



# Acceptance criteria for validating mathematical proofs used by school students, university students, and mathematicians in the context of teaching

Daniel Sommerhoff<sup>1</sup> · Stefan Ufer<sup>1</sup>

Accepted: 6 March 2019 / Published online: 18 March 2019  
© FIZ Karlsruhe 2019

## Abstract

Although there is no generally accepted list of criteria for the acceptance of proofs in mathematical practice, judging the acceptability of purported proofs is an essential aspect of handling proofs in daily mathematical work. For this reason, school students, university students, and mathematicians need to hold certain acceptance criteria for proofs, which are closely tied to their epistemology, notion of proof, and concept of evidence. However, acceptance criteria have so far received little academic attention and results have often been related to mathematicians' acceptance criteria in a research context. Still, as the role of proof changes during mathematical education and differs from research, it can be assumed that acceptance criteria differ substantially between educational levels and different communities. Thus, we analyzed and compared acceptance criteria by school students, university students, and mathematicians in the context of teaching. The results obtained reveal substantial cross-sectional differences in the frequency and choice of acceptance criteria, as well as major differences from those acceptance criteria reported by prior studies in research contexts. Structure-oriented criteria, regarding the logical structure of a proof or individual inferences, are most often referred to by all groups. In comparison, meaning-oriented criteria, such as understanding, play a minor role. Finally, social criteria were not mentioned in the context of teaching. From an educational perspective, the results obtained underline university students' need for support in implementing acceptance criteria and suggest that a more explicit discussion of the functions of proof, their evidential value, and the criteria for their acceptance may be beneficial.

**Keywords** Mathematical proof · Acceptance criteria · Enculturation · Argumentation · Proof validation · Socio-mathematical norms

## 1 Introduction

Mathematical proof can be considered a key method of mathematics as a scientific discipline, particularly with regard to its function as bearer of mathematical knowledge and evidence (Mariotti 2006). Proof and argumentation in general play a central role in mathematics education (e.g., Hanna and Jahnke 1996; Heinze and Reiss 2009) and are implemented in standard documents worldwide (e.g., CCSSI 2010). Typically, these emphasize informal mathematical reasoning and argumentation in lower grades, focusing

on understanding and explanation as underlying functions (Hanna 2000). With increasing educational level, arguments are systematized, formalized, and directed towards underlining the validity of a given claim, resulting in the concept of mathematical proof (e.g., Heinze and Reiss 2009). Furthermore, the idea of axioms, an emphasis on the role of deductive inferences, and the use of symbolic notation extend the concept of proof with respect to its systematization and validation functions. Overall, the concept of proof, its functions, and the according norms and values change substantially over time and differ between contexts and communities.

Empirical studies have repeatedly shown that handling proof is difficult for learners (e.g., Healy and Hoyles 2000), yet proof appears to be particularly demanding at the transition from school to university (e.g., Clark and Lovric 2009; Coriveau and Bednarz 2017; Moore 1994; Selden 2011; Selden and Selden 2003; Tall 2008; Weber 2003). This difficulty is often attributed to a shift in the role of mathematical proof

---

✉ Daniel Sommerhoff  
sommerhoff@math.lmu.de

<sup>1</sup> Chair of Mathematics Education, Ludwig-Maximilians-Universität München, Theresienstr. 39, 80333 Munich, Germany

between school and university, when for example validity becomes the main function (e.g., Tall 2008). Moreover, these studies have underlined that the adaptation and enculturation to the new norms and values is particularly difficult for students.

