ORIGINAL ARTICLE

Fostering first-year pre-service teachers' proof competencies

Leander Kempen[1] · Rolf Biehler[1]

Abstract

In the study presented, as we report in this paper, we describe our theoretical and practical consideration to engage first-year pre-service teachers in proving activities in the context of a transition-to-proof course. We investigated how students argued to verify a claim of elementary number theory on entering university and compared the results to their performance in the final examination of the course. Subsequently, we elaborate on the following results: On entering university, students do not seem to be capable of using algebraic variables as a heuristic to engage in reasoning. However, after learning about different kinds of proofs and the symbolic language of mathematics, students give evidence of starting to value mathematical language and of enhancing their proof competencies.

Keywords Transition to university · Generic proof · Variables · Proving

1 Introduction

Since mathematical proof constitutes a fundamental concept in mathematics, the learning about proof and proving has to play an essential role in university courses on mathematics. The University of Paderborn offers the course "Introduction into the culture of mathematics" to help pre-service teachers to accomplish the transition to university and to come to terms with mathematical proof. In this context, the symbolic language of mathematics plays a unique role. As Mason et al. (2005) stress, algebra is about expressing generality. Accordingly, the language of algebra has to be used and promoted in this context. However, the use of this 'language' has to be learned, too, as school algebra has not necessarily stressed the aspect Mason et al. are highlighting. One aim of the course mentioned above is to introduce and to promote the use of the symbolic language of algebra in a meaningful way. Moreover, we aim to provide the pre-service teachers with kinds of proofs that they can use in class later on.

First, we describe our course concept for pre-service teachers (lower secondary school). For the design of the course, the way students reason on entering university is most important, and we present the results of our relevant investigations. Finally, we analyze students' reasoning[1] at the end of the course for the purpose of describing changes

compared to their proving attempts at the beginning of the course. Accordingly, we reflect on the impact of the course. The research presented here is part of a broader research project in the context of the dissertation of the first author under the supervision of the second author (Kempen 2019).

2 Theoretical background

In this section, we recapitulate common problems with mathematical proof students have on entering university, and describe possible differences concerning mathematical proof at school and university (Sect. 2.1). We elaborate on our concept of generic proofs and discuss the aspect of evidence (Sect. 2.2). Finally, we highlight some features of the symbolic language of mathematics that were emphasized in our course (Sect. 2.3).

2.1 Students' proof competencies in the transition to university

By summarizing international studies, Selden (2012, p. 398 ff.) extracts the following problems of first-year students with mathematical proof: the correct use of symbolic mathematical language, a nonstandard view of proof (e.g., what constitutes a proof), the selection of helpful

✉ Leander Kempen
  kempen@khdm.de

1 Universität Paderborn, Paderborn, Germany

---

[1] In this paper, "reasoning" is used as a superordinate concept that comprises all kinds of possible approaches to verify a claim.

representations when proving, and the knowledge on how to read and check proofs. Here the question arises how students' prior experiences with mathematical proof relate to the requirements at university. As Hemmi (2006, p. 132 ff.) and Kempen (2019, p. 244 ff.) found, students have little prior experience with mathematical proof when entering university. Following the results of Kempen's (2019, p. 246 ff.) research, first-year students mainly link the concept of proof with the proof of Thales' theorem, the Pythagorean theorem and the binomial formulas. Here, the differences concerning the concept of proof at school and at university level become apparent: Particularly in school geometry, proofs make use of a figure to perform deductive reasoning. In elementary arithmetic, many proofs utilize simple calculations using variables, as, for example, the proofs of the binomial formulas do. Neither definitions nor theorems are used to perform reasoning. First-year students' views of proof seem to be mainly connected to the use of a proof figure and simple calculations performed on given 'letters'. However, when constructing mathematical proofs at university, even in a course for first-years, students need to make use of definitions and to apply theorems about abstract concepts and to explicitly perform deductive reasoning (Selden and Selden 2007).

The following questions arise: How do first-year students argue when being asked to verify a statement? Are they used to performing deductive reasoning when constructing a proof? Do the students make use of symbolic language when proving a claim? If so, do they struggle with its correct use? We address these questions in this paper.

2.2 Generic proof

Starting with findings in students' proof attempts (Balacheff 1988) and the distinction between generality and genericity (Mason and Pimm 1984), the idea of generic proofs has become popular in the international discussion about proof and proving.

> A generic proof aims to exhibit a complete chain of reasoning from assumptions to conclusion, just as in a general proof; however, […] a generic proof makes the chain of reasoning accessible to students by reducing its level of abstraction; it achieves this by examining an example that makes it possible to exhibit the complete chain of reasoning without the need to use a symbolism that the student might find incomprehensible (Dreyfus et al. 2012, p. 204).

Mason and Pimm give the following figure as an example of a generic proof (see Fig. 1).

The above generic proof consists only of one concrete example making use of figurate numbers (dots). In this
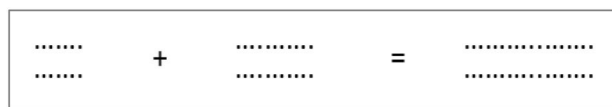

Fig. 1 A generic proof for the claim that the sum of two even numbers is always even (Figure similar to Mason and Pimm 1984, p 284)

example, the reader has to detect the overall scheme, which is also described by the authors:

> It serves to remind us of an image or perception of even numbers as numbers which can be displayed as two matching rows of dots. Since in both numbers the dots pair up, so too will they in the sum, formed by amalgamating the dots (p. 284).

It is this overall scheme that makes the examples go beyond empirical evidence. As Reid and Vallejo Vargas (2018, p. 247) point out, two kinds of evidence have to be detected in a generic proof to pass beyond empirical evidence and to achieve a valid general proof: Evidence of generality and evidence of reasoning. The first kind of evidence is about the awareness that the scheme detected in the concrete examples has a general quality. The second kind of evidence "points to the mathematical reasons for why the structure can be extrapolated for other cases from the example(s) given, and it is based […] on the ground knowledge the community shares at that point" (p. 247). Accordingly, a generic proof comprises concrete examples so that the reader can detect a general overall scheme. Due to this general ("generic") scheme, these examples are also called "generic examples". The general overall scheme makes the generic examples go beyond empirical evidence and form a proper proof. (The question if generic proofs can be considered as valid mathematical proofs is still discussed in the community, see Reid and Vallejo Vargas 2018, and Leron and Zaslavski 2013). In sum, a generic proof consists of (1) a presentation of generic examples and (2) a general argument that verifies the claim in general. Following these considerations, Biehler and Kempen (2013) suggest a conceptual refinement concerning the distinction between generic examples and generic proofs. For a generic proof, the detected argument and its generality, given in the context of one or more concrete (generic) examples, have to be explicated for the reader. We provide an example of a generic proof for the claim that the sum of an odd natural number and its double is always odd:

Generic proof:

$$1 + 2 \cdot 1 = 3 \cdot 1 = 3$$

$$17 + 2 \cdot 17 = 3 \cdot 17 = 51$$

Comparing the equations, one can recognize that the result must always be three times the initial number. Since three

times an odd number is always odd, the result is an odd number.

