CrossMark

Pre-Service Teachers' Benefits from an Inquiry-Based Transition-to-Proof Course with a Focus on Generic Proofs

Leander Kempen [1] · Rolf Biehler [1]

Abstract

At the University of Paderborn, the course "Introduction into the culture of mathematics" is required for all first-year students who enter the study program for future mathematics teachers at lower secondary level (grade 5–10). In this inquiry-based transition-to-proof course, we use four different kinds of proofs (the generic proof with numbers, the generic proof with figurate numbers, the proof with figurate numbers using geometric variables, and the so-called "formal proof") to engage students in exploration, reasoning, and proving. In this paper, we report findings from an empirical study in winter term 2014/15 (pre- and posttest) concerning proof validation and acceptance. We used different kinds of 'reasoning' taken from Healy and Hoyles (2000) to assess students' proof validation. At the beginning of the course, about a third of the students judged the purely empirical verifications and wrong algebraic operations as correct proof. These forms of reasoning being judged as correct proofs decreased greatly in the posttest. To investigate proof acceptance, the students had to rate different aspects – such as "conviction", "explanatory power", or "validity" - of the four kinds of proofs. "Proof acceptance scales" with very high reliabilities (Cronbach's $\alpha > .864$) were constructed using factor analysis. While in the pretest most of the students did not accept the generic proofs and the proof with geometric variables as general valid verifications, their acceptance increased during the course. However, in the posttest, the ratings of the different aspects vary greatly concerning the four kinds of proofs.

Keywords Transition to university · Generic proof · Proof acceptance · Proof validation

✉ Leander Kempen
  kempen@khdm.de

  Rolf Biehler
  biehler@khdm.de

[1] University of Paderborn, Warburger Straße 100, 33098 Paderborn, Germany

Introduction

Many universities offer transition-to-proof courses in order to address students' diffi-
culties with mathematical proof. Inquiry-based transition-to-proof courses seem to be
especially promising, discussing proving-techniques while emphasizing the process
aspect of mathematics (Selden et al. 2015; Smith 2006). It is the emphasis on the
process aspect of mathematics (exploration, identification of assumptions, proving, and
considerations of different kinds of explanations) that characterizes an inquiry-based
approach to transition-to-proof (see National Research Council 1996). We developed
an inquiry-based transition-to-proof course for first-year student teachers using four
different kinds of proofs to foster students' proof competencies: the generic proof with
numbers, the generic proof with figurate numbers, the proof with figurate numbers
using geometric variables (see Kempen and Biehler 2016), and the so-called "formal
proof". In this course, we pursued three objectives: (1) To enhance students' transition
to the mathematical (so-called) formal proof, (2) to promote the mathematical symbolic
language in a meaningful way and (3) to equip students with intellectually-honest[1]
ways of proving that can also be used in their future teaching at the school level.

The ability to understand and to construct mathematical proof is a key competency
in university mathematics. The understanding of the meaning, significance and function
of the mathematical symbolic language is an undeniable prerequisite for dealing with
higher mathematics. The teaching of appropriate forms of mathematical proof for
school mathematics provides the possibility to include valid forms of reasoning into
the classroom, i.e. to establish a culture of reasoning and proving in an intellectually-
honest and propaedeutic way. The course is evaluated and refined in a design-based
research scenario. Plomp (2010, p. 9) describes this methodology as follows: Design-
based research is "the systematic study of designing, developing and evaluating
educational interventions (such as programs, teaching-learning strategies and materials,
products and systems) as solutions for complex problems in educational practice, which
also aims at advancing our knowledge about the characteristics of these interventions
and the processes of designing and developing them". In this paper, we focus on the
fourth implementation of the course in 2014/15. As in other implementations of the
course, we investigated students' benefits concerning proof competencies, beliefs and
acceptance in detail. In this paper, we report on our findings concerning proof valida-
tion and proof acceptance, a concept on which we will also elaborate.

In the teaching of mathematical proof, the study of Healy and Hoyles (2000) gave
some important insights about 14- and 15-year old students' conceptions of different
kinds of reasoning. Here it became clear that learners' mathematical socialization and their
acceptance of forms of reasoning have to be considered in the teaching of proof. While
generic proofs are said to be a pedagogical tool to engage students in reasoning and
proving and to foster proof competencies (see Dreyfus et al. 2012; Leron and Zaslavski
2013; Rowland 1998; compare also the concept of "transparent pseudo-proofs" in Malek
and Movshovitz-Hadar 2011) it seems surprising that students' acceptance of these kinds
of reasoning has not been investigated in detail yet. In this contribution, we want to add

---

[1] The intellectual-honesty principle states that the notion of proof in school mathematics should be concep-
tualized so that it is honest to mathematics as a discipline and honoring students as mathematical learners (see
Kirsch 2000; Stylianides 2007).

new and current results to the ongoing discussion about pre-service teachers' conceptions of proof and their acceptance of different (pedagogically oriented) kinds of proofs.

Theoretical Background

Students' Proof Validation

Following Selden and Selden (2017, p. 340), proof validation involves the reading of and the reflection on proof or proof attempts to determine their correctness. In this sense, proof validation is related to the correctness of a proof, i.e. the arguments used and its inherent logic. Proof validation has been investigated concerning undergraduate students (e.g. Selden and Selden 2003; Sommerhoff et al. 2016), mathematicians (e.g. Weber 2008), and pupils (e.g., Healy and Hoyles 1998; Healy and Hoyles 2000). Since we are interested in students' proof validation and proof conceptions when entering university, the results of the Healy and Hoyles study appear especially important for us. In their study of approximately 2500 high-attaining 14- and 15-year old students, Healy and Hoyles (1998, 2000) found some remarkable results concerning proof conceptions. In this study, students' were asked to judge four different kinds of reasoning (a narrative proof,[2] an empirical verification, a wrong algebraic proof and a correct argumentation with variables) concerning the aspects validity (e.g. "shows that the statement is always true") and explanation (e.g. "shows you why the statement is true") (Healy and Hoyles 2000, p. 403). Concerning the two given empirical verifications, more than half of the students gave correct evaluations (54% and 60%), stating that these arguments only rely on the subset of cases. Less than one-third of the students expressed that these empirical arguments had no explanatory power at all. While the correct algebraic proof was judged correctly concerning its validity by 40% of students, only 11% attested the highest rate of explanation. In the case of the wrong algebraic proof, only 12% gave correct judgements concerning its validity and only 3% attested the highest degree of explanatory power. The highest validity rating was achieved in the case of the narrative argument, where 68% of the students gave correct evaluations. The highest percentage of full explanatory power (42%) was also achieved in the case of the narrative argument.

Another important result of the study touches upon students' proof conceptions. These students held two different conceptions of proof: the arguments they considered would achieve the best mark from their teachers differed from the arguments they would adopt for their own approach. While the students chose algebraic arguments for achieving the best mark, the empirical arguments were preferred for their own approach. These empirical arguments were also found to be more convincing and explanatory.

Reiss et al. (2000) adapted a part of Healy and Hoyles' study. They chose four different arguments about a proof problem as to whether or not a given triangle is isosceles, and gave it to 81 secondary school students (German Gymnasium, the most academic type of secondary school in Germany). These students were to assess the explanatory power, correctness and generality of four different arguments: a correct

---

[2] Following Healy and Hoyles (2000, p. 400 f.), a narrative proof is considered as "an argument that suggest[ed] underlying reasons and explanations written in a narrative everyday style".

formal proof, a correct narrative proof, an empirical argument and a formal circular solution. Reiss et al. (2000, p. 116) summarized their findings:

> We interpret the findings, particularly those on assumed teacher preference, as an indication that the majority of students consider a correct (noncircular), formally presented proof to be the mathematically accepted norm. However, they have not entirely adopted this norm in their own attempts at proof or their understanding of convincing mathematical arguments. The majority feel that the correct narrative proof is the best way of explaining geometrical content to their classmates, while one third of respondents selected the purely empirical argument as the best explanation.

Students' Proof Acceptance

Since proof validation can be considered to be an objective question about the correctness of a proof (see section "Students' proof validation"), it is also necessary to discuss a subjective perception when talking about the acceptance of a proof. This idea of a subjective category is linked to the concept of relative conviction of Weber and Mejia-Ramos (2015). As Weber and Mejia-Ramos (2015, p. 16) point out, an argument can improve the relative conviction[3] about the truth of a claim. In this sense, a person's proof acceptance is not only affected by the (perceived) correctness of proof, but also by its representation, the extent of validity, conviction and verification, as we will point out below. Accordingly, we will elaborate on a concept of "proof acceptance" that is based on different views articulated in several studies. Finally, this elaboration will lead to a definition of proof acceptance that will also serve as an operationalization here.

The term "proof acceptance" is often used in the literature when students judge given arguments as "correct proofs". First, it appears obvious that students ratings of concrete argumentations as "correct proofs" depend on their individual conception of mathematical proof. Following this idea, Reid and Knipping (2010, p. 66) summarize some findings from the literature concerning factors that may influence persons' acceptance of arguments as proofs. Their enumeration comprises the form of the argument and the familiarity with the conjecture, the familiarity of the methods used and the use of diagrams and concrete examples. Thus, in the so-called acceptance of different kinds of proofs, the form (i.e. representation) of the argument and the use of concrete examples or diagrams appear to play a crucial role. Dreyfus (2000) found that most secondary school teachers in his study ($n = 44$) easily validated a formal symbolic mode of representation as proof, but they rarely appreciated other arguments presented in verbal or visual mode or done with generic examples. These teachers even tended to perceive narrative proofs as deficient because of the lack of mathematical symbolic language. Tabach et al. (2010, 2011) investigated the mathematical knowledge of high school teachers and concluded that some teachers over-value the generality of formal presented arguments and under-value the generality of verbal ones.