To examine students' difficulties at the transition to university and the differences regarding proof related norms and values in school and university communities, school students', university students', and mathematicians' *acceptance criteria* for proofs can be analyzed. These are individual criteria used to decide if a purported proof is acceptable or not. They can be assumed to reflect the social norms and values regarding mathematical proof. Acceptance criteria are necessary for validating and constructing proofs and can thus be ascribed a central role for handling proofs successfully. They can also be seen as important indicators for students' understanding of mathematics, its epistemology as a science, and their understanding of mathematical evidence (e.g., Mariotti and Balacheff 2008). However, so far there are few quantitative studies regarding the use of acceptance criteria for mathematical proofs, in particular, neither cross-sectional studies nor those comparing teaching and research contexts. Previous research has either focused on conceptions of proof more generally (e.g., Harel and Sowder 1998; Hemmi 2006) or on criteria used by mathematicians within a research context. However, based on theoretical considerations and first qualitative analyses (e.g., Weber 2008) these are unlikely to mirror the acceptance criteria in a teaching context or those held by school or university students.

As indicated above, there appears to be a shift in the acceptability of proof, both at the transition from school to university and from university to research, which has yet to be addressed by research. The presented study thus evaluated whether school students (at the end of their school studies) and university students detect violations of central norms for mathematical proofs when validating proofs and whether they can provide according justifications. Furthermore, the acceptance criteria used by school and university students were systematically compared to analyze differences in the local norms regarding proof between both contexts. This comparison is complemented with the investigation of research mathematicians' use of acceptance criteria in the context of undergraduate university teaching. Finally, acceptance criteria used by mathematicians in the context of teaching were compared with those reported from the context of research (e.g., Hanna 1989; Heinze 2010).

## 2 Acceptance criteria for mathematical proofs in different communities

The terms mathematical reasoning, arguing, and proving, respectively reason, argument(ation), and proof, are often vaguely defined and possess multiple interpretations (e.g.,

Reid and Knipping 2010). In particular, they are still widely discussed (e.g., Aberdein 2009; Balacheff 2008; Manin 2010), and perspectives of what constitutes an argument or proof vary both in practice and research. Still, most authors agree that (of these terms) mathematical proof represents the narrowest concept, often interpreted as referring to mathematical arguments that meet certain *socio-mathematical norms* (Yackel and Cobb 1996), for example regarding their formal representation. However, Hanna and Jahnke (1996, p. 878) asserted that there are no “generally accepted criteria for the validity of a mathematical proof” (in mathematical practice).

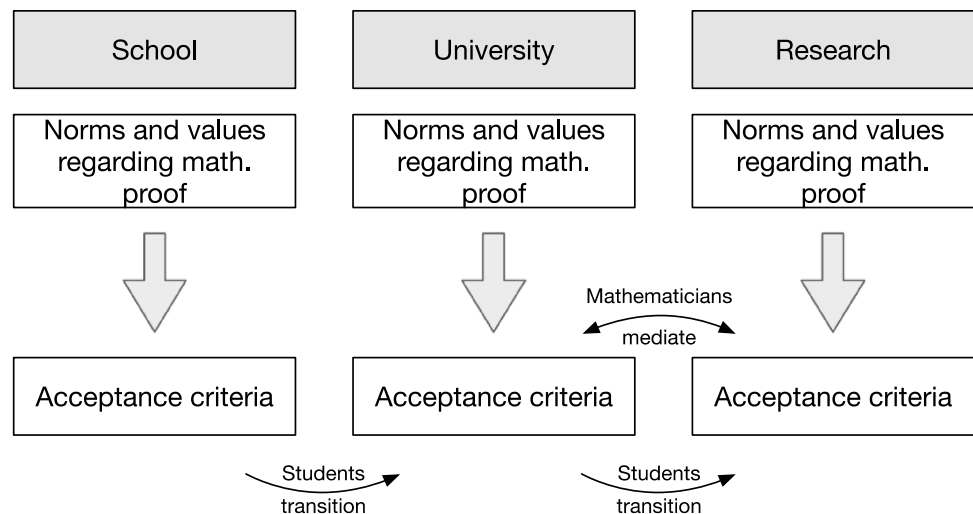
Still, judging the acceptability of mathematical proofs is an essential aspect of engaging with proofs. Accordingly, school students, university students, and mathematicians must hold certain individual criteria that they use *locally* to judge and justify the acceptability of proofs. In the following, these will be referred to as *acceptance criteria* for mathematical proofs, and these entail all criteria that are used in mathematical practice to judge whether a proof is acceptable or not. In particular, criteria for accepting or rejecting a proof are partially intrinsically linked, as some criteria can be regarded as logically ‘necessary’ or ‘sufficient’: For example, the use of circular reasoning is sufficient in order to reject a proof. At the same time, the absence of circular reasoning is a necessary condition for accepting a proof. Thus *circular reasoning* may be used as a criterion for accepting and for rejecting a proof.