The explanation of the general argument seems to be necessary to complete the proof because the generic examples point only to a general idea that is 'hidden' in a concrete context. The narrative explicates the writer's thoughts and arguments for the reader. The concrete numbers used in generic examples serve as 'variables', as a particular representation of a class. This is not only true for the numbers used in concrete examples, but also for the concrete figurate numbers used in Mason and Pimm's example (see Fig. 1). When describing the overall argument in a generic proof, one has to refer to these implicit variables. One possibility is to use word variables for referring to a general term like 'a natural number', 'its square', 'its successor' or 'two matching rows of dots'. Also, word variables referring to a universal quantification such as 'always' can stress the generality of the argument.

The concept of generic proofs is closely related to other concepts of proof in mathematics education (see Biehler and Kempen 2016, for an overview). In the concept of *operative proofs* (Wittmann 2009), the exploration of the mathematical problem by performing operations on "quasi-real" mathematical objects (p. 254) plays a crucial role. Here, students are to detect an invariant aspect when performing the operations on concrete objects. This invariant aspect serves as the overall argument in the proof. In this sense, operative proofs can be considered to be generic proofs: The concrete examples on which the operations are performed serve as generic examples, and the invariant aspect of the operations completes the generic proof. However, there are generic proofs that do not fit the concept of operative proofs. For example, Tall (1979) discusses the proof by contradiction of the irrationality of $\sqrt{2}$. This proof makes use of the aspect that in a possible representation of $\sqrt{2}$ as a rational number, the numerator and the denominator will always have a common divisor. This argument can be considered as a generic proof for the claim that the square root of a prime number is always irrational. Since the proof does not arise from an operative setting, this proof can hardly be considered to be an operative proof.

However, one has to be careful about the question of whether a scheme detected in concrete examples may be transferred to all possible cases and therefore can function as a 'general' argument. Following this problem, the relationship between generic proofs and the concept of preformal proofs (Blum and Kirsch 1991) has to be considered, too. In order to be sure about the correctness of an argument detected in a concrete context, Blum and Kirsch require that the argument used can be formalized. In this sense, the proof becomes "preformal". However, in contrast to a generic proof, preformal proofs may also make use of arguments from physics or experiences from the real world (cf. Blum and Kirsch 1991; Blum 1998).

Generic proofs seem to offer several advantages in the learning of mathematical proof, e.g.:

- Generic proofs offer the possibility of constructing mathematical proofs without the use of mathematical symbols. Because of this aspect, generic proofs seem to be appropriate kinds of proofs for mathematics courses at school.
- In the context of generic proofs, the investigation of concrete examples gets highlighted and becomes a natural part of the proving process. These activities may lead to a better understanding of the given claim. Besides, the difference between purely empirical verifications and general proofs can be highlighted.

Despite the numerous theoretical considerations made concerning the concept of generic proofs in the literature, only a little effort has been made to investigate students' actual handling of generic proof. Karunakaran et al. (2014) use generic proofs to foster pre-service teachers' proof competencies. However, the students in their study struggle with the link between the particular examples and the general case. Here, the need for further investigation appears: How do students deal with the concept of generic proof, and do students grasp the difference between purely empirical verifications and generic proofs? For investigating students' generic proof constructions, new research instruments are needed.

2.3 Learning the symbolic language of mathematics

Considering concrete examples as generic examples, the concrete numbers used serve as a kind of 'variable'; the numbers used in the generic proof given above serve as a "characteristic representation of its class" (Balacheff 1988, p. 219). However, since in this case the generality is hidden in a concrete context, a generic proof can easily be compared to a proof making use of (algebraic) variables:

Let $n \in \mathbb{N}$. We have:

$$(2n - 1) + 2 \cdot (2n - 1) = 6n - 3 = 2 \cdot (3n - 1) - 1,$$

where $(3n - 1) \in \mathbb{N}$.

So the result is an odd number QED.

If one agrees with this use of variables, there is no further need for explication. The use of the variable n and the correct computations ensure the generality and validity of the proof. It is this use of algebraic variables for formulating a claim and proving it, that we consider to be an appropriate image of the symbolic language of mathematics for first-year pre-service teachers at the beginning of our course.

This comparison of a generic proof with numbers to the corresponding proof with algebraic variables serves as an example for how to highlight the advantages and benefits of symbolic language (cf. Malle 1993; Mason et al. 2005, p. 1 ff.):

1. Algebra makes it possible to communicate general incidents. While the generality detected in concrete examples has to be 'advocated', the use of algebraic variables is about communicating general incidents.
2. The symbolic language of mathematics inherits a control function concerning the validity and generality of arguments. When making use of concrete examples to find an argument, the question arises if the arguments found rely on specific properties of the concrete numbers used. Accordingly, the argument cannot be considered as general or valid. However, when performing the computations used in the generic examples again by using algebraic variables, the result affirms the validity and generality of the arguments found.
3. The symbolic language of mathematics may lead to total conviction and generality because the computation making use of variables conveys both kinds of evidence mentioned above: the evidence of awareness of generality and the mathematical evidence of reasoning. Arguments with the use of other symbols, numbers or diagrams may not achieve this conviction.

As mentioned above, students at university level also struggle with the correct use of variables (Selden and Selden 2007). Accordingly, we also aim to investigate how far students' way of using variables interferes with their proof constructions.

3 The conception of the course "Introduction into the culture of mathematics"

We now give a short example from the course "Introduction into the culture of mathematics" (Sect. 3.1) and highlight four different kinds of proof (Sect. 3.2). Finally, we introduce some specific tasks to foster students' proof competencies and to provide experiences with different notational systems (Sect. 3.3).

3.1 Investigations of examples and different kinds of proof

The course starts with the problem "Someone claims the sum of three consecutive numbers is always divisible by three. Is this correct? If so, why?" We provide students with three different 'strategies' for dealing with this claim: (1)

testing the claim with several examples, (2) testing the claim with several examples to get some insights and to find a generic argument that can be used in a generic proof and (3) formalizing the claim and performing algebraic manipulations to find an argument.

Strategy (1): Testing the claim with several examples.[2]

$1 + 2 + 3 = 6$ is divisible by three. $500 + 501 + 502 = 1503$ is divisible by three. The claim seems to be true.

Strategy (2): Testing the claim with several examples to get some insights and to find a generic argument that can be used in a generic proof.

