. http://doi.org/10.1007/0-306-47215-5

Melhuish, K., Thanheiser, E. and Guyot, L. (2020) 'Elementary school teachers' noticing of essential mathematical reasoning forms: justification and generalization', *Journal of Mathematics Teacher Education*, Vol. 23, pp. 35–67. http://doi.org/10.1007/s10857-018-9408-4

Miles, M. B., Huberman, A. M. and Saldaña, J. (2014) *Qualitative data analysis: A methods sourcebook*, 3rd edition, Thousand Oaks, CA: SAGE.

Minner, D., Levy, A. and Century, J. (2010) 'Inquiry-based science instruction – what is it and does it matter? Results from a research synthesis years 1984 to 2002', *Journal of Research in Science Teaching*, Vol. 47, No. 4, pp. 474–496. http://doi.org/10.1002/tea.20347

Moore, T. J., Johnson, A. C. and Glancy, A. W. (2020) STEM integration: A synthesis of conceptual frameworks and definitions', in Johnson C. C., Mohr-Schroeder, M. J., Moore, T. J. and English, L. D. (ed.) *Handbook of Research on STEM Education*, New York: Routledge, pp. 3–16. http://doi.org/10.4324/9780429021381-2

National Research Council (1996) *National science education standards*, Washington, DC: National Academy Press.

Naylor, S. and Keogh, B. (2010) *Concept Cartoons in Science Education*, 2nd Edition [CD-ROM], Sandbach: Millgate House Education.

Naylor, S., Keogh, B. and Downing, B. (2007) 'Argumentation and primary science', *Research in Science Education*, Vol. 37, No. 1, pp. 17–39. http://doi.org/10.1007/s11165-005-9002-5

Ping, I. L. L., Halim, L. and Osman, K. (2020) 'Explicit teaching of scientific argumentation as an approach in developing argumentation skills, science process skills and biology understanding', *Journal of Baltic Science Education*, Vol. 19, No. 2, pp. 276–288. http://doi.org/10.33225/jbse/20.19.276

Prokop, P., Prokop, M. and Tunnicliffe, S. D. (2007) 'Effects of keeping animals as pets on children's concepts of vertebrates and invertebrates', *International Journal of Science Education*, Vol. 30, No. 4, pp. 431–449. http://doi.org/10.1080/09500690701206686

Rocard, M., Csermely, P., Jorde, D., Lenzen, D., Walberg-Henriksson, H. and Hemmo, V. (2007) *Science education now: a renewed pedagogy for the future of Europe*, Luxembourg: Office for Official Publications of the European Communities, [Online], Available: http://www.eesc.europa.eu/resources/docs/rapportrocardfinal.pdf [4 Jun 2021].

Rokos, L. and Samková, L. (2020) 'Coding classroom talk from the perspective of formative assessment and inquiry-based education: a communication model for mathematics and science lessons', in *Proceedings of ICERI2020 Conference*, Seville, pp. 1170–1179. http://doi.org/10.21125/iceri.2020.0318

Roubíček, F. (2014) *The set of four geometric Concept Cartoons for assessing future primary school teacher's knowledge*, [Internal material, unpublished].

Samková, L. (2019) 'Investigating subject matter knowledge and pedagogical content knowledge in mathematics with the Concept Cartoons method', *Scientia in educatione*, Vol. 10, No. 2, pp. 62–79. http://doi.org/10.14712/18047106.1548

Samková, L. (2020) 'Observing how future primary school teachers reason about general statements', in *Proceedings of the 17th International Conference on Efficiency and Responsibility in Education (ERIE 2020)*, Prague, pp. 263–271.

Samková, L. and Tichá, M. (2017) 'On the way to observe how future primary school teachers reason about fractions', *Journal of Efficiency and Responsibility in Education and Science*, Vol. 10, No. 4, pp. 93–100. http://doi.org/10.7160/eriesj.2017.100401

Selden, A. (2012) 'Transitions and proof and proving at tertiary level', in Hanna, G. and de Villiers, M. (ed.) *Proof and Proving in Mathematics Education*, Dordrecht: Springer, pp. 391–420. http://doi.org/10.1007/978-94-007-2129-6_17

Simon, M. A. and Blume, G. W. (1996) 'Justification in the mathematics classroom: A study of prospective elementary teachers', *Journal of Mathematical Behavior*, Vol. 15, No. 1, pp. 3–31. http://doi.org/10.1016/s0732-3123(96)90036-x

Stevens, A. L., Wilkins, J. L. M, Lovin, L. H., Siegfried, J., Norton, A. and Busi, R. (2020) 'Promoting sophisticated fraction constructs through instructional changes in a mathematics course for PreK-8 prospective teachers', *Journal of Mathematics Teacher Education*, Vol. 23, No. 2, pp. 153–181. http://doi.org/10.1007/s10857-018-9415-5

Trowbridge, J. E., and Mintzes, J. J. (1988) 'Alternative conceptions in animal classification: A cross-age study', *Journal of Research in Science Teaching*, Vol. 25, No. 7, pp. 547–571. http://doi.org/10.1002/tea.3660250704

Tuset, G. A. (2019) 'Preservice teachers' geometrical discourses when leading classroom discussions about defining and classifying quadrilaterals', in *Proceedings of CERME11*, Utrecht, pp. 3523–3530.

Valovičová, Ľ., Ondruška, J., Zelenický, Ľ., Chytrý, V. and Medová, J. (2020) 'Enhancing computational thinking through interdisciplinary STEAM activities using tablets', *Mathematics*, Vol. 8, No. 12, 2128. http://doi.org/10.3390/math8122128

de Villiers, M. (1994) 'The role and function of a hierarchical classification of quadrilaterals', *For the Learning of Mathematics*, Vol. 14, No. 1, pp. 11–18.

Vízek, L. and Samková, L. (2021) 'Observing how future primary school teachers reason about quadrilaterals', in *Proceedings of the 18th International Conference on Efficiency and Responsibility in Education (ERIE 2021)*, Prague, pp. 159–167.

Yen, C.-F., Yao, T.-W. and Chiu, Y.-C. (2004) 'Alternative conceptions in animal classification focusing on amphibians and reptiles: A cross-age study', *International Journal of Science and Mathematics Education*, Vol. 2, No. 2, pp. 159–174. http://doi.org/10.1007/s10763-004-1951-z

Zazkis, R. and Chernoff, E. J. (2008) 'What makes a counterexample exemplary?', *Educational Studies in Mathematics*, Vol. 68, pp. 195–208. http://doi.org/10.1007/s10649-007-9110-4