---

[3] "*Relative Conviction*: An individual has relative conviction in a claim if the subjective level of probability that one attributes to that claim being true exceeds a certain threshold to provide a warrant for some future actions." (Weber and Mejia-Ramos 2015, p. 16; emphasis in original)

When a person is rating a mathematical proof, several aspects have to be considered, which is reflected in different studies. First, a proof must fulfill *verification*, i.e. the proof shows that the statement is always true (compare Healy and Hoyles 2000; Weber 2014). Another important aspect when reading a proof is the extent of *explanation*, i.e. the proof shows you why the statement is true (compare Healy and Hoyles 2000; Hanna 2018). Then the subjective category of *(relative) conviction* must be considered (see above): is someone completely convinced by this argument? (see Weber 2010; Weber and Mejia-Ramos 2015). Finally, in the study of Martin and Harel (1989) students were asked to rate whether each verification of a statement was a valid mathematical proof. In this sense, proof acceptance is about the label of 'proof'; i.e., if someone would call a verification proof.

It is appropriate for the term "proof acceptance" to combine the different aspects that the reader is to perceive when reading and understanding a proof. These aspects are related to the various functions of proofs (e.g., de Villiers 1990) as verification, explanatory power, conviction, and also the perceptions or interpretations of the argument by the reader as a purely empirical check of examples, the elimination of any doubt about the validity of the statement being proved and the judgement as "proof". In this study 'proof acceptance' is conceptualized as the extent to which an individual perceives verification, conviction and explanation when reading a mathematical proof combined with the extent, the reader does consider the reasoning to be a "correct mathematical proof". For this investigation, new instruments are needed.

In their study, Healy and Hoyles (see above) use different items to assess the validity and the explanatory power of different kinds of reasoning. But in having a closer look at the items, the questions seem to bear more than 'just' verification and explanatory power. We interpret the items of Healy and Hoyles (2000, p. 403) to touch upon the following aspects of mathematical proof: (a) *correctness* ("has a mistake in it"), (b) *verification* ("shows that the statement is always true"), (c) *misunderstanding as a single check of examples* ("only shows that the statement is true for some even numbers"), and (d) *explanatory power* ("shows you why the statement is true"). We followed this idea of asking for different aspects of mathematical proof to meet with our concept of proof acceptance. We combined the dimension mentioned above with characteristics of the four types of proofs herein (see below). Finally, we developed the following dimensions to assess students' proof acceptance: "verification" (the objective confirmation about the truth of a statement), "interpretation as purely empirical verification", "possible existence of counterexamples", "(relative) conviction", "importance of variables", "explanation", "interpretation as testing of concrete cases" and the "correctness" of a proof.[4]

The Course "Introduction into the culture of mathematics"

The course "Introduction into the culture of mathematics" was developed by the second author and was held for the first time in 2011/2012 as a requirement

---

[4] The concrete items to evaluate these dimensions are given in section "Research Instruments".

for first-year secondary pre-service teachers. We used this first implementation of the course and the following three to refine and to evaluate the course in a design-based research scenario. Accordingly, the empirical study presented in this paper is a part of a larger empirical study that forms the core of the Ph.D. project of the first author.

About 160 pre-service teachers attend this course every year, which comprises (per week): 2 h of lecture, 2 h of tutorial in groups (about 30 students each) and 2 h of "plenary tutorial", where didactically- enhanced and reflected solutions of the weekly homework are presented. During the course, the students explore mathematical issues (e.g., figurate numbers) and learn to construct generic proofs and formal proofs. An integral part of the concept of the course is the usage of four different kinds of proofs (see Kempen and Biehler 2016). In the following section, we will describe these proof types and the related aims in detail.

Four Different Kinds of Proofs

In the teaching of mathematics, different kinds of proofs have been suggested and discussed by mathematics educators and mathematicians (Dreyfus et al. 2012). The concept of the generic proof has become a prominent didactical tool both for the secondary and the tertiary level (e.g., Rowland 2002; Stylianides 2010). Dreyfus et al. (2012, p. 204) describe this kind of proof as follows:

> 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.

A crucial point in the concept of the generic proof is the identification of the generic argument and the understanding (or acceptance) of the general character of the argumentation. Biehler and Kempen (2013) discuss establishing norms for the construction of a generic proof at university, to ensure students' understanding and acceptance of the concept. Following this pedagogical view, a generic proof consists of generic examples followed by valid narrative reasoning. Therefore, this concept of a generic proof is due to a pedagogical context (see also Reid and Vallejo Vargas 2018). There are valid operations performed on concrete examples that illustrate why the statement is true in these examples. Afterward, it must be explained why this argument also fits all possible cases and therefore is a general verification. We communicated this norm to our students in the course "Introduction into the culture of mathematics". Our students were asked to present concrete (generic) examples and to explain their inherent argument with the generic character verbally when constructing a generic proof. We consider this norm as a matter of sociomathematical norm in the sense of Yackel and Cobb (1996).
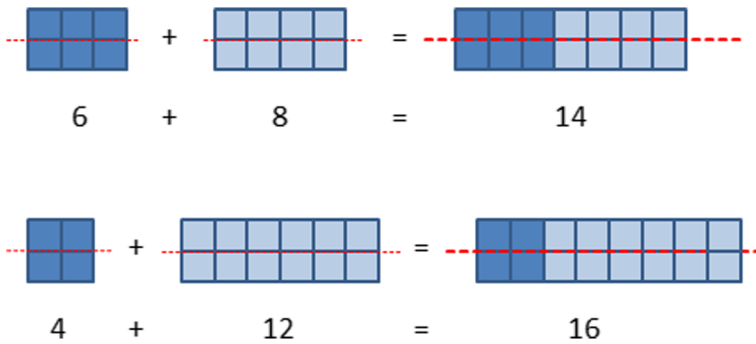
Fig. 1 The sum of two even numbers represented by figurate numbers

We will now give examples of the four proof types we distinguish in our course. We start with generic proofs:

Claim: The sum of two even numbers is always even

The generic proof with numbers:

$$4 + 6 = 2 \cdot 2 + 2 \cdot 3 = 2 \cdot (2 + 3)$$
$$8 + 12 = 2 \cdot 4 + 2 \cdot 6 = 2 \cdot (4 + 6)$$
$$2 + 14 = 2 \cdot 1 + 2 \cdot 7 = 2 \cdot (1 + 7)$$

Every even number can be written as two times a natural number. By using the distributive law, the sum of two even numbers equals two times the sum of two natural numbers. Since two times any natural number is even, the result will always be an even number.

While in this generic proof the concrete examples involve (natural) numbers, it is also possible to work with figurate numbers. As Mason and Pimm (1984) argue, diagrams can be helpful to foster the perception of the immanent generality. In addition to this view, the use of diagrams or geometric representations are said to be useful to fulfill the transition to algebra, i.e. to convey a meaningful concept of algebraic variables (e.g., Flores 2002). Following these considerations, we use also the notational system of figurate numbers to construct generic proofs. An example for the generic proof with figurate numbers is shown below.

The generic proof with figurate numbers[5]:

By using figurate numbers, every even number can be represented by two equally long rows of dots. By adding two equal numbers, one will always obtain two resulting equally long rows of dots. This means that the result will always be an even number (Fig. 1).

---

[5] (1) In the figure corresponding to the generic proof with figurate numbers, we are using little squares instead of 'dots'. In this case, the use of squares makes it easier to grasp the geometric shape mentioned in the text. (2) In this proof, we explain the fact that every even number can be represented by two equally long rows of dots. But we do not explicitly mention that every (natural) number represented by two equally long rows of dots is even. We considered this fact to be clear for the reader of the proof.

In generic proofs, the generality must be conveyed in concrete examples and be expressed in a narrative reasoning, but the use of variables implies the generality in so-called "formal proofs".

The formal proof:

Let $a, b \in \mathbb{N}$, a and b and even. Then a and b can be written as: $a = 2n$ and $b = 2m$ with $n, m \in \mathbb{N}$. We have: $a + b = 2n + 2m = 2 \cdot (n + m)$. Since $(n + m) \in \mathbb{N}$, the sum is an even number. Q.e.d.

In the literature, some proofs with figurate numbers use little dots to represent an arbitrary number. This is a way of expressing a variable more explicitly in the notational system. This notational variation may help to identify the generality of the argument better than with using "simple figurate numbers" and therefore an additional narrative explanation is not necessary, similar to a formal proof. We learned from earlier implementations of the course that an explicit distinction of this fourth type of proof is helpful for the students. We illustrate such a proof below with the same proposition (Fig. 2).

A proof with geometric variables:

In proofs with algebraic or geometrical variables, we do not ask for additional narrative reasoning, because the variables and their use are meant to express the generality. This is why the students do not have to write down their argument when constructing a proof with geometric variables if they find the diagram sufficient. But when constructing a generic proof with numbers or figurate numbers, the students are to verbally explain the argument and the immanent generality. In this way, we want to highlight the use of variables to express generality in the sense of Mason et al. (2005) in contrast to the use of single (generic) examples, where one would explain the generality. A single example is a matter of a concrete context, where one might identify a generic argument. But here, we ask the students to explicate this broader argument explicitly. In mathematics, variables are meant and used to represent a broader context, i.e., to express generality (see above). The proofs using figurate numbers and geometric variables can be considered as to what Nelsen (1993 and 2000) calls a 'proof without words': "Generally, proofs without words (PWWs) are pictures or diagrams that help the reader see *why* a particular mathematical statement may be true, and also to see *how* one might begin to go about proving it true" (Nelsen 2000, p. ix; author's emphasis).

The Learning Sequence in the Course: Brief Overview

The content of the course is divided into six chapters: (1) Discovery and proof in arithmetic, (2) figurate numbers, (3) sequences and mathematical induction, (4) propositional logic and proof types, (5) equations, and (6) functions.
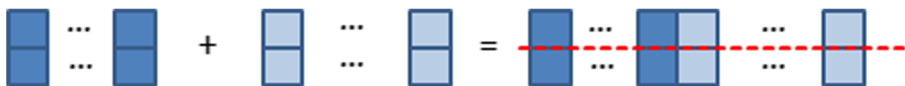

Fig. 2 A proof with "geometric variables" and figurate numbers

The first chapter broaches the issue of exploration, discovery and proving in particular. The exploration begins with the question: "Someone claims: The sum of three consecutive natural numbers is always divisible by three. Is this correct?". Discussing the value of testing and investigating concrete examples, the concept of the generic proof is presented by the lecturer and discussed by the whole group. By formalizing the generic argument, one comes closer to the idea of a formal proof and the algebraic variables are introduced in a meaningful way to express generality (see Mason et al. 2005). In comparison of the generic proof and the formal proof, it becomes possible to discuss the meaning of generality when proving a universal statement. During the course, the question of generality in the case of the generic proof is an important matter and must be discussed in detail (see Kulpa 2009; Mason and Pimm 1984). We further try to generalize the initial statement and investigate the following:

(C2) The sum of 2 consecutive natural numbers is always divisible by 2.
(C4) The sum of 4 consecutive natural numbers is always divisible by 4.
(C5) The sum of 5 consecutive natural numbers is always divisible by 5.
(C6) The sum of 6 consecutive natural numbers is always divisible by 6.
…
(Ck) The sum of $k \in \mathbb{N}$ consecutive natural numbers is always divisible by k.