A discovery can be made: The sum is always three times the middle number: $4 + 5 + 6 = 3 \cdot 5$ etc. Why is this the case? Is this always the case? 4 is 1 less than 5, 6 is 5 plus 1, these 1 s compensate each other! This insight can be used to formulate a generic proof (with natural numbers):

$$1 + 2 + 3 = (2 - 1) + 2 + (2 + 1) = 3 \cdot 2$$

$$500 + 501 + 502 = (501 - 1) + 501 + (501 + 1) = 3 \cdot 501$$

One can always write the sum of three consecutive numbers as: (middle number $- 1$) $+$ middle number $+$ (middle number $+1$). *Since this sum equals three times the "middle number", the sum is always divisible by three.*

By formulating the full generic proof, the claim gets justified. Here, we stress the features of a generic proof: For a generic proof one has to detect an overall (generic) argument in the context of concrete examples. This is what makes the examples generic. Then one has to explicate the argument and to explain why it can be transferred to all possible cases (see Sect. 2.2). Here, the differences between purely empirical verifications (strategy 1) and generic proofs can be highlighted.
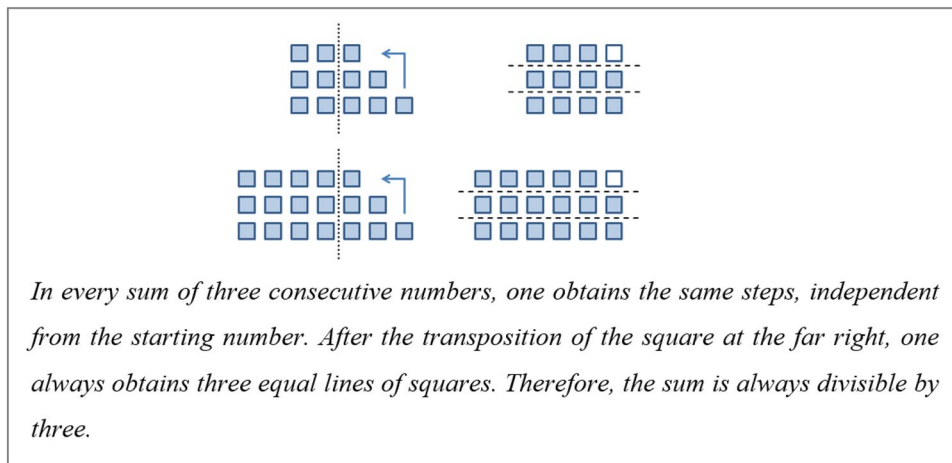
Strategy (3): Formalizing the claim and performing algebraic manipulations.

Another strategy is to introduce algebraic variables to cope with the given claim: Let $m \in \mathbb{N}$ be the initial number, then the sum can be expressed as $m + (m + 1) + (m + 2)$. We use algebraic manipulations and check whether we can rearrange the variables in a way that shows that the sum is a multiple of 3: $m + (m + 1) + (m + 2) = 3m + 3 = 3 \cdot (m + 1)$.

Compared to the generic proof, the formulation of the problem using variables combined with simple calculations may be considered to be easier. However, as we aim to teach our students about the deductive structure of mathematics, the mere use of variables does not suffice. In a formal proof, as advocated in our course, all variables used have to be defined correctly, and references to prior theorems or

---

[2] The testing of the claim helps the students to understand the claim and to get an idea about its truth [cf. the structure of the proving process given in Boero (1999, p. 2)].

Fig. 2 Generic proof with figurate numbers about the sum of three consecutive numbers

In every sum of three consecutive numbers, one obtains the same steps, independent from the starting number. After the transposition of the square at the far right, one always obtains three equal lines of squares. Therefore, the sum is always divisible by three.

definitions have to be stated. Accordingly, the following proof would be considered:

Let $m \in \mathbb{N}$. We have

$$m + (m + 1) + (m + 2) = 3m + 3 = 3 \cdot (m + 1).$$

Since $(m + 1) \in \mathbb{N}$, the result is divisible by three (def. 1.1).[3]

When using algebraic variables, the letters fulfil the task of highlighting an overall scheme and therefore provide the evidence of awareness of generality. Accordingly, there is no need to describe the computations performed with the use of word variables when the reader 'understands' the use of variables. Also, the result of the computations cannot rely on a specific property of the concrete numbers used. In this sense, no extra verbalization and explanation seem to be necessary when using (algebraic) variables in a proof.

These three strategies are also discussed in the notational system of figurate numbers. Accordingly, a generic proof with figurate numbers can be constructed:

A Generic proof with figurate numbers:

In this case, the evidence of awareness of generality is indicated by the vertical lines and the arrows in the two concrete examples (Fig. 2). These aspects are expressed in the narrative by the phrasings "in every sum", "one always obtains", and "is always". The mathematical evidence of reasoning is based on 'intuitive-evident' facts in the context of figurate numbers: When representing the sum of three consecutive natural numbers, each corresponding line will have one more square than the former. When one transfers one square from the longest line to the shortest, all three lines will have the same number of squares. Accordingly, the sum will be divisible by three.

In line with the algebraic variables, we introduce 'geometric (discrete) variables' to represent an 'arbitrary' number of dots or squares (Fig. 3). Now it becomes possible to formulate a proof to the given claim using geometric variables (Fig. 4).

Proof with geometric variables: As is the case in the formal proof, we do not ask for further explanations about the generality when using geometric variables. The variables are meant to express the general validity of the argument. Since there is no reliance on a concrete 'number', the argument necessarily displays a general quality. In this sense, the geometric variable ensures the evidence of awareness of generality. For the mathematical evidence of reasoning, the same aspects mentioned for the generic proof with figurate numbers have to be considered. Taken together, we are using four different kinds of proof, as shown in Fig. 5.

3.2 Tasks for generalizing and for formulating general incidents

Having been presented with the different kinds of proofs, the students have to work on different tasks in which they have to construct the different kinds of proofs.

In the following tasks, students have to identify a typical pattern in the examples. Having detected this pattern, they have to formulate a corresponding general claim using word-variables (to express their finding in more familiar 'language') as well as algebraic variables. When formulating the general claim, students experience several advantages of symbolic mathematical language: When using word-variables they have to look for appropriate terms (words like 'successor') for describing the common pattern, and they have
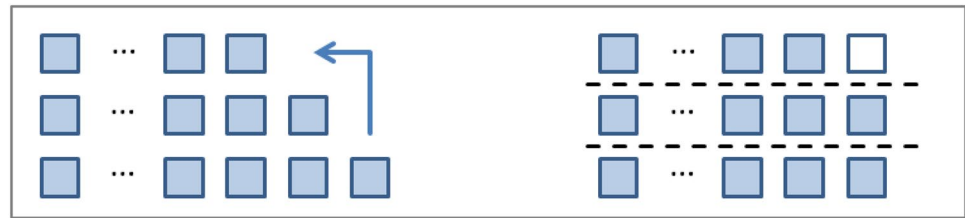
Fig. 3 A geometric variable

---

[3] Definition 1.1: $n \in \mathbb{N}$ is divisible by $a \in \mathbb{N}$, iff there is a number $q \in \mathbb{N}$ with $n = a \cdot q$.

Fig. 4 A proof using geometric variables and figurate numbers

	arithmetic/ algebra	**figurate numbers**
generic proofs	*generic proof with natural numbers*	*generic proof with figurate numbers*
proofs with variables	*formal proof*	*proof with geometric variables*

Fig. 5 Four different kinds of proofs used in the course "Introduction into the culture of mathematics"

to explain in detail (i.e., to use more words and to stress the generality using the word 'always'). When using algebraic variables, the formulation of the claim becomes easier. The algebraic variables can serve as a good starting point for the 'formal' proof later on.

Task 1: Consider the following equations:

$$1^2 + 1 + 2 = 2^2 \qquad 2^2 + 2 + 3 = 3^2 \qquad 3^2 + 3 + 4 = 4^2$$

Name and generalize the principle that is shown in the examples.

(a) Formulate the general principle with the use of word-variables. (*When one squares a natural number and adds the number itself and its successor, the sum will always equal the square of the successor.*)
(b) Formulate a general principle by using algebraic variables. (*For all $n \in \mathbb{N} : n^2 + n + (n + 1) = (n + 1)^2$.*)

Task 2: Consider at the following equations:

$$3^2 - 1 = 8 = 8 \cdot 1 \qquad 5^2 - 1 = 24 = 8 \cdot 3 \qquad 7^2 - 1 = 48 = 8 \cdot 6$$

Name and generalize the principle that is shown in the examples.