### 2.1 The local proof culture and enculturation

That there is no generally accepted list of criteria for the acceptability of mathematical proofs is plausible from a perspective that sees proof as embedded in a social practice. For example, Manin (2010) claimed that “a proof only becomes a proof after the social act of ‘accepting it as a proof’” (p. 45). Accordingly, acceptance criteria for mathematical proofs have to be seen as *socio-mathematical norms* (Reid and Knipping 2010; Yackel and Cobb 1996). That is, these criteria are, often implicitly, negotiated socially within certain communities of practicing mathematicians. Thus, acceptance criteria may vary between communities, as illustrated by discussions about the acceptability of computer-assisted proofs.

It is certainly possible to distinguish a multitude of mathematical communities, that is, groups of people jointly practicing mathematics (see further Lave and Wenger 1991). Still, looking at the usual trajectories when learning to prove, three types of communities can be distinguished (Fig. 1): school communities, university (education) communities, and research communities. Usually, students pass through these sequentially, yet reaching a mathematical research community happens neither necessarily nor particularly often.

**Fig. 1** Illustration of three communities in a ‘learning to prove’-trajectory, students’ transitions, and mathematicians’ mediating role between communities



Within school communities, first proofs are often introduced in 7th–9th grade in the context of geometry (e.g., Balacheff 2010), for example as a specific kind of argumentation. In school, proofs are used locally and are not part of a global deductive, axiom-driven theory. Moreover, didactical proof conceptions, such as empirical-inductive arguments are often accepted in school, especially in lower grades, and are stressed as valuable for school learning (Hanna and Jahnke 1996). Within university communities, however, mathematics is understood as a proof-based, axiomatic-deductive system that is “recreated” in first-year lectures (Tall 2008). This systematic approach requires proofs focusing on validity. Finally, research communities operate inside the axiomatic-deductive system created at university, for example when formulating and proving novel theorems. This regularly involves complex and highly-specialized proofs, whose epistemic status may be unclear for some time and which is also often not verifiable by all practicing mathematicians (cf. Devlin 2003).

Overall, becoming a member of one of the three substantially different types of mathematical communities requires an *enculturation process* (see further Schoenfeld 1992), that is, a process in which an individual acquires cultural goods such as acceptance criteria for mathematical proofs, by being exposed to and interacting with a community.

Today, little is known about learning to prove as an enculturation process. There appears to be no specific model to describe this process, and only very first evidence on enculturation processes regarding the norms and values underlying mathematical proofs exists. A longitudinal study (Bieda et al. 2006) showed that enculturation in school is an ongoing process happening over years. The surveyed students used general arguments increasingly often from 6th to 9th grade and appeared to have an increasing understanding of the value of arguments for justifying general mathematical statements. A study by Perrenet and Taconis (2009), focusing on enculturation effects upon entering university in the

context of problem solving, found significant shifts in students’ beliefs and behavior towards those of the local community, represented by their teachers. Accordingly, it can be assumed that enculturation processes not only are happening when changing communities, but they may be especially profound at these transitions. Furthermore, empirical data by Perrenet and Taconis (2009), theoretical considerations, and qualitative data (e.g., Lave and Wenger 1991) suggest that enculturation processes aim at an increasing adaptation of newcomers (students) to the established norms and values in the particular community and, in the context of mathematical proofs, ultimately to those from a research context.