The refutation of (C2) or (C4) makes it possible to discuss the value of a counterexample in the case of a universal statement. Here, it is not only possible to refute the statement, but also to prove a general one: "The sum of 2 consecutive natural numbers is never divisible by 2". After comparing all the results concerning the statements, the final conjecture is formulated and proved: The sum of $k \in \mathbb{N}$ consecutive natural numbers is divisible by k if and only if k is an odd number.

In the second chapter, we investigate different figurate numbers (e.g., triangular numbers, square numbers, and pentagonal numbers). The students are to find structures and relationships between the different figurate numbers and to prove these with different kinds of proofs. While in the first chapter, the arithmetic was the place for conjecturing and proving and figurate numbers were used as an alternative notational system to prove the statements, in the second chapter it is vice versa. The figurate numbers are the place for conjecturing and proving and the students may use concrete examples with numbers or the algebraic symbolic language to prove the different findings. In this sense, it is also intended that students may experience the power of symbolic mathematical language and the use of algebraic variables. In the remaining chapters, the mathematical proof always plays a central role. We use the four different kinds of proofs throughout the course. In the homework, we make use of so-called multiple proof tasks (e.g., Dreyfus et al. 2012, p. 198; Leikin 2009), where students have to prove a single statement with all four kinds of proofs.

We consider this course to be inquiry-based, because our students are to investigate mathematical phenomena (e.g. expressed in concrete examples) and to formulate conjectures. These conjectures have to be proven or refuted afterward. Accordingly, the students are to solve new and unfamiliar problems continuously (compare Rasmussen and Kwon 2007).

Research Questions

In the transition to higher mathematics, students face different challenges in their first semesters at university. Here, the new forms of reasoning and the operation on a higher formal level can be main hurdles in this transition (e.g., Selden 2012; Gueudet 2008). But in order to accomplish the transition to mathematical proof, it is necessary to investigate students' effective concepts of proof when entering university and their acceptance of different kinds of reasoning recommended in the literature. Also, since proof is a multidimensional construct, one has to take into account different aspects to investigate "proof acceptance". An important assessment was done by Healy and Hoyles (1998, 2000) when investigating middle school students' judgement on different kinds of reasoning (see section "Students' proof validation"). Before assessing students' proof acceptance, we wanted to apply the assessment done by Healy and Hoyles (students' judgement of four different kinds of reasoning, see above) to our students to provide a basis for a discussion of proof acceptance. For an evaluation of our course, the questions of the proof-validation test in the pretest (at the beginning of the course) were also asked in the posttest (at the end of the course). In this sense, we investigated a part of pre-service teachers' understanding of mathematical proof when entering university and the impact of our course (compare research question 1). Since the participants of our course comprise both first-year students and more advanced students, it appears valuable to examine these two subgroups separately. In this way, we might obtain a closer look at pre-service teachers' conceptions of proof when entering university and possible changes in their conceptions due to the courses at university they complete.

To investigate students' proof acceptance and their interpretation of the four different types of proofs used in the course, students were asked to judge one of each type of proof concerning the aspects "verification", "interpretation as purely empirical verification", "possible existence of counterexamples", "conviction", "importance of variables", "explanation", "interpretation as testing of concrete cases" and "correctness". These items were asked in the proof questionnaire used in the pre- and the posttest for evaluating the course's benefits. The corresponding research question 2 offers some considerable insights into pre-service teachers' understanding of the four different kinds of proofs. Finally, students' proof acceptance scores were calculated. Our investigation aims to describe students' proof acceptance when entering university (at the beginning of our course) and possible changes in students' proof acceptance during our course (see research question 3).

The research questions are:

(1)   How do pre-service teachers judge different kinds of reasoning (a narrative proof, an empirical verification, a wrong algebraic proof and a correct argumentation with variables) at the beginning and at the end of the course?

   a.   Are there meaningful differences concerning the students in their first semester and students in a more advanced semester?

(2)   How do pre-service teachers rate the different kinds of proofs (the generic proof with numbers, the generic proof with figurate numbers, the proof with figurate numbers using geometric variables and the so-called formal proof) concerning different aspects (verification, interpretation as purely empirical verification,

existence of counterexamples, conviction, importance of variables, explanation, interpretation as testing of concrete cases and correctness) at the beginning and at the end of the course?

(3) How does students' proof acceptance of the four kinds of proofs (resulting from their proof ratings) change during the course?

Methodology

Participants

The participants in our study are the pre-service teachers that attended the course "Introduction into the culture of mathematics" at the University of Paderborn in winter term 2014/15. The course was framed by a pre- and a posttest. By using personalized codes, students' performances and attitudes were tracked. The sample consisted of $N = 74$ pre-service teachers who participated in the pre- and the posttest. We focus our analysis on this subset of our students.

Research Instruments

The pre- and the posttest consisted (inter alia) of (1) a multiple choice proof-validation test adapted from Healy and Hoyles (2000) and (2) a proof acceptance survey, where the students assessed different kinds of proofs.

(1) **Proof-validation test**

In the multiple-choice test, the students were asked to rate four different kinds of reasoning taken from Healy and Hoyles (2000, p. 401). We translated the selected proofs and modified them slightly to emphasize the different immanent aspects (see Fig. 3). For each proof, students were asked if it is a "correct proof" or" no correct proof". We consider this investigation of students' proof validation as an important basis for our research concerning the broader concept of proof acceptance.

In the senior class, Katja, Leon, Maria und Nisha had to prove the following conjecture:

(2) **The proof acceptance questionnaire**

Our questionnaire included asking students to rate different aspects of concrete given proofs. These aspects included verification, interpretation as purely empirical verification, existence of counterexamples, conviction, explanation, testing of concrete cases and correctness (the concrete items are shown below). We adapted the idea of assessing different aspects (Healy and Hoyles 2000; Weber 2010) and formulated corresponding statements (see below). We selected one of each of the four kinds of proofs mentioned above to be rated by the students. (These concrete proof productions to be rated are shown below.) We made the decision to choose four proofs of different claims, because one type of proof should not influence the acceptance of the other ones of the same claim. Of course, a disadvantage is that the students also judge the concrete proof of that claim and not only the type of proof in general.

In the senior class, Katja, Leon, Maria und Nisha had to prove the following conjecture:

When you multiply any 3 consecutive numbers, your answer is always a multiple of 6.

	correct proof	no correct proof
Katja's answer A multiple of 6 must have factors of 3 and 2. If you have three consecutive numbers, one will be a multiple of 3 as every third number is in the three times table. Also, at least one number will be even and all even numbers are multiples of 2. If you multiply the three consecutive numbers together, the answer must have at least one factor of 3 and one factor of 2.	☐	☐
Leon's answer $1 \cdot 2 \cdot 3 = 6$ $2 \cdot 3 \cdot 4 = 24$ $4 \cdot 5 \cdot 6 = 120$ $6 \cdot 7 \cdot 8 = 336$	☐	☐
Maria's answer x is any natural number $$x \cdot (x + 1) \cdot (x + 2) = (x^2 + x) \cdot (x + 2)$$ $$= x^3 + x^2 + 2x^2 + 2x$$ Cancelling the x's gives $\quad 1 + 1 + 2 + 2 = 6$	☐	☐
Nisha's answer[6] The product of three consecutive numbers can be represented by: $n \cdot (n + 1) \cdot (n + 2)$, where n is any natural number. This product is divisible by 2, because: First case: n is an even number: n is divisible by two and so is the product. Second case: n is an odd number: $(n + 1)$ is an even number and divisible by two. So the product is divisible by two. The product is divisible by three, because if you have three consecutive natural numbers, exactly one will be a multiple of three. Since the product $n \cdot (n + 1) \cdot (n + 2)$ is divisible by two and 2 and 3, the product is a multiple of 6.	☐	☐

From the above answers, choose one that would be closest to what you would do if you were asked to answer this question.	
From the above answers, chose the one to which your teacher would give the best mark.	

Fig. 3 The multiple choice proof-validation test adapted from Healy and Hoyles (2000, p. 401)

The students were asked to rate the four concrete given proofs (see below) concerning the aspects named above on a six-level Likert scale ([1] "totally disagree" … [6] "totally agree"). The statements to be rated are:

The reasoning…

(i) shows that the statement is true in every possible case. ["true"; item concerning *verification*]

(ii)   convinces me that the statement holds in every case. ["conv."; item concerning *conviction*]
(iii)   shows that the statement is true for every time and 100%. ["100%"; item concerning *verification*]
(iv)   explains why the statement is true. ["explan."; item concerning *explanation*]
(v)   is a correct and valid proof. ["corr. Proof"; item concerning *correctness*]
(vi)   verifies the statement only for some concrete cases, but not in general. ["example"; item concerning *interpretation as a single check of examples*]
(vii)   is not universally valid, since a counterexample can still exist. ["counterex."; item concerning *existence of a counterexample*]
(viii) is just a test of single cases and not a general verification. ["cases"; item concerning *interpretation as a single check of examples*]
(ix)   is not a valid verification without the use of variables. ["variables"; item concerning *the importance of variables*]
(x)   has to be represented in a more formal way to totally convince me. ["more formal"; item concerning *the demand for a formal representation*]

The items (i), (iv), (vi) are adapted and slightly modified from Healy and Hoyles (2000, p. 403). The items (ii), (iii) and (v) were formulated to include several functions that proofs may fulfill and the items (vii) and (viii) to challenge misinterpretations. The items (ix) and (x) broach the issue of a formal representation that might be desired by a reader of a proof.
   The concrete proofs to be rated are the following:

   The generic proof with numbers ["genN"] - claim: The sum of an odd natural number and its double is always odd.

$$1 + 2 \times 1 = 3 \times 1 = 3, \;\; 5 + 2 \times 5 = 3 \times 5 = 15, \;\; 13 + 2 \times 13 = 3 \times 13 = 39$$

The sum of an odd natural number and its double equals three times the initial number. Since the initial number is an odd number, one obtains the product of
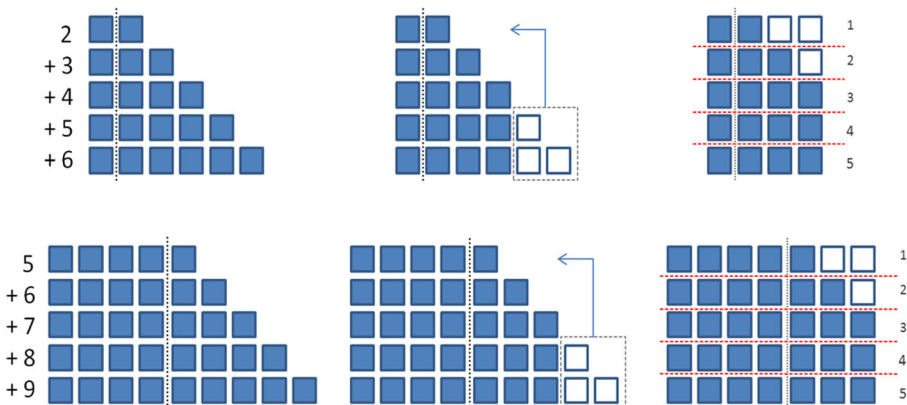

Fig. 4 The sum of five consecutive numbers represented by figurate numbers

two odd numbers. Since the product of any two odd numbers is always odd, the result will always be an odd number.

The generic proof with figurate numbers ["genFig"]- claim: The sum of five consecutive natural numbers is always divisible by five.

In the representation of the sum of five consecutive natural numbers by figurate numbers, one always obtains the same shape of stairs on the right side. By transforming these stairs (taking the edge at the bottom right and putting it above) one always obtains five equal rows. So the result will always be divisible by five (Fig. 4).

The formal proof ["FP"]- claim: For all natural numbers a, b, c: If b is a multiple of a and c is a multiple of a, then (b + c) is a multiple of a.

Let a, b, c be natural numbers. Since b is a multiple of a, there exists a natural number n with: $n\,a = b$. Since c is a multiple of a, there exists a natural number m with: $m\,a = c$. We have: $b + c = n\,a + m\,a = (n + m)$. Since $(n + m)$ is a natural number, $(b + c)$ is a multiple of a. Q.e.d.

The proof with geometric variables [GV] - claim: The square of an even natural number is always divisible by four (Fig. 5).

Pilot Testing and Refinement of the Research Instruments

The Proof-Validation Test The adaption of the proof item from Healy and Hoyles (2000) has been used in the context of the project "KLIMAGS" (e.g. Blum et al. 2015). We refined this instrument by changing the following aspects: We included the answer of Nisha, modified as mentioned above and took the Healy and Hoyles (2000) version of Leon's answer, which had been slightly changed in the "KLIMAGS" study. Finally, we changed the categories "proof/no proof" from the KLIMAGS study to
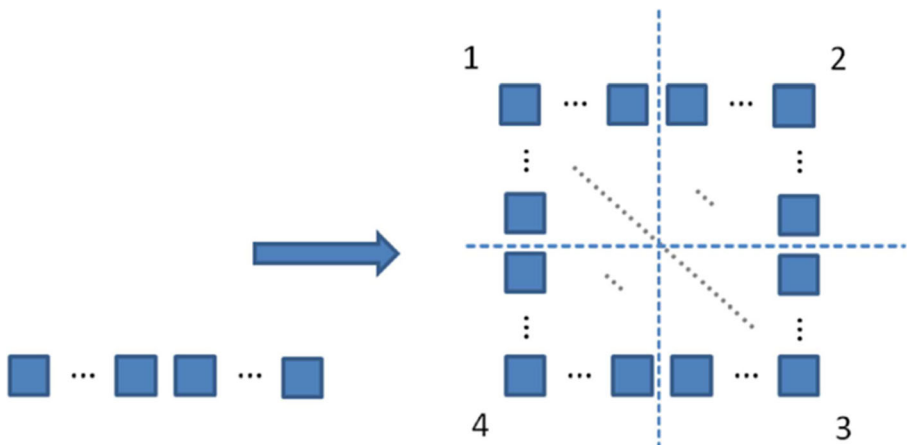

Fig. 5 A proof with "geometric variables" and figurate numbers

"correct proof/no correct proof" to avoid misunderstandings. This modified version in the proof-validation test was successfully piloted in winter term 2013/14.

The Proof Acceptance Survey We piloted the statements to be rated (items (i) – (x), see section "Research Instruments") on a six-level Likert scale with six different proofs in winter term 2013/14. Finally, we discussed the selection of the concrete proofs for the final questionnaire with several mathematics educators.

We investigated students' ratings concerning each aspect and used explorative factor analysis to evaluate our concept of proof acceptance. Finally, by using explorative factor analysis and considering the item discrimination, it became clear that the items concerning the different aspects of mathematical proof contribute to one factor, the construct we call "proof acceptance". Accordingly, the score of one proof acceptance scale is the mean of someone's ratings of the ten items used. A high scale value represents a high level of acceptance concerning a given 'proof'; a low value can be considered as its refusal. Such an acceptance scale was constructed for each proof being rated. The reliabilities of the constructed scales out of these eight items were very high (all Cronbach's alpha > .825).[6]

Data Collection

The pretest took place in the opening session of the course in winter term 2014/15. The posttest was conducted during the second to last session. By using personalized codes, 74 students were tracked from the pre- to the posttest. These 74 students constitute our dataset. When deepening our analysis, we will split this dataset to discuss the students in their first semester of university ($n = 37$) and the students in a more advanced semester[7] (n = 37) separately. The questionnaire included several questions concerning beliefs and competencies about argumentation and proof. Due to the length of this paper, we will only focus on findings about students' proof validation and proof acceptance.

Results

We will first report findings about students' proof validation. Then students' choices of the argument for achieving the best marks by the teachers and the choice for argument being similar to their own approach are discussed. Afterward, we will discuss students' proof acceptance. The results concerning the four different kinds of proofs are considered separately. In addition, we will compare the different kinds of proofs concerning the aspects "conviction" and "explanatory power", since these dimensions may be considered main aspects of mathematical proof (e.g., Hersh 1993). The results may open the discussion about the concept of proofs that explain with regard to the explanatory power of the mathematical symbolic language. Since we are also interested in possible existing differences between students in their first semester at university and

---

[6] The ratings to items (vii) – (x) were turned for data analysis (a rating as "1" became "6", a rating as "2" became "5", …). Accordingly a high value in the resulting scales points to a high level of proof acceptance.
[7] Even though the course is meant for teaching first-year students, some participants of the course are not in their first semester of university. Some of these changed their study program, some omitted the course in their first semester and others were repeating the course.

more advanced students, we will provide our results for all students who took both the pre- and the posttest separately for first year students ["first-year students"] and more advanced students ["more advanced"]. The more advanced students may have failed the first time or are new students in this study program but may have studied in another program (which may have included some math courses) before, or they have started this program in the summer term and attend our course in their second semester, although it is designed for first semester students.

Students' Proof Validation – Answer to the Research Question 1

The percentages of students validating the different arguments as correct proofs are shown in Table 1 for the pre- and the posttest (see also Fig. 6).

Of all students in the pretest, 74.3% judged the narrative argument ("Katja") as a correct proof. The purely empirical verification ("Leon") was considered a correct proof by 17.6% of students and the wrong argumentation ("Maria") by 27.0% of students. It is remarkable that about 17% percent of the students judged the checking of examples as correct proof and that about a third of them accepted the wrong formal reasoning in the case of Maria. The correct argumentation with variables was judged as correct proof by 89.2% of students and was the proof with the highest approval.

As shown in the posttest, the rating of the narrative argumentation as "correct proof" did not change. While the percentage decreased in the case of the purely empirical verification ($p = .012$; McNemar-Test) and the wrong argumentation, it increased in the case of the correct proof with variables.

In the subgroup of the first-year students, the narrative argumentation was rated nearly the same as the argumentation with variables. But as shown in the posttest, the percentage of the narrative argument decreased (pretest: 83.8%, posttest: 75.7%), while it increased in the other case. Both the wrong argument and the purely empirical examples were considered as proof by about a third of the students in the pretest, and about 10% still judged these arguments as correct proofs in the posttest. The decrease of first-year students accepting the empirical argument ("Leon") is significant ($p = .021$; McNemar-Test). Nearly all students in a more advanced semester judged the correct argumentation with variables as correct proof, but only 64.9% in the case of the narrative argument. Surprisingly, 27% of these students still accepted the wrong argument in the pretest, whereas the examples were not judged as correct proof (5.4%).

Table 1 Percentage of students validating the proof production as "correct proof" (pre- and posttest)

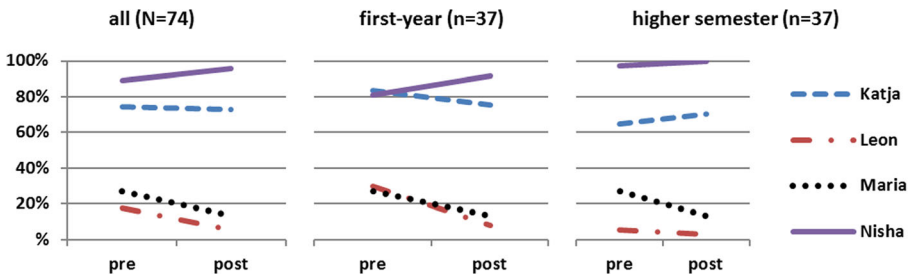|	all (N = 74)		first year students (n = 37)		more advanced (n = 37)	
	pre	post	pre	post	pre	post
Katja (narrative)	74.3	73.0	83.8	75.7	64.9	70.3
Leon (examples)	17.6*	5.4*	29.7*	8.1*	5.4	2.7
Maria (false)	27.0	13.5	27.0	13.5	27.0	13.5
Nisha (variables)	89.2	95.9	81.1	91.9	97.3	100

"*": $p < .05$ (McNemar-Test)

Fig. 6 Percentage of students validating the arguments as "correct proof" (pre- and posttest)

The results concerning students' choices for achieving the best mark and being the nearest to their own approach are shown in Fig. 7. Nearly all students chose the correct argument with algebraic variables ("Nisha": 89.9%) as the argument that would probably achieve the best mark by the teacher. But in the case for their own approach, students were almost as likely to choose the correct narrative proof ("Katja": 37.3%) as the correct proof with variables ("Nisha": 40.4%). In the subgroup of the first-year students, even more chose the narrative one ("Katja": 37.5%), but also 21.9% named the empirical argument ("Leon") as being the nearest to their own approach. The more advanced students preferred the argument with variables ("Nisha": 51.6%).