(a) Formulate the general principle by using word-variables. (*The square of an odd number minus 1 always equals a multiple of eight.*)
(b) Formulate the overall principle with the use of algebraic variables. (*For all $a \in \mathbb{N}$ there is a $b \in \mathbb{N}$ satisfying. $(2a + 1)^2 - 1 = 8 \cdot b$*).

Some may have discovered that the factors before 8 form the sequence of the triangular numbers. This argument could have led to a different generalization: some may have discovered that the factors before 8 form the sequence of the

triangular numbers.[4] This argument could have led to a different generalization: $(2n + 1)^2 - 1 = 8 \cdot D_n$, where D_n is the nth triangular number).

3.3 Multiple proof tasks

Tasks in which a claim has to be proven with several proofs are called 'multiple proof tasks' (Leikin 2009, p. 31). We adapted this idea and asked for generic proofs and formal proofs in the context of figurate numbers, too. When looking for different kinds of proofs for one claim, some exploration has to be done to find adequate starting points (ibid. p. 179). Moreover, students are to work in different 'notational systems' (algebra, numbers, and figurate numbers), so that they can experience the (dis-)advantages of each system. While the use of concrete numbers seems to be an easy and intuitive approach, the question about the generality of the argument has to be tackled. While the use of figurate numbers offers some graphical or demonstrative answers to a claim, one always needs to have an 'idea' how to manipulate them, how to arrange and group them. In the case of the symbolic language of mathematics the following advantages can be experienced, broadening the advantages mentioned in Sect. 2.3:

- This notational system has a universal character, i.e., it can be applied to every mathematical problem.
- Often, one does not need to have a particular 'idea' to construct an argument, as is the case in generic proofs. After formulating a claim using algebraic variables, simple (also experimental) computations may lead to the answer of a problem.

---

[4] D_n is the nth triangular number: $1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$.

- The power of algebra becomes evident in the calculation following valid rules.
- The generality follows immediately from the use of (algebraic) variables and the rules for calculations.
- It is an 'easy' and short way to communicate arguments, as the reader usually understands the language.

We give the following task as an example of a multiple proof task:

Task 3: Consider the following claim:

The sum of six consecutive natural numbers is always odd.

Prove the claim with a…

(a) generic proof with numbers.
(b) generic proof with figurate numbers.
(c) proof with geometric variables.
(d) formal proof.

The reader can easily imagine how these proofs can be performed.

4 Research questions

In the context of our study, we are interested in how far students argue to verify a statement when beginning their university studies. Here also the question arises of whether the students make use of generic examples (research question 1). In comparison to their performance at the beginning of the course, we also want to investigate students' performance at the end of the course. The comparison of these two data sets may lead to further insights concerning the impact of our course (research question 2). Finally, we want to investigate how students cope with the four kinds of proofs used in the course (research question 3).

(1) How do students argue when being asked to verify a theorem of elementary number theory at the beginning of the course?

    (a) Are there common mistakes concerning the use of algebraic variables?
    (b) Do students make use of generic examples?

(2) How do students argue when being asked to verify a theorem at the end of the course?
    (a) What differences compared to the former results can be observed?
(3) To what extent do students succeed in constructing the four kinds of proofs at the end of the course?

5 Methodology

5.1 Participants

The participants in our study are pre-service teachers attending the course "Introduction into the culture of mathematics" at the University of Paderborn in the winter term 2014/15. In the first session of the course, the students were asked to take a pretest about their knowledge of proof and proving. For describing students' reasoning at the beginning of the course, we only refer to the participants in their first university semester (n = 71; male: 28, mean age = 19.68 years and female: 43, mean age = 20.24 years). These students have not been exposed to and therefore had not been influenced by other courses at the university yet. Concerning the second research question, we refer to the 51 first-year pre-service teachers (19 male and 32 female) who participated both in the pretest and in the final examination of the course. By using an anonymous and personalized code, it was possible to link students' data. When discussing students' performances concerning the four kinds of proof, we look at the results of all first-year students participating in the final examination of the course (n = 52; 19 male and 33 female).

5.2 Research instruments

The pretest was completed in the first session of the course, and the final examination of the course consisted (inter alia) of a task in which students were asked to verify the statement about the sum of two odd numbers (see below). By using the same task, it becomes possible to investigate and compare students' performances in more detail. Finally, the students were asked to prove one statement in the final examination with all four different kinds of proof. In the following, we describe these research instruments in detail.

(1) The task to investigate students' reasoning.

We used the following task to investigate students' reasoning:

The sum $11 + 17$ is an even number.
Is this true for every sum of any two odd numbers?
Argue convincingly!

This task seems appropriate for investigating first-year students' reasoning because it is easy to understand and can be answered by utilizing basic knowledge of arithmetic and algebra (see Brunner 2013, p. 193). Moreover, different ways for solving the task are possible. The example at the beginning of the task illustrates the claim and may

open the way for further checks of examples and explorations. The formulation "Is this true" explicates the inherent generality of the claim. Finally, we deliberately asked to "give convincing reasons" to avoid any connotations with the word proving, in case associations with this word might affect students' work in some way. On the contrary, we intended to provide a wider context and to open the task for any argument, not necessarily making use of algebraic variables. The term "convincing" is used to make students explicate all their arguments.

(2) The task to investigate students' proof constructions.

In the final examination, students were asked to work on the multiple proof task about the sum of six consecutive numbers (task 3, see above). We chose this claim because it is accessible to all four kinds of proof.

5.2.1 The set of categories to investigate students' reasoning and students' proof productions and its refinement in a pilot study

As mentioned in Sect. 2.2, we needed a set of categories to categorize proof constructions concerning all four kinds of proof used in the course. We started the development of this research instrument by using deductive-inductive coding (see Kuckartz 2012, p. 69). We started to work with the combination of the set of categories of Bell (1976) and Recio and Godino (2001) to categorize students' answers to the proving task about the sum of two odd numbers. We realized that we could hardly apply Bell's "dependence" category (Table 3) and therefore omitted it. Furthermore, we found it necessary to subclassify the category "pseudo": Many students answered the task by just repeating the theorem (C3), by rephrasing the theorem (C4) or by mentioning non-relevant or wrong aspects (C5). We also wanted to make a note of the answers that could count as complete explanations but involved minor formal inaccuracies to distinguish these answers from really perfect ones. We finally came up with a set of ten categories (Table 1). For the present study, we felt the need to come up with a set of categories that could be applied to students' answers to the reasoning task and also to all students' proof constructions. Therefore, the following changes were made: The empirical categories "illustration" (C1) and "empirical verification" (C2) were combined, as it was hard to distinguish these two categories in the context of the four kinds of proof. The differentiation of the "pseudo" category was revoked because this distinction did not make sense when at the same time investigating proof constructions that made use of figurate numbers. What is more, the

categories "argumentation[5] with gap" (K4) and "complete explanation" (K5) were constructed by combining former categories, to have universal ones that can be applied to all four kinds of proof. Examples to illustrate each category are shown in Table 2.