Regarding the three adjacent mathematical communities, mathematicians can be seen as intermediary persons: They are part of a community of *university mathematics teachers and learners* and part of a community of *mathematics researchers* (Fig. 1). These communities may differ, both in terms of their members and their norms and values regarding mathematical proofs. An important aspect of mathematicians’ work is thus to mediate between both communities, as research practice informs and constrains classroom practice (Dawkins and Weber 2017).

## 2.2 Research on acceptance criteria<sup>1</sup>

Acceptance criteria can be seen as more or less pragmatic instantiations of certain norms and values regarding mathematical proof, which are held by the (local) mathematical

<sup>1</sup> Assuming a social nature of acceptance criteria for proof, relevant research naturally has to be seen as specific to the associated context (e.g., country, community, classroom,...). Thus, interpreting, transferring, and generalizing research results has to be done carefully and ideally requires empirical underpinning. Therefore, references to results from multiple communities and countries are given in this paper.

community (e.g., the members of a school or university; see also Dawkins and Weber 2017), for the daily mathematical work with proofs (see Fig. 1, vertical direction).

However, it is still a philosophical debate, how ‘local’ these norms, values, and according criteria are (or should be) and whether what constitutes a proof depends on absolute (global) (e.g., Azzouni 2004; Hilbert 1931) or on local, social criteria within a specific mathematical community (e.g., Reid and Knipping 2010; Yackel and Cobb 1996). Both sides have a point: it is a reasonable question, which global criteria could be defined (normatively) for every ideal, complete proof (derivation). Still, given that derivations are rarely produced in mathematical practice (and the connection of a real proof text to a corresponding derivation is far from trivial to describe; see below and Tanswell 2015), it is also reasonable to investigate which (socially formed, local) criteria students and mathematicians use in their own practice to accept or reject a proof.

Both perspectives can be used to create a list of possible acceptance criteria for mathematical proofs: a deductive, logical-structural perspective emanating from the concept of proof as an ideal proof, or a social-descriptive perspective examining mathematical practice and the acceptance criteria, norms, and values reported or used within.

A starting point to describe acceptance criteria from a logical-structural perspective is the notion of *formal proofs* or *derivations*, that is finite sequences of sentences, each of which are either axioms, given premises, consequences from the preceding sentences by valid rules of inference, or assumptions that are justified subsequently (e.g., Azzouni 2004; Hilbert 1931, p. 489). Derivations have been considered the ideal standard for proof for some time, but were critiqued and discussed heavily in the literature (e.g., Hanna 2000). Aberdein (2009) underlined that “not all—indeed hardly any—mathematical proofs are strict formally valid logical derivations” (p. 1), especially not within mathematical practice and teaching. Some authors thus propose an *in-principle formalizability* as a characteristic of acceptable proofs (e.g., Alama and Kahle 2013). That is, an acceptable proof may contain gaps, as long as it can be transformed into a corresponding formal proof. Still, this view has also been challenged on plausible grounds (Tanswell 2015). If accepted as a characteristic, however, it would imply that a proof should (a) only use axioms or premises, already proven arguments, or arguments that are easy to prove, (b) only use allowed rules of inference, and (c) have a logical structure that ends at the claim to be proved. These criteria correspond to the three categories of knowledge about the role of mathematical proofs and their acceptance (*methodological knowledge*; Heinze and Reiss 2003), which can be interpreted as acceptance criteria, as follows. *Proof scheme* refers to the types of inference that are acceptable in mathematical proofs. *Logical chain* refers to the local validity of

each individual step of a proof, in particular, that warrants have been shown in previous steps or are part of the *shared knowledge* (e.g., Lakatos 1963). While the latter criteria focus on local aspects of a proof, *proof structure* focuses on the overall structure of a proof and the question of whether it reduces the claim completely to axioms, prerequisites and proved statements, and does not, for example, contain circular reasoning (*petitio principii*).