Students' Proof Acceptance

In this section, we will first discuss the results of the different items concerning each proof separately. Afterward, we will compare the four proofs among each other with regard to the aspects" conviction" and "explanatory power". The results concerning the constructed proof acceptance scales are presented subsequently.

Students' Ratings Concerning the Four Kinds of Proofs – Answer to Research Question 2

The results (medians) of the items concerning the four different kinds of proofs in the pre- and posttest are shown in Table 2 and Figs. 8 and 9.

Generic Proof with Numbers In the pretest, most of the students did not agree that this reasoning shows that the statement is true ("true": median of 3). On the contrary, the
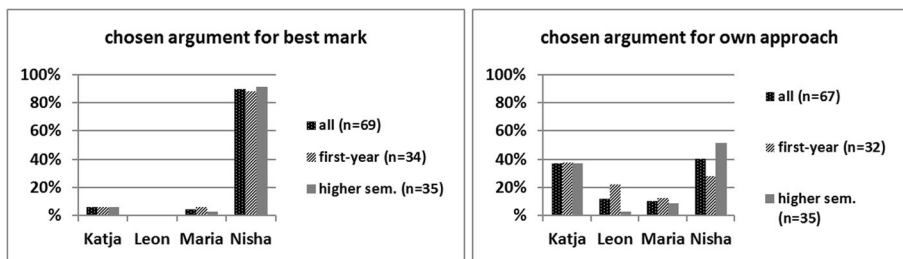

Fig. 7 Arguments chosen for achieving the best mark (left) and for students' own approach (right). Percentages of students' choices (whole group, first-year students and students in a more advanced semester)

Table 2 Medians of the proof acceptance items concerning the four different kinds of proofs in the pre- and posttest (ratings on a six-level Likert scale ([1] "totally disagree" … [6] "totally agree"); significance of the differences of the mean from the pre- to the posttest (Wilcoxon-Test): **: $p < .001$, *:$p < .01$, (*):$p < .05$

	generic proof with numbers		generic proof with figurate numbers		proof with geometric variables		formal proof[a]	
	pre	post	pre	post	pre	post	pre	post
true	3.0	5.0**	4.5	6.0**	3.0	5.0**	6.0	6.0
conviction	3.0	5.0**	5.0	6.0**	3.0	5.0**	6.0	6.0
100%	2.0	4.0**	2.5	5.0**	2.0	4.0**	5.0	6.0(*)
explanation	4.0	5.0**	5.0	6.0**	3.0	5.0**	6.0	6.0
correct proof	2.5	5.0**	3.0	6.0**	3.0	5.0**	6.0	6.0
example	5.0	2.0**	4.0	2.0**	4.0	2.0**	1.0	1.0
counterexample	5.0	2.0**	3.5	2.0**	4.0	2.0*	1.0	1.0
single cases	5.0	3.0**	4.0	2.0**	4.0	2.0**	1.0	1.0
variable	5.0	2.0**	3.5	2.0**	4.0	2.0**	–	–
more formal	5.0	4.0**	4.0	2.0**	5.0	3.0**	–	–

[a] In the case of the formal proof, the items concerning the missing use of variables ("variable") and the demand for a more formal representation ("more formal") were omitted, because these items seemed to be superfluous in this context

argumentation was considered as a check of several examples ("example": median of 5 and "cases": median of 5) and students agreed that there may still exist a counterexample to the statement ("counterexample": median of 5). Thus, this reasoning was not considered convincing ("conv.": median of 3) and only slightly explanatory ("explan.": median of 4). Accordingly, students did not agree that due to this reasoning the statement has to be 100% true ("100%": median of 2) and the reasoning was not judged as a "correct proof" (median: 2.5).

The results of the posttest show that students' perception of the generic proof with numbers had changed. Here, the generic proof with numbers was perceived as a
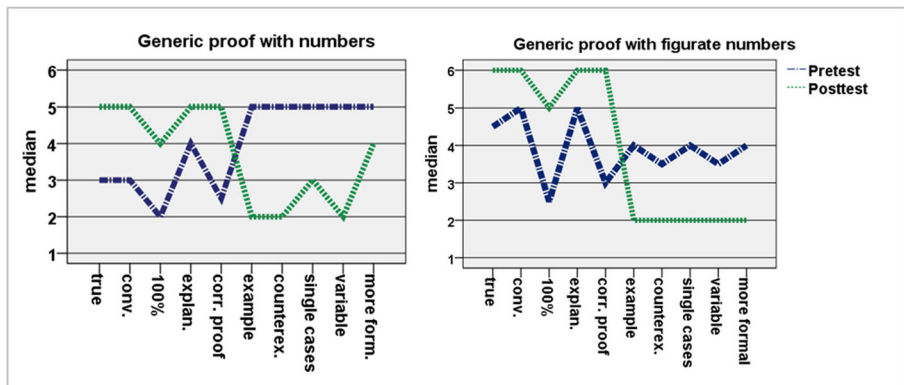

Fig. 8 Students' acceptance of the generic proof with numbers (left) and the generic proof with figurate numbers (right) - Results of the different items (medians) in the pre- and posttest ($n = 74$)
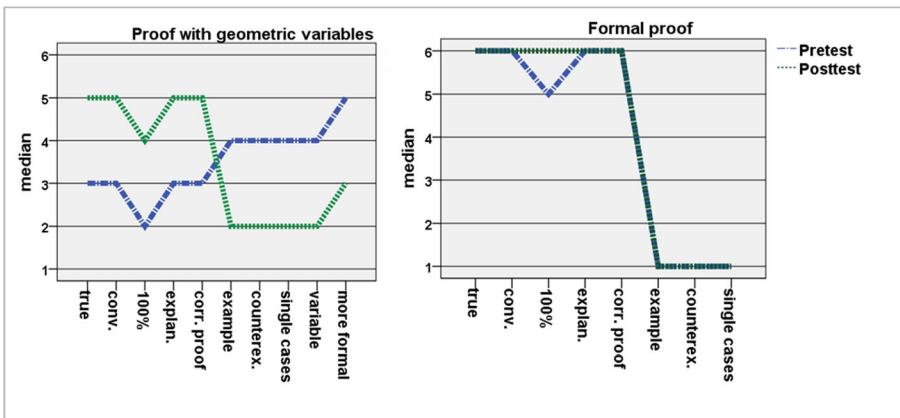
Fig. 9 Students' acceptance of the proof with geometric variables (left) and the formal proof (right) - Results of the different items (medians) in the pre- and posttest ($n = 74$)

convincing and explanatory argument (both median: 5) and even as a correct proof (median: 5) by most of the students. It was agreed that it shows that the statement is true (median: 5) but students still hesitated that the statement has to be 100% true (median: 4). Now, the generic proof was less considered as a purely empirical verification ("example"; median: 2 and "cases"; median: 3), but here the results show a high variation.

Generic Proof with Figurate Numbers In the pretest, students' ratings of the different items concerning the generic proof with figurate numbers showed a very high variation. While this reasoning was mostly considered convincing and explanatory (medians of 5), most of the students did not agree that the validity of the statement was proved 100% (median of 2.5). The medians of the items concerning the interpretation as a simple check of examples and the possibility of the existence of a counterexample are 4 and 3.5, so there was no clear statement expressed by the students. Finally, this reasoning was not considered as a correct proof (median of 3), but also in this case, the result shows a high variation.

The results of the posttest showed that students mostly agreed about the different aspects concerning this reasoning at the end of the course. The variation of the results decreased and the ratings displayed clear opinions. Now, the generic proof with figurate numbers was perceived as a general verification ("true": median of 6; "counterexample": median of 2 and "100%": median of 5) and the interpretation as a single check of examples was mainly rejected ("example" and "cases": median of 2). Finally, the reasoning was judged as a "correct proof "with a median of 6.

Proof with Geometric Variables The results concerning the proof with geometric variables are shown in Fig. 9 and Table 2. In the pretest, the high variations of the results illustrate students' different opinions about this argument. In summary, this proof was mainly not rated as a correct proof ("correct proof": median of 3) and the students did not agree that due to this reasoning the statement has to be true in general ("true": median of 3 and "100%": median of 2). Most of the students did not consider this argument to be convincing or explanatory (both with a median of 3). But in the posttest, the students state clear positions. In the posttest the students agreed that the

statement has to be true and that the argument is convincing (both median of 5). The interpretation as a single check of examples was rejected, and the proof was considered as explanatory and as a correct proof by the students (both with a median of 5).

Formal Proof In the pretest, the students stated clear positions in the case of the formal proof (see Fig. 9 and Table 2). The formal proof was perceived as a "correct proof" that shows that the statement is true. The interpretation as a purely empirical verification was rejected and this argument was both considered convincing and explanatory (medians of 6). In the posttest, all medians achieved the maximum (6) respectively the minimum (1) of the Likert scales. (As mentioned above, we did not use the items "variable" and "more formal" in the case of the formal proof, compare Fig. 9 right.)

Comparison of the Four Proofs: Students' Observed Conviction In this section, we refer to the results concerning "conviction" and "explanatory power" mentioned above to compare the four different kinds of proofs.