When using the set of categories given in Table 2 in the pilot study in the winter term 2013/2014, all students' proof productions were coded by two raters. Also, students' answers to the proving task "sum of two odd numbers" in the winter term 2014/15 were coded twice. The correspondent inter-rater reliability concerning the set of categories is shown in Table 3.

5.3 Data collection

In the first session of the course, students were asked to work voluntarily on an anonymous pretest. We informed the students that their performance would not affect their grade in the course in any way. We used personalized and anonymous codes to link the students' data from the pretest with their performances in the final examination of the course. In this final examination, students had to work on the proving task about the sum of six consecutive natural numbers and again on the problem of the sum of any two odd numbers. Using the same task, it became possible to compare students' reasoning before and after attending the course.

6 Results

In the following, we present the results concerning students' reasoning at the beginning of the course (Sect. 6.1) and students' reasoning in the final examination at the end of the course (Sect. 6.2). Finally, we elaborate on students' proof construction in the final examination of the course (Sect. 6.3).

6.1 Students' reasoning at the beginning of the course

In the questionnaire at the beginning of the course, students were asked to work on the problem about the sum of two odd numbers (see above). We investigated students' answers We were interested in data concerning (1) the quality of reasoning, (2) the way students argued, and (3) common mistakes when using algebraic variables.

(1) The quality of reasoning.

Surprisingly, only 10% of students' answers could be rated as 'complete explanation' (see Fig. 6). Also, the high percentages of pseudo answers (32%) and purely-empirical verifications (14%) are remarkable. Finally, only 19% of the answers of the first-year students contained valid arguments ["arg. with gap" + "complete expl."].

---

[5] In this paper, the word "argumentation" is meant to describe an approach, where a person applies some warrant on data to support a conclusion (in the sense of Toulmin 1958).

Table 1 The development of the set of categories to describe the quality of students' reasoning

Bell (1976, p. 28 f.)	Recio and Godino (2001, p. 86)	Kempen and Biehler (2014, p. 427 ff.)	The present study
Diverse	The answer is very deficient (confused, incoherent)	C0: no argument is given	K0: no argument is given
Empirical argumentation	The student checks the proposition with examples, without serious mistakes	C1: Illustration ("The truth of the general statement is illustrated by several examples.")	K1: empirical ("The student gives further examples for the sum of two odd numbers, but no (deductive) arguments are mentioned")
Extrapolation: Truth of general statement inferred from a subset of the relevant cases [...]. The basis of the inference is clearly empirical	The student checks the proposition with examples and asserts its general validity	C2: Empirical verification ("The truth of the general statement is inferred from a subset of examples.")	
Deductive reasoning		C3: pseudo-verification by just repeating the statement to be proved ("The answer is given by stating the statement that the sum of any two odd numbers is always even.")	K2: pseudo ("The arguments used are redundant, irrelevant or wrong.")
Dependence: Attempts to make a deductive link between data and conclusion, but fails to achieve any higher category		C4: pseudo-verification by pseudo argumentation ("The verification is done by an explanation that merely paraphrases the statement that the sum of two odd numbers is always even.")	
Relevant, general restatement: [...] represents the situation as a whole, in general terms, as if aware that a deductive connection exists but unable to expose it		C5: pseudo (3)—pseudo-argument, mathematically wrong ("The arguments given are either not relevant for the task or mathematically wrong.")	
Relevant, collateral details: [...] mentions relevant aspects which could form part of a proof [...] but fails to build them into a connected argument; is fragmentary		C6: fragmentary ("The answer contains relevant aspects that could form part of a proof, but the student fails to build up a coherent argument.")	K3: fragmentary ("The student mentions relevant aspects, but fails to build up a coherent argument.")

Table 1 (continued)

Bell (1976, p. 28 f.)	Recio and Godino (2001, p. 86)	Kempen and Biehler (2014, p. 427 ff.)	The present study
Connected, incomplete: Has a connected argument with explanatory quality, but is incomplete	The student justifies the validity of the proposition, by using other well-known theorems or propositions, using partially correct procedures	C7: connected arguments with an unrecoverable gap ("The student gives a connected argument with explanatory quality, but the argumentation includes an unrecoverable gap.") C8: connected argument, but incomplete ("The student gives a connected argument with explanatory quality, but the argumentation is incomplete. Here one could close the created gap by adding some sentences.")	K4: argumentation with a gap ("The argument given includes a gap so that the generality of the claim is not verified.")
Connected, S[a]: failing only because it appeals to facts or principles which are no more generally agreed than the proposition itself *Complete Explanation*: Derives the conclusion by a connected argument from the data and generally agreed on facts or principles	The student gives a substantially correct proof, which includes an appropriate symbolization	C9: complete explanation (a)—with minor (formal) inaccuracies ("The student derives the conclusion by a connected argument and from generally agreed facts or principles. Just because of minor (formal) inaccuracies the explanation is not a perfect validation.") C10: complete explanation	K5: complete explanation ("The truth of the claim is shown by making use of valid mathematical arguments.")

[a]Following Bell (1976, p 29), the name "Connected, S" is related to a connected deductive approach, where a "side-step" is made appealing to other facts or principles

Table 2 Set of categories for describing the quality of students' reasoning (categories, explanations and illustrations)

Name	Explanation	Illustration (taken from students' work)
K99: not processed	–	–
K0: no argument is given	The student gives a single vote about the truth of the statement	Yes
K1: empirical	The student gives further examples for the sum of two odd numbers, but no (deductive) arguments are mentioned	I tried several examples $3+3=6$ even $13+7=20$ even The sum of two odd numbers is always even
K2: pseudo	The arguments used are redundant, irrelevant or wrong	Yes. It is true for any two odd numbers because the result will always be divisible by two
K3: fragmentary	The student mentions relevant aspects but fails to build up a coherent argument	Any odd number can be represented by $2n+1$
K4: argumentation with a gap	The argument given includes a gap so that the generality of the claim is not verified	Yes, because an odd number can be represented by $2n+1$ and $(2n+1)+(2n+1)=2(2n+1)$ is even
K5: a complete explanation	The truth of the claim is shown by making use of valid mathematical arguments	Any odd number can be represented by $2n+1$, $n \in \mathbb{N}_0$. $(2n+1)+(2n+1)=2n+2m+2=2(n+m+1)$, $(n+m+1) \in \mathbb{N}$ So the result will always be an even number

Table 3 Inter-rater reliability (Cohen's kappa) concerning the set of categories used to categorize students' proof constructions for the four different kinds of proofs used in the course

	Cohen's kappa
"Task 3" about the sum of six consecutive numbers	
Generic proof with numbers	0.804
Formal proof	0.823
Generic proof with figurate numbers	0.783
Proof with geometric variables	0.756
Task "sum of two odd number"	0.818
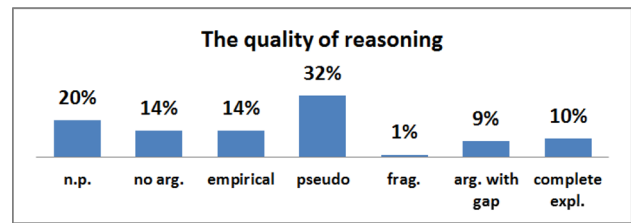
Fig. 6 The quality of first-year pre-service teachers' reasoning in the pretest (n = 71)

(2) The way students argue.