A social-descriptive approach to acceptance criteria for proofs has been taken by Hanna (1989), claiming that “mathematicians accept a new theorem only when some combination of the following holds” (p. 879): they understand the theorem; the theorem is significant enough; the theorem is consistent with the body of accepted results; the author has an unimpeachable reputation; there is convincing mathematical argument for it. Partially based on these findings, Heinze (2010) conducted an online-study with 40 mathematicians and provided evidence that the three most frequently reported reasons for mathematicians to accept a proof in a research context are that they checked it themselves, that it was produced by colleagues with high standards, and that it was published long ago without contradiction.

Another social-descriptive approach has been taken by Dawkins and Weber (2017), who identified four *values* that they perceived as held by the mathematical community, relating to the following: (a) use of a priori arguments, (b) the a-contextuality of knowledge and justifications, (c) a desire to increase understanding of mathematics, and (d) the desire for a set of consistent proof standards. Based on these values, they formulated multiple *norms* upholding these values, for example that justifications are deductive and do not admit rebuttals.

Based on these sources, different categories of acceptance criteria emerge: First, based on the findings by Hanna (1989) and Heinze (2010), there appear to be several *social acceptance criteria*, for example relating to the reputation of the author. These can be related to an *authoritarian proof scheme* and underline the social dimension in the acceptance of proofs.

Besides social criteria, which base the acceptability of proofs on others’ decisions and behavior, two other categories emerge: *Structure-oriented criteria*, that is, criteria meant to ensure the formal validity of the proof, and *meaning-oriented criteria*, meant to ensure goals such as understanding, the consistency with one’s own prior mathematical knowledge, or more aesthetically related goals such as beauty or simplicity.

The difference between the latter categories can be illustrated based on an argument by Weber and Mejia-Ramos (2015), who differentiated between relative and absolute conviction. *Structure-oriented criteria* can be interpreted as aiming to establish absolute conviction, that is, to assert that the given statement is valid. In contrast,

*meaning-oriented criteria* try to establish relative conviction, that is, trust that the statement holds with a high probability. For example, an iconic representation may establish high relative conviction, yet may not establish absolute conviction. Accordingly, the iconic argument may be rejected as a proof based on *structure-oriented criteria*, but likely not by *meaning-oriented criteria*.

Finally, this distinction also aligns with research on the functions of mathematical proofs. Whereas *structure-oriented criteria* relate to the verification function, *meaning-oriented criteria* relate to functions such as explanation or understanding (de Villiers 1990; Hanna and Jahnke 1996).

Although some research on the notion of proof in general (Hemmi 2006) and the acceptance of proofs in particular (Harel and Sowder 1998; Healy and Hoyles 2000) exist, there is little quantitative evidence on acceptance criteria held by school and university students, and only a bit more regarding mathematicians in a research context. In the research context, some acceptance criteria appear to relate more to the statement or theorem to prove (e.g., ‘they understand the theorem’) than to the actual proof. It is likely that these criteria are particularly important in mathematicians’ research practice, where often *statement* and *proof* are questioned. However, statements in teaching settings are often either simply accepted as true by the students or implied to be true (e.g., ‘Prove that...’-tasks), thus creating an epistemologically different situation. Accordingly, differences in the acceptance criteria used by researchers as contrasted with students or mathematicians in a teaching context can be expected.

### 2.3 Proof validation and acceptance criteria

Acceptance criteria for proofs can be regarded as important during the construction of proofs (see Ufer et al. 2009), as they serve as benchmarks to check whether the proof is acceptable or if parts need to be improved. Still, the actual acceptance criteria remain mostly implicit. *Proof validation*, in contrast, comprising the reading of purported proofs aimed at judging their acceptability (e.g., Alcock and Weber 2005; Selden and Selden 2003) explicitly requires using and verbalizing acceptance criteria when giving a justification for the judgement. Thus, data from proof validation may provide important information about acceptance criteria. Not only the justifications may help to identify the relevant acceptance criteria, but also the judgments (accept or reject) may indicate their proper implementation. In particular, the conjunction of both sources appears most beneficial.