In the pretest, students rated the generic proof with numbers ["genN"] and the proof using geometric variables ["GV"] as the proofs that were the least convincing with a median of 3 (see Table 3 and Fig. 10). The generic proof with figurate numbers ["genFig"] had a median of 5. The results of these three proofs show a high variation. Here, the formal proof ["FP"] (median: 6) is rated the highest. Looking at the posttest, the high acceptance of the aspect "conviction" is remarkable. Both the generic proof with numbers and the proof using geometric variables have a median of 5 and the generic proof with figurate numbers and the formal proof have a median of 6. Overall, students' observed conviction remained quite high in the formal proof and greatly increased in the other proofs. At the end of the course, the generic proof with figurate numbers was rated as convincing as the formal proof.

Comparison of the Four Proofs: Students' Observed Explanatory Power With regard to "explanatory power", the generic proof with numbers and the proof with geometric variables achieved the lowest medians in the pretest (medians of 4 and 3). The generic proof with figurate numbers had a median of five. Concerning the explanatory power, the

Table 3 Statistical data concerning the item "conviction" ["genN": generic proof with numbers, "genFig": generic proof with figurate numbers, "GV": proof with geometric variables, "FP": formal proof]

Students' observed conviction in the pre- and posttest: Statistical Data

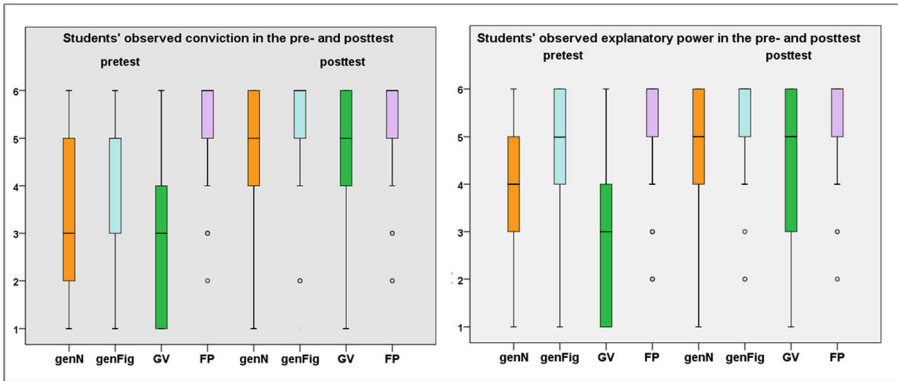|	pretest				posttest			
	genN	genFig	GV	FP	genN	genFig	GV	FP
n	74	74	68	72	74	74	68	72
mean	3.32	4.38	2.96	5.35	4.66	5.22	4.49	5.51
median	3.00	5.00	3.00	6.00	5.00	6.00	5.00	6.00
SD	1.664	1.411	1.688	1.050	1.581	1.162	1.568	.964
min	1	1	1	2	1	2	1	2
max	6	6	6	6	6	6	6	6

Fig. 10 Boxplots concerning the results of the items "conviction" (left) and "explanatory power" (right) for all four kinds of proofs in the pre- and the posttest ["genN": generic proof with numbers, "genFig": generic proof with figurate numbers, "GV": proof with geometric variables, "FP": formal proof]

formal proof was considered the best (see Table 4 and Fig. 10). In the posttest, the generic proof with numbers and the proof with geometric variables were rated higher than in the pretest, but the results showed a high variation. With regard to "explanatory power", the generic proof with figurate numbers and the formal proof were rated the highest.

The Proof Acceptance Scales – Answer to Research Question 2

By using exploratory factor analysis and considering the item discrimination, we constructed scales concerning "proof acceptance" (see section "Pilot testing and refinement of the research instruments"). The statistical data concerning the proof acceptance scales for the four kinds of proofs (in the pre- and posttest) are shown in Table 5.

Students' proof acceptance score concerning the generic proof with numbers (mean of 2.79) was quite low at the beginning of the course (see Fig. 11). The acceptance of the proof with geometric variables was nearly the same. The generic proof with figurate

Table 4 Statistical data concerning the item "explanatory power"

Students' observed explanatory power in the pre- and posttest: Statistical Data

	pretest				posttest			
	genN	genFig	GV	FP	genN	genFig	GV	FP
n	74	74	68	72	74	74	69	73
mean	3.82	4.50	2.85	5.15	4.85	5.41	4.19	5.58
median	4.00	5.00	3.00	6.00	5.00	6.00	5.00	6.00
SD	1.511	1.274	1.730	1.206	1.382	.905	1.743	.832
min	1	1	1	2	1	2	1	2
max	6	6	6	6	6	6	6	6

Table 5 Statistical data concerning proof acceptance scales

| | Proof-Acceptance-Scale | | | | | | | |
| | pretest | | | | Posttest | | | |
	acc_genN	acc_gen Fig	acc_GV	acc_FP	acc_genN	acc_gen Fig	acc_GV	acc_FP
n	74	74	67	72	74	74	67	72
mean	2.79	3.67	2.96	5.15	4.27	4.85	4.34	5.50
median	2.50	3.50	2.88	5.63	4.50	5.29	4.63	6.00
SD	1.18	1.27	1.27	1.02	1.45	1.27	1.36	.80
skew	.82	.28	.72	−1.05	−.51	−1.24	−.66	−1.78
kurtosis	.415	−.826	.336	−.192	−.890	.740	−.356	2.511
min	1.00	1.00	1.00	2.63	1.00	1.00	1.00	2.63
max	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00
Cronbach's alpha	.886	.912	.896	.939	.938	.928	.930	.951

numbers was more accepted (mean: 3.67) and the formal proof was the proof with the highest acceptance score (mean: 5.15). All differences between the means in the pretest are highly statistically significant ($p < .001$; t-test), except for the difference between the generic proof with numbers and the proof with geometric variables.

Compared to the pretest, all acceptance scores increased. The generic proof with numbers and the proof with geometric variables achieved the mean 4.27 and 4.34 respectively, which can be considered a positive acceptance. The generic proof with figurate numbers had an even higher score in the posttest (mean: 4.85) and the formal proof still had the highest mean (5.50). In the posttest, all differences between the means are highly statistically significant ($p < .001$; t-test), except for the difference between the generic proof with numbers and the proof with geometric variables. The
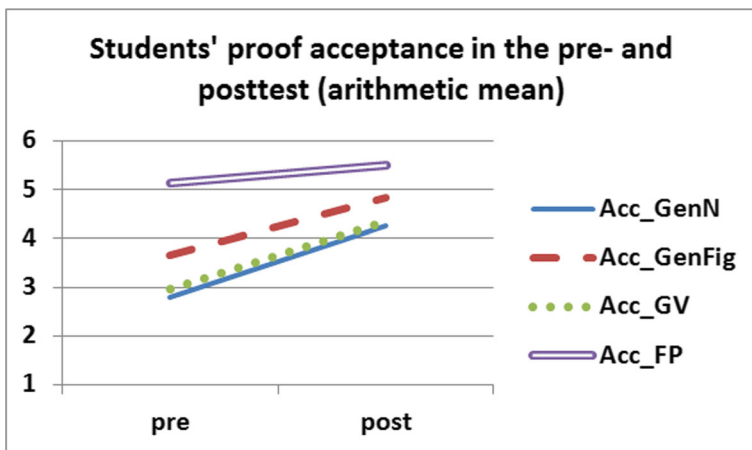

Fig. 11 Students' proof acceptance in the pre- and posttest (arithmetic mean) ["genN": generic proof with numbers, "genFig": generic proof with figurate numbers, "GV": proof with geometric variables, "FP": formal proof]

Table 6 Students' proof acceptance in the pre- and posttest (arithmetic mean); significance of the differences between the pre- and posttest (t-test) and effectsize (Cohen's d)

	n	pre	post	p value	Cohen's d
Acc_GenN	74	2.79	4.27	<.001	1.13
Acc_GenFig	74	3.67	4.85	<.001	.94
Acc_GV	67	2.96	4.34	<.001	1.06
Acc_FP	72	5.15	5.50	.003	.37

changes of the arithmetic means from the pre- to the posttest are highly statistically significant (p < .001; t-test) for all kinds of proofs, except for the increase of the formal proof (pre-Test: 5.16; post-test: 5.50) with $p = .003$ (see Table 6).

Since the analysis of the arithmetic means showed only a general increase in proof acceptance, it appeared necessary to also examine the individual development of the scores. By subtracting the acceptance score of the pre- and the posttest for every student, an individual change score was calculated ("change_acc_*prooftype*"). The resulting scales run (theoretically) from −6 to +6, where a positive score implies an increase concerning proof acceptance from the pre- to the posttest. These individual change scores are shown in Figs. 12 and 13.

With regard to the generic proof with numbers, 71.2% of the individual acceptance scores increased. The mean of 1.47 and the maximum of 5 also illustrate students' increased acceptance. In the case of the generic proof with figurate numbers, 72.6% of individual scores increased. The mean of the scale is 1.19 and its maximum of 4.38 illustrates the change in students' proof acceptance. Regarding the proof with geometric variables, 72.1% of the acceptance scores increased (mean: 1.37 and maximum: 4.88). The results concerning the formal proof are also remarkable: The score increased for 49.3% of the students and stayed constant for 34.2% of them. This result is due to students' maximum acceptance of the formal proof (see above). The high variation in students' change score shows the heterogeneity of their individual benefits from the course. While some students' proof acceptance (nearly) increased by five points, the proof acceptance of others decreased slightly.
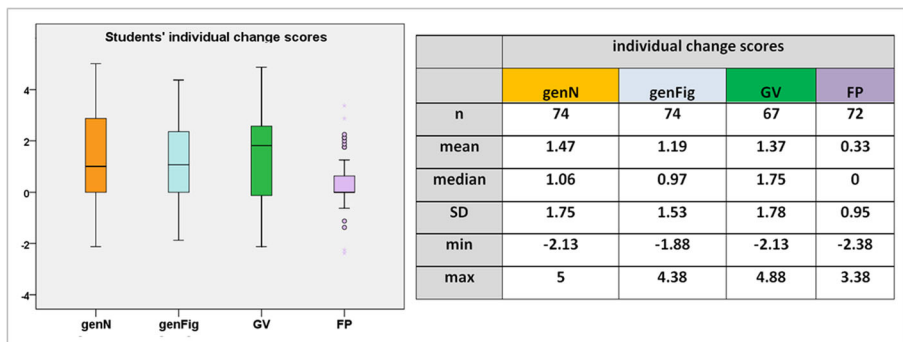