After having categorized students' answers, we were able to distinguish eight different types of (correct and wrong) arguments (see Table 4). The corresponding results are shown in Fig. 7. Here, it also becomes clear how these different ways of arguing correspond to the categories of the quality of students' reasoning.

Only 6.8% of the students used formalization with algebraic variables. Surprisingly, 19.7% answered the question about the sum of two odd numbers by stating or rephrasing the theorem that the sum of two odd numbers is always even.

Only two students made use of a generic example when trying to answer the given task: one student in the context of the argument "using the digits" and one concerning "the overlaps cancel each other".

(iii) Common mistakes in the use of algebraic variables in the pretest.

Both of the students making use of the formalization "$2n+1$" only used one variable to represent the sum of any two odd numbers [e.g. $(2n+1)+(2n+1)$]. Out of the seven students using "$n+1$" to work on the task, four used only one variable to represent the sum. So, we can assume that students are not familiar with the use of variables to solve proving tasks from elementary arithmetic.

Table 4 Different ways students are arguing in the task "Sum of two odd numbers"

Abbr.	Name of category	Explanation
n. p	Not processed	The task is not processed
A0	No argument	There is no argument given
A1	"Inductive"	The truth of the claim is inferred from a subset of examples without further (deductive) argument
Pseudo-arguments		
A2	"Stating the theorem"	The theorem is stated that the sum of any two odd numbers is always even
A3	"Redundant, irrelevant, wrong"	The arguments given are redundant, irrelevant, or even wrong
Arguments without formalization		
A4	"The overlaps cancel each other"	Within the argumentation the following argument is used: Every odd number has an 'overlap' compared to the former even number. If you sum up two odd numbers, one adds these 'overlaps' too. Since these two overlaps cancel each other, the result equals an even number
A5	"Using the digits"	Within the argumentation the following argument is used: When adding two odd numbers, it suffices to have a look at the digits at the right. Here one adds two odd numbers between 1 and 9. The result will always be an even number
A6	"Even and odd numbers alternate"	Within the argumentation, the following argument is used: Considering the natural numbers, even and odd numbers alternate. One may interpret the sum of two odd numbers as walking along the number line, starting at an odd number and walking forward an odd number. So I will always stop on an even number
Arguments with formalization		
A7	Formalization using "2n + 1"	The odd numbers are represented in the form "$2n+1$"
A8	Formalization using "n + 1"	The odd numbers are represented in the form "$n+1$", where n represents an even number

Fig. 7 Results concerning students' way of reasoning in the pretest
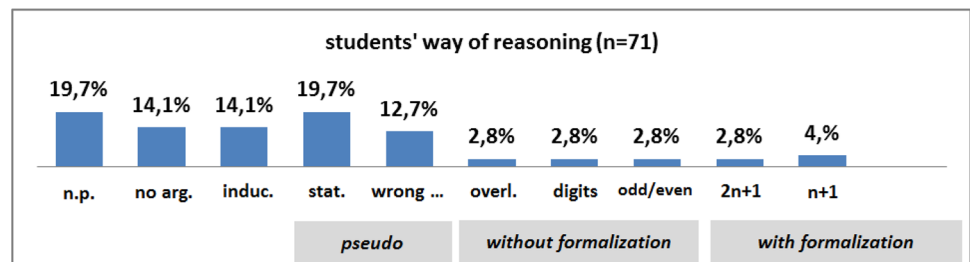
Table 5 Results concerning the "quality of reasoning" in the pretest and the final examination [%] (first-year pre-service teachers that could be tracked from the pretest to the final exam, n=51)

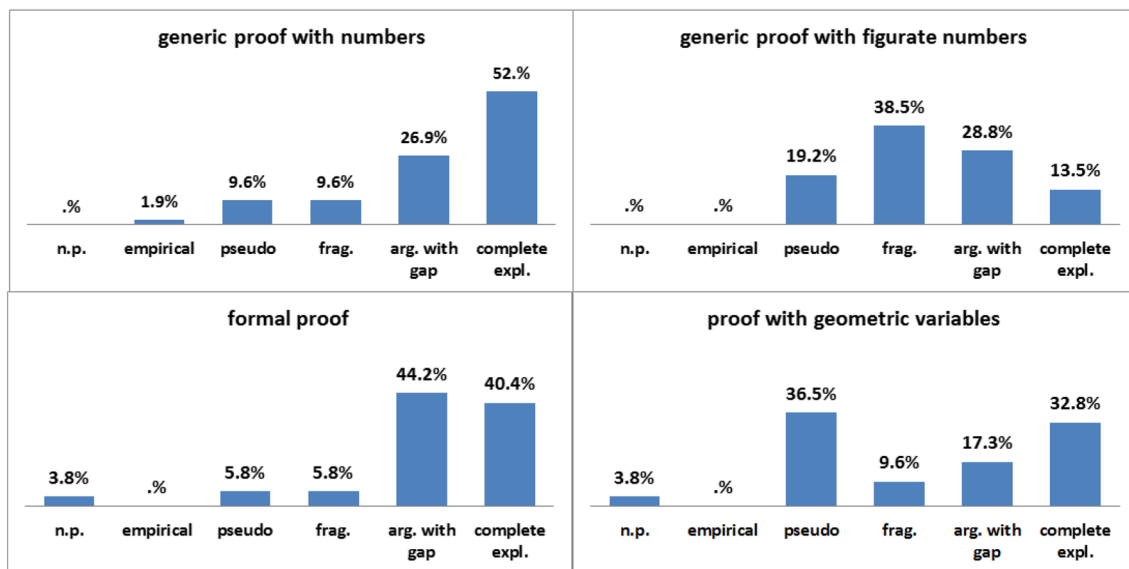|	Pretest	Final exam
K99: not processed	17.2	0
K0: no argument is given	21.2	0
K1: empirical	5.8	0
K2: pseudo	32.7	7.8
K3: fragmentary	1.9	2.0
K4: argumentation with a gap	15.4	52.9
K5: a complete explanation	5.8	37.3
Sum	100	100

6.2 Students' reasoning at the end of the course

In the final examination of the course, students were asked to work again on the same problem as in the pretest about the sum of any two odd numbers. For these data, we used the same sets of categories for investigating (1) the quality of reasoning, (2) the way students argue, and (3) common mistakes when using algebraic variables. The data set consisted of the 51 first-year students whose data could be linked from the pretest to the final examination.

(1) The quality of reasoning.

Table 6 Results concerning students' way of reasoning in the final exam [%] (first-year pre-service teachers)

		First-year students (n=51) [%]
n. p	Not processed	–
A0	No argument	–
A1	Inductive	–
Pseudo		
A2	Stating the theorem	5.9
A3	Wrong, irrelevant	2
Arguments without formalization		
A4	Overlaps	5.9
A5	Digits	–
A6	Odd and even numbers	–
Arguments with formalization		
A7	"2n+1"	84.3
A8	"n+1"	2
A9	"n+n=2n"	–
Sum		100
Sum "with formalization" [A7+A8+A9]		86.3

Fig. 8 Results concerning proof construction from the final examination of the course (n=52)

While only 21.2% of the students' answers in the pretest could be rated as meaningful argumentations (K4 + K5), this percentage is about 90.2% in the final test of the course (see Table 5). However, only 37.3% of all argumentations achieved a "complete explanation" because of (minor) inaccuracies or mistakes. In the final test of the course, there were no more answers without any arguments or purely empirical verifications. While about a third of the students' answers were pseudo answers in the pretest,

we had a much lower rate in the final examination. From a normative point of view, the 'low' rate of complete explanations (37.3%) has to be seen critically. Investigating the way students were arguing in the final examination will partly explain this phenomenon.