### 2.4 School students

In a large study with high-attaining students, Healy and Hoyles (2000) showed that students had problems correctly validating proofs. Although a majority (between 54 and 60%) were able to correctly judge the (un)acceptability of empirical arguments, about a third failed to do so. Furthermore, students accepted arguments given in symbolic notation more frequently than narrative proofs. These results are partially underlined by an interview study with 22 students by Bieda and Lepak (2014), unveiling that besides deductive reasoning, also understanding and being convincing are important acceptance criteria for school students.

### 2.5 University students

Based on the prominent role of proof during university education, there are several studies focusing on students’ proof validation. Analyses by Alcock and Weber (2005) and Selden and Selden (2003) reveal mediocre results for students’ judgments, which often focused on structure-oriented criteria relating to the logical chain, such as the proper use of definitions or missing or incorrect warrants. Still, Selden and Selden (2003, p. 27) highlighted that although students check multiple aspects of the purported proofs, their first judgments “yield no better than chance results”, implying problems in the implementation of their acceptance criteria. Furthermore, results reveal that undergraduates also have problems in justifying their judgments. Multiple studies suggest that students focus on surface features and local properties of proofs such as the individual validity of single inferences (e.g., Inglis and Alcock 2012; Selden and Selden 2003).

### 2.6 Mathematicians

In contrast, mathematicians appear to consider local and global aspects of proofs (Weber 2008; Weber and Mejía-Ramos 2011), thus considering structure-oriented criteria corresponding to the proof structure and logical chain. However, Inglis and Alcock (2012) were not able to confirm mathematicians’ attendance to global aspects of proofs using eye-movement data (Weber et al. 2013).

A study on proof validation by Inglis et al. (2013) with research mathematicians underlined the variance in mathematicians’ acceptability judgments of proofs in the context of teaching. Mathematicians referred to the logical chain of the proof, the overall proof structure, as well as to explicit errors; these are acceptance criteria that can be regarded as structure-oriented.

Moreover, Weber (2008) and Weber and Mejía-Ramos (2011) showed that mathematicians use a variety of modes of reasoning during proof validation, some being rather formal



corresponding to structure-oriented criteria, others rather focusing on meaning-oriented criteria (e.g., understanding why individual examples work; Weber 2008, p. 441ff.). Moreover, several authors (e.g., Heinze 2010; Weber 2008; Weber and Mejía-Ramos 2011) have argued that the conception of mathematics as a purely deductive-axiomatic practice is idealized and unrealistic and that mathematicians obtain conviction from both empirical and authoritative sources. Still, it is unclear if this result extends also to the context of teaching, or if mathematicians focus on formal aspects when teaching proof and proving.

### 3 The current study

Handling mathematical proofs is part of school students', university students', and mathematicians' practice. For this reason, they have to be able to evaluate given or self-constructed proofs. Their criteria for judging these proofs reflect the (implicit) norms and values of the local mathematical community. Although acceptance criteria appear to be highly important for the successful handling of mathematical proofs, there is currently neither a satisfactory theoretical description of acceptance criteria for mathematical proofs (within mathematical practice) nor is there sufficient empirical evidence on the acceptance criteria used by school students, university students, and mathematicians in a teaching context.

As the transition from school to university appears to be crucial for learning to prove and enculturation into the university community, the present study focuses on the acceptance criteria used by students at the end of school, university mathematics students in their first semesters, and mathematicians in the context of teaching, and compares these groups cross-sectionally.

The study was guided by the following main questions:

(RQ 1) Do school and university students detect violations of central norms for mathematical proofs when validating proofs? Do they provide justifications for their judgments that *match* the violations contained in the given purported proofs?

Based on prior results, we expected problems by both groups when validating proofs. Moreover, we expected pronounced differences in their judgments and justifications based on the violated norm.