Students' individual change scores		individual change scores			
		genN	genFig	GV	FP
	n	74	74	67	72
	mean	1.47	1.19	1.37	0.33
	median	1.06	0.97	1.75	0
	SD	1.75	1.53	1.78	0.95
	min	-2.13	-1.88	-2.13	-2.38
	max	5	4.38	4.88	3.38

Fig. 12 Boxplots and statistical data of the individual change in proof acceptance
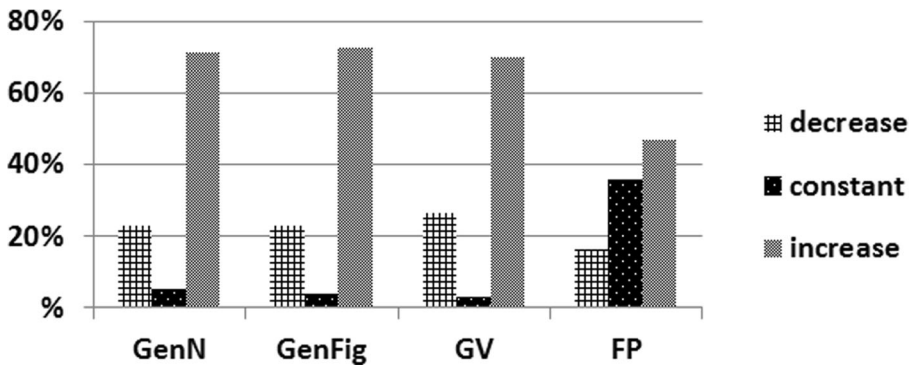
Fig. 13 Students' individual change acceptance score, summarized in the categories "decrease", "constant" and "increase"

Summary and Discussion

In this study, we investigated pre-service teachers' proof validation and proof acceptance and how they benefited from attending an inquiry-based transition to proof course.

In the beginning of the course, 17.6% of the students rated the testing of several examples as "correct proof". But after attending the course, this percentage decreased significantly to 5.4%. For the subset of the students in their first semester, the percentage is even more remarkable, decreasing from 29.7% in the pretest to 8.1% in the posttest. The presented argument containing an algebraic error was rated by 27% of the students as correct proof in the beginning of the course. Here it is not possible to explain their choice, as it can be due to several factors (e.g., they did not notice the error, they did not understand the algebraic formulation at all or they may have overestimated the value of the algebraic symbols). But at the end of course, the percentage decreased to 13.5%. One has to note that this rate still appears quite high for university students. Comparing the validations of the two correct reasonings, the one with the use of algebraic variables achieved more approval as correct proof, both in the pre- and the posttest. These results give insights into students' understanding of the concept of proof when entering university. About a third of the students starting their university studies hold inadequate conceptions of mathematical proof; these students consider a single check of several examples to be a valid proof. Also, the wrong algebraic proof was judged as correct by about a third of the students. It appears as if their mathematics courses prior to university did not provide the students with sufficient concept knowledge about mathematical proof. The question arises, what features of our course may have had the desirable effects? First, we want to emphasize our use of examples. The students of the course were to explore several claims and we wanted them to test the claims on concrete examples first. Thus, the investigation of concrete examples became part of the proving process (in the sense of Boero 1999). On the basis of these investigations, a generic proof might have been constructed. Accordingly, we used the concept of generic proofs to emphasize continuously the difference between a check of examples and a valid general proof. Throughout the course, the students were to construct generic proofs and formal proofs. When constructing a generic proof and a formal proof to one claim, the several benefits of the mathematical symbolic language become obvious: While in a generic proof the students ought to explicitly write the

generic argument and to explain its validity, the use of variables expresses the generality of the argument in the formal proof. The perception that formal proofs might be easier to write might lead to the effect that the reasoning with algebraic variables achieved more approval as correct proof at the end of the course.

With regard to proof acceptance, students were asked to rate four given kinds of proofs concerning the following aspects: verification, interpretation as purely empirical verification, existence of counterexamples, conviction, explanation, testing of concrete cases and correctness. In the beginning of the course, the generic proof with numbers and the one with figurate numbers were mostly considered as a single check of examples. Many students did not see the general argument that becomes apparent in the generic examples and that is formulated in the narrative reasoning following these examples. Accordingly, students did not value its explanatory power. But after attending the course, students did value the general verification given by the two generic proofs and rejected the interpretation as a mere check of some examples. In the case of the proof with geometric variables, the students did not state clear positions at the beginning of the course and many of them did not consider this kind of reasoning as explanatory. In the posttest, most of the students agreed that the statement has to be true due to the proof and that the argument is convincing. The interpretation as a mere check of examples was rejected. But in the posttest, this kind of proof was not considered as explanatory by the students. The formal proof was considered the most convincing and explanatory argument, both in the pre- and the posttest.

Here, one can detect students' understanding of the different kinds of proofs. It appears as if students are not used to understand (i.e. to read and validate) proofs expressed in concrete examples. This result is somewhat surprising, because these kinds of generic proofs are recommended ways to perform reasoning and proving at school (e.g. Leiß and Blum 2006). However, our students do not appear to be used to such proofs. Students also struggled with proofs making use of figurate numbers. It may seem obvious to conclude that students struggle with the use of this kind of notation system. Figurate numbers are used in elementary school, for example, to discuss mathematical phenomena as even and odd numbers. Even in middle school, figurate numbers are used in the context of algebra, variables and sequences. But here, the students starting their university studies do not appear to be used to these symbols. We interpret these findings as suggesting that the use of any notational system has to be learned and practiced (compare Dörfler 2008; Jahnke 1984). In reference to the course, one might again emphasize the use of the four kinds of proofs. The students were asked to construct and to compare the four kinds of proofs nearly the entire semester so they could get used to each notational system. When proving one claim with all four kinds of proofs, the advantages of each kind of proof and also one's individual preferences can be experienced. Here, the results concerning the explanatory power of the four kinds of proofs are worth discussing. Generic proofs are said to be proofs that explain (compare Hemmi 2006, p. 44) and Healy and Hoyles (2000) found that the students in their study found narrative proofs more explanatory than formal ones. First, it seems as if our students were not used to the concept of generic proof in the beginning of the course. Therefore, they might have not grasped the inner validity and the explanatory power of the generic argument. But in the posttest, students gave higher ratings to the formal proof concerning explanatory power. One might consider several explanations for this fact. The students in our study were older, so age might have influenced the results. The math classes in school may promote the use of variables continuously in higher grades, so our first-year students had more time to become familiar with this notational

system then the 14 and 15 year old students in the study of Healy and Hoyles. But one might also consider the length of the proofs. While a generic proof contains a narrative reasoning to be read and understood, the formal proof is quite shorter and deals with fewer characters. Accordingly, the shorter proof might be considered the more explanatory one, because there are fewer steps to follow in the proof.

For measuring an overall proof acceptance, a proof acceptance scale was constructed for all four kinds of proof. Using this scale, it is possible to compare students' proof acceptance of the four different kinds of proofs and to measure the changes from the pre- to the posttest. While all acceptance scores increased during the semester, the formal proof always gained the highest acceptance by the students, followed by the generic proof with figurate numbers. To gain more insight, students' individual change (of proof acceptance) scores were also calculated. In these individual results, the proof acceptance scores for three types of proof (the two generic proofs and the proof with geometric variables) of about 70% of the students increased. On the other hand, the scores decreased in about 20% of the students. A possible explanation for this result may be the effect that an increased proof acceptance concerning one kind of proof may lead to a decrease concerning another kind. This interesting observation requires more research. However, we want to emphasize that the proof acceptance scale appears to be a valuable research instrument. With the help of our questionnaire, it became possible to investigate a broader concept of proof understanding, which we call proof acceptance. As shown in our pretest results, students did not accept the generic proofs when starting our course, but the formal proof achieved very high acceptance values. It appears as if the formal proof represents the paradigm of a mathematical proof. Also, the formal proof was also considered the most convincing and explanatory argument. During the course, all acceptance scores increased. But at the end of the course, the formal proof still achieved the highest scores. Accordingly, our course succeeded in broadening students' understanding of mathematical proof. The promotion of the investigation of examples as a natural part of the proving process and the use of generic proofs seem to have worked against the misconception that a simple check of examples does constitute a proof.

A limitation of our study is that students' proof ratings do not only rely on the general kind of proof (generic proof, formal proof, etc.) but are also related to the concrete proof. As we explained earlier, the topics of the presented proofs were different because we wanted to avoid interactions between the judgements of the different types. Therefore, it would be imprudent to overestimate or to generalize these findings in an inappropriate way. These findings may open further discussion about several aspects in the learning and teaching of mathematical proof. We would like to add some questions and research challenges. The phenomenon mentioned in the literature that learners do accept examples as mathematical proof (compare Reid and Knipping 2010, p. 59 ff.) is partly confirmed in this study. However, 17.6% percent of the pre-service teachers rate the mere use of examples as correct proof in the beginning of our course. Future research questions based on this result include: Do the students consider this empirical verification as a correct proof because they do not recognize the general claim inherent in the statement? Or are the students personally convinced by the examples and therefore mark "correct proof"? Do they have a misconception of mathematical proof that leads them to their rating or do they identify a general pattern in the examples that makes them recognize a generic proof in some sense? These questions, also raised by Reid and Knipping (2010, p. 59 ff.) and Weber and Mejia-Ramos (2015), should be addressed in future research.

The use of generic proofs and the use of figurate numbers are listed in the literature in the context of appropriate forms of reasoning for school mathematics. As we have shown above, the pre-service teachers did not perceive the generality quality of the arguments and they seemed to have trouble with the use and meaning of figurate numbers. This creates a potential discussion regarding if and how students at school reason with these kinds of proofs that are not formulated in a formal way.