(2) The way students argue.

In the final examination of the course, the vast majority of the first-year students used some form of formalization, most students in the shape "2n + 1" (84.3%) (see Table 6).

All other ways of arguing nearly disappeared. The argumentation about the overlaps of odd numbers was as frequently used as the answer "stating the theorem" (5.9%). All three students giving the argument "overlaps" made use of at least one generic example.

(3) Common mistakes in the use of algebraic variables in the final exam.

Out of the 40 first-year students making use of the formalization "$2n + 1$", 19 students (47.5%) used only one variable to represent any two odd numbers.

6.3 Students' proof productions in the final examination of the course

In the final examination of the course, the students had to prove the following statement with the four kinds of proof used in the course (see Sect. 3.2): "The sum of six consecutive natural numbers is always odd". We used the same set of categories as in the previous analysis to investigate students' proof productions. The results are shown in Fig. 8.

About half of the first-year students (52%) succeeded in constructing a complete generic proof with numbers. In sum, meaningful arguments are given by 78.9% of the students ["arg. with gap" + "complete explan."]. Concerning formal proof, meaningful arguments were given by 84.4% of the students, but only 40.5% achieved a "complete explanation". This result is due to our concept of formal proof (see Sect. 3.1).

To achieve a complete formal proof, we asked for a reference to a definition or a theorem at the end of the proof for confirmation that the result really is an odd number. In the case of the proofs with figurate numbers, students seemed to struggle with this kind of notational system. Here, the percentages of pseudo and fragmentary answers increased enormously.

One caveat has to be mentioned concerning students' motivation in working on these tests. Of course, their motivation for constructing the best argumentation possible was probably quite a lot higher in the final examination of the course. However, in the pretest, students did try to give meaningful argumentations, as most answers were also quite detailed.

7 Discussion

In this section, we answer our research questions [RQ] (Sect. 7.1) and discuss the results in a broader context (Sect. 7.2).

7.1 Summary of findings

RQ1: How do students argue when being asked to verify a theorem of elementary number theory at the beginning of the course? (a) Are there common mistakes concerning the use of algebraic variables? (b) Do students make use of generic examples?

In the pretest at the beginning of the course, only 10% of the first-year students ($n = 71$) were able to give a justification for the claim about the sum of any two odd numbers that we could rate as "complete explanation". While 9% of the answers to the task could be accepted as "argumentation with gap", 32% were "pseudo" answers (stating the theorem about the sum of any two odd numbers or giving wrong or irrelevant facts). 14% of the students gave purely empirical arguments. 20% of the first-year students did not attempt to answer the task. Considering the way students argued, only two students (6.8%) used algebraic variables to answer the given task. Both students made the mistake of using only one variable to represent any two odd numbers. The use of examples in a somehow generic way could be detected in only two argumentations.

The high percentage of answers without any correct argument may give a hint of the fact that students in Germany are not familiar with these kinds of proof tasks when entering university. Since only 6.8% of the students make use of the symbolic language of mathematics, it seems as if the students are not capable of using an algebraic variable as a heuristic to perform the kind of reasoning required here. Besides, the use of generic examples does not seem to be a heuristic that first-year students use intuitively.

RQ2: How do students argue when being asked to verify a theorem at the end of the course? (a) What differences compared to the former results can be observed?

In the final test of the course, nearly all students used formalization in the form "$2n + 1$" to verify the statement about the sum of any two odd numbers (84.3%). However, due to a deficient usage of algebraic variables, only 37.3% achieved a "complete explanation". (Out of the 40 first-year students making use of the formalization "$2n + 1$", 19 students (47.5%) used only one variable to represent any two odd numbers.[6]) Only a few answers still consisted of the pseudo-answer "stating the theorem" (5.9%). Three students (5.9%) made use of a generic proof using the "overlaps between the even and odd number".

Compared to the results of the pretest, the answers not containing any argument or using empirical evidence

---

[6] An explanation for this result may be that we did not use many tasks where students had to make use of two variables to prove a given claim. This problem should be considered in the next implementation of the course.

disappeared. The percentage of "pseudo" answers (stating the theorem or naming wrong or irrelevant facts) decreased enormously from about a third to 7.8%. Taken together, the answers containing meaningful argumentations ["arg. with gap" + "complete explanation"] rose from 21.2 to 90.2%.

To sum up, students' performance in verifying a theorem of elementary arithmetic increased from the pretest to the final examination of the course. The high percentage of answers making use of algebraic variables (mostly "2n + 1") illustrates students' preference for using algebraic variables at the end of the course. However, even after attending the course, first-year students struggled with the correct use of variables to fulfil a complete verification. Compared to the pretest, the number of answers containing generic examples did not increase significantly.

RQ3: To what extent do students succeed in constructing the four kinds of proofs at the end of the course?

At the end of the course, 78.9% of the students gave valid arguments when constructing the generic proof with numbers, and 52% succeeded in achieving a "complete explanation". Only one student gave a purely empirical-verification. It seems as if the majority of students grasped the idea of a generic proof and were able to work with this concept.

Concerning formal proof, valid arguments were given by 84.6% of the students; a "complete explanation" was achieved by 40.4%. This quite low percentage seems to be astonishing because the given task was quite easy. However, due to our concept of formal proof used and taught in the course, we asked for a reference to a definition or a theorem for the confirmation that the result obtained (e.g., $\dots = 6n + 15 = 2 \cdot (3n + 7) + 1$) is an odd number[7]. Not meeting this norm led to a gradation to the category "argumentation with gap". However, overall, the students seem to be able to work with algebraic variables to prove a claim of elementary arithmetic at the end of the course.

Only 13.5% of the students succeeded in constructing a generic proof with figurate numbers. 38.5% of the proof attempts were rated as "fragmentary", 19.2% as "pseudo" answers. In the case of the proof with geometric variables, "pseudo" answers were given by 36.5% of the students, and a complete explanation was achieved by 32.8%. We would like to recall that a "pseudo" answer in these cases refers to a proof attempt where no meaningful structure or use of figurate numbers can be detected. The category "fragmentary" is used when the compilation of figurate numbers cannot be followed by the given answer, but the configuration obtained could be used for further arguments. Surprisingly, students still struggled with the use of figurate numbers even

at the end of the course. It would be interesting to investigate whether problems with the use of figurate numbers and with the concepts of the different kinds of proof used affect each other.

7.2 The results viewed from a broader perspective

We presented a learning sequence for a special kind of transition-to-proof course for pre-service teachers. Four different kinds of proofs were used to foster first-year pre-service teachers' proof competencies, to ease their transition to university and to equip them with kinds of proof they can also use in class later on. In this sense, we presented a conception for a university course specially designed for pre-service secondary (middle school) teachers. We intend to highlight this effort as an example of innovation of the content of existing university curricula, from the viewpoint of competence-orientation and target audiences. The course is conceptualized explicitly to continue students' prior experiences with mathematical proof and to foster their proof competencies when entering university (see research question 1). In this sense, this kind of didactical consideration for higher-education may serve as an example for the international community.