(RQ 2) Which acceptance criteria do school students, university students, and mathematicians refer to when accepting or rejecting purported proofs in the context of teaching?

As there are currently no quantitative data available for these groups, no specific hypotheses were made. Still, based on prior research on proof validation, we expected that all three groups would refer to structure-oriented criteria as well as meaning-oriented criteria. Furthermore, based on the reported problems at the transition from school to university and the shift in the role of proof, we expected university students to refer more to structure-oriented criteria than school students do, but also to struggle to implement them. Finally, based on the findings by Perrenet and Taconis (2009), we expected that the alignment of criteria with those of mathematicians would increase from school to university students.

(RQ 3) Which acceptance criteria do mathematicians indicate as suitable for ensuring or rejecting the acceptability of proofs in the context of university teaching? What are differences and commonalities when this context is compared with the research context?

Proofs by students typically correspond to propositions that are presumed to hold. Accordingly, we expected that the criteria offered by mathematicians in the teaching context would deviate from those in the research context, where the acceptability of the statement and its proof are in question. Such differences are also supported by results from Weber (2008). Furthermore, we expected that mathematicians would not use social acceptance criteria in the teaching context, in contrast to the research context.

### 4 Method

To answer these questions, questionnaire data were analyzed from all three groups. For school students, data were gathered from future mathematics students ( $N_1 = 114$ , 59 male, 47 female, 8 NA) in a voluntary preparation course for their university mathematics studies (bachelor, secondary school teaching degree). This sample was chosen as these students were assumed to be well enculturated into school mathematics, as they had completed the whole length of school education and had not started their university mathematics studies yet. Moreover, participants had decided to study mathematics at university, allowing a better comparison with the university student sample than a random school student sample. The university student group was surveyed in the first session of a voluntary course on proving, which consisted of mathematics students ( $N_2 = 66$ , 24 male, 41 female, 1 NA; bachelor, secondary school teaching degree) at the end of their first or third semester. Mathematicians' data were gathered from research mathematicians from German universities ( $N_3 = 273$ , 217 male, 50 female, 6 NA; 170

**Fig. 2** The proposition used in the study, an example purported proof, and according items

The purpose of the following tasks is **not** to prove the proposition yourself, but instead to judge the four purported proofs created by students.

**Proposition**

The product of three arbitrary consecutive integers is divisible by 3.

**Andrew's proof**

Let  $a \in \mathbb{Z}$  be an arbitrary integer. Accordingly the two consecutive integers can be written as  $a + 1$  and  $a + 2$ .

We are interested in the product

$$a \cdot (a + 1) \cdot (a + 2)$$

Expanding the term yields:

$$a \cdot (a + 1) \cdot (a + 2) = a \cdot (a^2 + 3a + 2) = a^3 + 3a^2 + 2a$$

Since the proposition should hold for an arbitrary integer  $a$ , the statement has to hold independently of  $a$ . We therefore only look at the coefficients of the term  $a^3 + 3a^2 + 2a$ . For the sum of these coefficients we get:

$$1 + 3 + 2 = 6$$

Accordingly since  $3|6$  holds, it also holds that  $3|a^3 + 3a^2 + 2a$  resp.  $3|a \cdot (a + 1) \cdot (a + 2)$ . Thus the product of three arbitrary consecutive integers is divisible by 3 and we have proven the proposition.

1. Is this "proof" a correct mathematical proof?

Yes  No  I don't know

2. Please give a short justification for your choice:

doctoral students, 53 post-docs, 16 lecturers, 31 professors, 3 NA) using an online questionnaire.

All groups received a questionnaire<sup>2</sup> with two sections, the first requesting demographic data. In the second section, participants were introduced to a proposition from elementary number theory (Fig. 2) and four purported proofs for this proposition that were presented as proofs by fellow students (for school and university students) or by university students in their first semester (for mathematicians). The questionnaire contained one acceptable proof as well as three unacceptable proofs (according to our socio-mathematical norms). The unacceptable proofs contained an incorrect warrant, circular reasoning, or inductive reasoning. Based on the definition of a formal proof, these can be seen as violations of core norms with regard to proofs. All three groups were asked to specify whether the purported proofs could be classified as a "correct mathematical proof" (closed format; Fig. 2) and subsequently to provide a justification for their judgment (open format).