Finally, we want to touch on the concept of "proofs that explain". In the literature, the concept of proofs that explain is often illustrated by proofs making use of geometric representations (e.g., Hanna 1990, p. 11). But in our study, the students valued the explanatory power of these kinds of proofs less than we expected on the basis of the literature. But they did value the explanatory power of the formal proof. As Jahnke (1984) argues, representations are neither self-evident nor self-explanatory. From a semiotic point of view, working in a notational system like figurate numbers has to be learned (as proposed in the context of diagram literacy in Diezmann and English (2001)). This view highlights the fact that proofs cannot be considered to be explanatory by themselves. Students have to learn to deal with any notational system. Accordingly, the question of explanatory power or conviction of a proof cannot be judged in a general way, but has to be answered individually regarding someone's prior experiences. However, based on our findings, it appears appropriate to consider the explanatory power of the mathematic symbolic language, to rethink the concept of proofs that prove and proofs that explain.

Compliance with Ethical Standards

Conflict of Interest Statement   On behalf of all authors, the corresponding author states that there is no conflict of interest.

Publisher's Note   Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

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 8th Congress of the European Society for* Research in Mathematics Education (pp. 86–95). Ankara: Middle East Technical University.

Blum, W., Biehler, R., & Hochmuth, R. (2015). KLIMAGS: Researching and improving mathematics courses for future primary teachers. In R. Biehler, R. Hochmuth, C. Hoyles & P. W. Thompson (Eds.), *Mathematics in undergraduate study programs: Challenges for research and for the dialogue between mathematics and didactics of mathematics. Oberwolfach Report No. 56/2014* (pp. 50–51). Oberwolfach.

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, 7/8*. Online: http://www.lettredelapreuve.org/OldPreuve/Newsletter/990708 Theme/990708ThemeUK.html. Accessed 27 Aug 2018.

de Villiers, M. (1990). The role and function of proof in mathematics. *Pythagoras, 24*, 17–24.

Diezmann, C., & English, L. (2001). Promoting the use of diagrams as tools for thinking. In A. Cuoco & F. R. Curcio (Eds.), *The roles of representation in school mathematics* (pp. 77–89). Reston, Virginia: National Council of Teachers of Mathematics.

Dörfler, W. (2008). Mathematical Reasoning: Mental Activity or Practice with Diagrams. In M. Niss (Ed.), *ICME 10 Proceedings, Regular Lectures* [CD-Rom]. Roskilde: IMFUFA, Roskilde University.

Dreyfus, T. (2000). Some views on proofs by teachers and mathematicians. In A. Gagatsis (Ed.), *Proceedings of the 2nd mediterranean conference on mathematics education* (Vol. I, pp. 11–25). Nikosia, Cyprus: The University of Cyprus.

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.

Flores, A. (2002). Geometric representations in the transition from arithmetic to algebra. In F. Hitt (Ed.), *Representation and mathematics visualization* (pp. 9–30). Mexico: North American Chapter of the International Group for the Psychology of Mathematics Education.

Gueudet, G. (2008). Investigating the secondary – Tertiary transition. *Educational Studies in Mathematics, 67*(3), 237–254.

Hanna, G. (1990). Some pedagogical aspects of proof. *Interchange, 21*(1), 6–13.

Hanna, G. (2018). Reflections on proof as explanation. In A. J. Stylianides & G. Harel (Eds.), *Advances in mathematics education research on proof and proving: An international perspective* (pp. 3–18). Cham: Springer International Publishing.

Healy, L., & Hoyles, C. (1998). *Justifying and proving in school mathematics, executive summary.* London: Institute of Education, University of London.

Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education, 31*(4), 396–428.

Hemmi, K. (2006). *Approaching proof in a community of mathematical* practice. Unpublished Doctoral Thesis, Stockholm University, Stockholm. Online: http://www.diva-portal.org/smash/get/diva2:189608/FULLTEXT01.pdf. Accessed 02 Aug 2018.

Hersh, R. (1993). Proving is convincing and explaining. *Educational Studies in Mathematics, 24*(4), 389–399.

Jahnke, H. N. (1984). Anschauung und Begründung in der Schulmathematik. In Gesellschaft für Didaktik der Mathematik (Ed.). Beiträge zum Mathematikunterricht (pp. 32–41). Bad Salzdethfurt: Franzbecker.

Kempen, L., & Biehler, R. (2016). Pre-service teachers' perceptions of generic proofs in elementary number theory. In K. Krainer & N. Vondrová (Eds.), *Proceedings of the 9th congress of the European Society for Research in mathematics education* (pp. 135–141). Prague: CERME.

Kirsch, A. (2000). Aspects of simplification in mathematics teaching. In I. Westbury, S. Hopmann & K. Riquarts (Eds.), *Teaching as a reflective practice. The German Didaktik Tradition* (pp. 267–284). Mahwah: Lawrence Erlbaum Associates.

Kulpa, Z. (2009). Main problems of diagrammatic reasoning. Part I: The generalization problem. *Foundations of Science, 14*(1–2), 75–96.

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, Taiwan: The Department of Mathematics, National Taiwan Normal University. Online: http://140.122.140.1/~icmi19/files/Volume_2.pdf. Accessed 27 Aug 2018.

Leiß, D., & Blum, W. (2006). Beschreibung zentraler mathematischer Kompetenzen. In W. Blum, C. Drüke-Noe, R. Hartung, & O. Köller (Eds.), *Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen* (pp. 33–80). Berlin: Cornelsen Scriptor.

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

Malek, A., & Movshovitz-Hadar, N. (2011). The effect of using transparent pseudo-proofs in linear algebra. *Research in Mathematics Education, 13*(1), 33–57.

Martin, W. G., & Harel, G. (1989). Proof frames of preservice elementary teachers. *Journal for Research in Mathematics Education, 20*, 41–51.

Mason, J., & Pimm, D. (1984). Generic examples: Seeing the general in the particular. *Educational Studies in Mathematics, 15*(3), 277–289.

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

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

Nelsen, R. B. (1993). *Proofs without words: Exercises in visual thinking.* Washington, DC: The Mathematical Association of America.

Nelsen, R. B. (2000). *Proofs without words 2: More exercises in visual thinking.* Washington, DC: The Mathematical Association of America.

Plomp, T. (2010). Educational design research: An introduction. In T. Plomp & N. Nieven (Eds.), An Introduction to Educational Research. (pp. 9–35). Enschede: John Wiley.

Rasmussen, C., & Kwon, O. N. (2007). An inquiry-oriented approach to undergraduate mathematics. *The Journal of Mathematical Behavior, 26*(3), 189–194.

Reid, D. A., & Knipping, C. (2010). Proof in mathematics education. In *Research, learning and teaching*. Rotterdam: Sense Publisher.

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.

Reiss, K., Heinze, A., & Klieme, E. (2000). Argumentation, proof and the understanding of proof. In A. Peter-Koop, K. Reiss, G. Törner, & B. Wollring (Eds.), *Developments in mathematics education in Germany. Selected papers from the annual conference on didactics of mathematics, potsdam 2000* (pp. 109–120). Hildesheim: Franzbecker.

Rowland, T. (1998). Conviction, explanation and generic examples. In A. Olivier & K. Newstead (Eds.), *Proceedings of the 22nd Conference of the International Group for Psychology of Mathematics Education* (Vol. 4, pp. 65–72). Stellenbosch: University of Stellenbosch.

Rowland, T. (2002). Generic proofs in number theory. In S. R. Campbell & R. Zazkis (Eds.), *Learning and teaching number theory* (pp. 157–183). Westport: Ablex.

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. (2003). Validations of proofs considered as texts: Can undergraduate tell whether an argument proofs a theorem? *Journal for Research in Mathematics Education, 34*, 4–36.

Selden, A., & Selden, J. (2017). A comparison of proof comprehension, proof construction, proof validation and proof evaluation. In R. Göller, R. Biehler, R. Hochmuth, & H.-G. Rück (Eds.), *Didactics of Mathematics in Higher Education as a Scientific Discipline – Conference Proceedings. khdm-Report 17–05* (pp. 339–345). Kassel: Universität Kassel Online: https://kobra.bibliothek.uni-kassel.de/bitstream/urn:nbn:de:hebis:34-2016041950121/5/khdm_report_17_05.pdf. Accessed 27 Aug 2018.

Selden, A., Selden, J., & Benkhalti, A. (2015). An analysis of transition-to-proof course students' proof constructions with a view towards course redesign. *Proceedings of the 17th Annual Conference on Research in Undergraduate Mathematics Education*. Online: https://www.researchgate.net/publication/268278836_An_Analysis_of_Transition-to-Proof_Course_Students'_Proof_Constructions_with_a_View_Towards_Course_Redesign. Accessed 30 July 2018.

Smith, J. C. (2006). A sense-making approach to proof: Strategies of students in traditional and problem-based number theory courses. *Journal of Mathematical Behavior, 25*, 73–90.

Sommerhoff, D., Ufer, S., & Kollar, I. (2016). Proof validation aspects and cognitive student prerequisites in undergraduate mathematics. In C. Csíkos, A. Rausch, & J. Szitányi (Eds.), *Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 219–226). Szeged: PME.

Stylianides, A. J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education, 38*(3), 289–321.

Stylianides, A. J. (2010). Engaging secondary students in reasoning and proving. *Mathematics Teaching, 219*, 39–44.

Tabach, M., Barkai, R., Tsamir, P., Tirosh, D., Dreyfus, T., & Levenson, E. (2010). Verbal justification - is it a proof? Secondary school teachers' perceptions. *International Journal of Science and Mathematics Education, 8*(6), 1071–1090.

Tabach, M., Levenson, E., Barkai, R., Tsamir, P., Tirosh, D., & Dreyfus, T. (2011). Secondary teachers' knowledge of elementary number theory proofs: The case of general-cover proofs. *Journal of Mathematics Teacher Education, 14*(6), 465–481.

Weber, K. (2008). How mathematicians determine if an argument is a valid proof. *Journal for Research in Mathematics Education, 39*(4), 431–459.

Weber, K. (2010). Mathematics majors' perceptions of conviction, validity. and proof. *Mathematical Thinking and Learning, 12*(4), 306–336.

Weber, K. (2014). Proof as a cluster concept. In C. Nicol, S. Oesterle, P. Liljedahl, & D. Allan (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. 5, pp. 353–360). Vancouver: PME.

Weber, K., & Mejia-Ramos, J. P. (2015). On relative and absolute conviction in mathematics. *For the Learning of Mathematics, 35*(2), 15–21.

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education, 27*(4), 458–477.