We showed how students' proof attempts changed from the pretest to the final examination of the course (research question 2). While almost no student made use of algebraic variables in the pretest, 86.3% of the students utilized them in the final examination. Accordingly, we argue that students learned about the use of algebraic variables as a significant tool for justifying claims of elementary arithmetic. Only very few students made use of generic examples to answer the task "sum of two odd numbers". In the case of the pretest, this might be due to the fact that the students did not know this concept. However, the concept of generic proofs does not seem to be an intuitive heuristic for our first-year students.

Students' proof performances at the end of the course (research question 3) gave more insights into the benefits of the concepts used in the course and their problems. While students more or less succeeded in constructing the generic proof and in performing reasoning with the use of algebraic variables, they still seemed to struggle with the use of figurate numbers, especially with the use of geometric variables. More research is needed to investigate the benefits and limitations of generic proofs (at university and school) and the use of figurate numbers in the context of mathematical proofs for learners. This paper offers research instruments for similar studies and results that could serve as a starting point for further investigations.

---

[7] A complete solution would have been for example $\dots = 6n + 15 = 2 \cdot (3n + 7) + 1$. Since $(3n + 7) \in \mathbb{N}$, the result is an odd number (Theorem 2.1).

References

Balacheff, N. (1988). Aspects of proof in pupils' practice of school mathematics. In D. Pimm (Ed.), *Mathematics, teachers and children* (pp 216–235). London: Hodder and Stoughton.

Bell, A. W. (1976). A study of pupils' proof-explanations in mathematical situations. *Educational Studies in Mathematics, 7*, 23–40.

Biehler, R., & Kempen, L. (2013). Students' use of variables and examples in their transition from generic proof to formal proof. In B. Ubuz, C. Haser & M. A. Mariotti (Eds.), *Proceedings of the Eighth Congress of the European Society for Research in Mathematics Education* (pp. 86–95). Ankara: Middle East Technical University. Retreived from http://cerme8.metu.edu.tr/wgpapers/WG1/WG1_Kempen.pdf. Accessed 12 Sep 2018.

Biehler, R., & Kempen, L. (2016). Didaktisch orientierte Beweiskonzepte—Eine Analyse zur mathematikdidaktischen Ideenentwicklung. *Journal für Mathematik-Didaktik, 37*(1), 141–179.

Blum, W. (1998). On the role of "Grundvorstellungen" for reality-related proofs: Examples and reflections. *Proceedings of CERME 3*. Retrieved from http://www.erme.tu-dortmund.de/%7Eerme/CERME3/Groups/TG4/TG4_Blum_cerme3.pdf. Accessed 12 Sep 2018.

Blum, W., & Kirsch, A. (1991). Preformal proving: Examples and reflections. *Educational Studies in Mathematics, 22*, 183–203.

Boero, P. (1999). Argumentation and mathematical proof: A complex, productive, unavoidable relationship in mathematics and mathematics education. *International Newsletter on the Teaching and Learning of Mathematical Proof*, July/August.

Brunner, E. (2013). *Innermathematisches Beweisen und Argumentieren in der Sekundarstufe 1. Mögliche Erklärungen für systematische Bearbeitungsunterschiede und leistungsförderliche Aspekte*. Münster: Waxmann.

Dreyfus, T., Nardi, E., & Leikin, R. (2012). Forms of proof and proving in the classroom. In G. Hanna & M. de Villiers (Eds.), *Proof and proving in mathematics education: The 19th ICMI Study* (pp. 191–214). Heidelberg: Springer.

Hemmi, K. (2006). *Approaching proof in a community of mathematical practice*. Stockholm University: Stockholm. Retrieved from http://www.diva-portal.org/smash/get/diva2:189608/FULLTEXT01.pdf. Accessed 18 Sep 2018.

Karunakaran, S., Freeburn, B., Konuk, N., & Arbaugh, F. (2014). Improving preservice secondary mathematics teachers' capability with generic example proofs. *Mathematics Teacher Educator, 2*(2), 158–170.

Kempen, L. (2019). *Begründen und Beweisen im Übergang von der Schule zur Hochschule*. Wiesbaden: Springer Spektrum.

Kempen, L., & Biehler, R. (2014). The quality of argumentations of first-year pre-service teachers. In S. Oesterle, P. Liljedahl, C. Nicol, & D. Allen (Eds.), *Proceedings of the 38th Conference of the International Group for the Psychology of Mathematics Education and the 36th Conference of the North American Chapter of the Psychology of Mathematics Education* (Vol. 3, pp. 425–432). Vancouver, Canada: PME.

Kuckartz, U. (2012). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. Weinheim/Basel: Beltz Juventa.

Leikin, R. (2009). Multiple proof tasks: Teacher practice and teacher education. In F. L. Lin, F. J. Hsieh, G. Hanna & M. De Villiers (Eds.), *ICMI Study 19: Proof and proving in mathematics education (Vol* 2, pp. 31–36). Taipei: The Department of Mathematics, National Taiwan Normal University.

Leron, U., & Zaslavski, O. (2013). Generic proving: Reflections on scope and method. *For the Learning of Mathematics, 33*(3), 24–30.

Malle, G. (1993). *Didaktische Probleme der elementaren Algebra*. Braunschweig/Wiesbaden: Vieweg & Sohn.

Mason, J., Graham, A., & Johnston-Wilder, S. (2005). *Developing thinking in Algebra*. London: The Open University.

Mason, J., & Pimm, D. (1984). Generic examples: Seeing the general in the particular. *Educational Studies in Mathematics, 15*, 277–289.

Recio, A. M., & Godino, J. D. (2001). Institutional and personal meanings of mathematical proof. *Educational Studies in Mathematics, 48*, 83–99.

Reid, D., & Vallejo Vargas, E. (2018). When is a generic argument a proof? In A. J. Stylianides & G. Harel (Eds.), *Advances in mathematics education research on proof and proving: An international perspective* (pp. 239–251). Cham: Springer International Publishing.

Selden, A. (2012). Transitions and proof and proving at tertiary level. In G. Hanna & M. de Villiers (Eds.), *Proof and proving in mathematics education: The 19th ICMI Study* (pp. 391–422). Heidelberg: Springer.

Selden, A., & Selden, J. (2007). Overcoming students' difficulties in learning to understand and construct proofs. *Making the connection: Research and practice in undergraduate mathematics*. MAA Notes.

Tall, D. (1979). Cognitive aspects of proof, with special reference to the irrationality of $\sqrt{2}$. In D. Tall (Ed.), *Proceedings of the 3rd International Conference on the Psychology of Mathematics Education* (pp. 203–205). Warwick: The Mathematics Education Research Center.

Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.

Wittmann, E. C. (2009). Operative proof in elementary mathematics. In F.-L. Lin, F.-J. Hsieh, G. Hanna & M. de Villiers (Eds.), *Proceedings of the ICMI Study 19 conference: Proof and Proving in Mathematics Education* (Vol. 2, pp. 251–256). Taipei: The Department of Mathematics, National Taiwan Normal University.