Mathematicians received a third section that introduced the scenario of a student tutor having problems grading first-year students' homework. Mathematicians were asked to state what the tutor had to focus on to make sure that a

purported proof is certainly "correct" or "incorrect" (two open items; "What does a student tutor have to look for when grading students' homework to determine if a purported proof is definitely (in)correct?").

#### 4.1 Verification of the employed purported proofs

As acceptance criteria for mathematical proofs are socially negotiated, the acceptance of a proof may differ between communities. To ensure that the purported proofs presented in this study can be 'rightfully seen' as (un)acceptable as intended, mathematicians' judgments were used as reference. The results obtained (Table 1) indicate that all proofs can be seen as (un)acceptable as intended, as there was almost unison agreement and variation was often justified by comments such as "the proof could be easily changed to a correct proof".

#### 4.2 Coding of acceptance criteria

The open items from the proof validation section were coded in two ways. First, it was analyzed whether they matched the errors implemented in the incorrect proofs (dichotomous rating), in order to judge whether they were based on the intended errors or on other aspects of the proofs.

Second, the justifications were segmented and coded to extract the used acceptance criteria. Coding was based on

<sup>2</sup> The questionnaire can be requested for research purposes from the authors.

**Table 1** Mathematicians' judgments on the acceptability of the purported proofs in the validation section

	Correct (%)	Incorrect (%)	Don't know (%)	Missing (%)
Acceptable proof	97.8	1.5	0.7	–
Proof with an incorrect warrant	2.9	88.3	7.3	1.5
Proof with circular reasoning	3.3	95.2	1.1	0.4
Proof with inductive reasoning	1.5	97.8	0.7	–

– no missing data

**Table 2** Main-categories used in the coding scheme

Deductive categories		Inductive categories	
Structure-oriented	Meaning-oriented	Criterial reasoning	Non-criterial
Proof structure	Understanding	Trivializations	Suggestions for improvement
Proof scheme	Consistency with prior knowledge	Unambiguousness	Mathematical proof
Logical chain	Aesthetics	All premises were used	Other
Counterexamples			

a deductively derived coding scheme emanating from the theoretical discussion above, complemented by inductively derived categories that occurred during the coding process (Mayring 2014). The coding scheme (Table 2) comprised 13 main-categories for acceptance criteria. As inductive categories, *trivializations* was added for comments relating to the unacceptability of generic justifications such as “this is trivial”, *unambiguousness* for comments requiring proofs to be unambiguous, and *all premises were used* for comments mentioning that in a mathematical proof, all premises are required to prove the proposition.

The coding scheme further comprised categories for responses that were interpreted as *non-criterial*, that is not allowing the determination of the underlying acceptance criteria. Here, categories were created for *suggestions for improvements* for general recommendations that could not be associated with a specific acceptance criterion, and *mathematical proof*, for claims that the purported proof is (not) a “mathematical proof” without further criteria as justification. The other non-criterial answers varied substantially (e.g., “if identical to the sample solution :D” or “It does not harm that tutors also participate in the lecture. Like this they stay up to date”)<sup>3</sup>, yet no additional coherent categories could be formed.

The answers from mathematicians' third section were coded analogously to the open items from the proof validation section. Further, they were additionally analyzed regarding comments on the impossibility of ascertaining the correctness of a proof, as such comments were noticed during the initial coding process.

Coding was carried out by a researcher and a student assistant. 20% of the data were double-coded with good inter-rater reliabilities ( $\kappa_{\text{segmentation}} \geq 0.70$ ;  $\kappa_{\text{match}} \geq 0.93$ ;  $\kappa_{\text{acc\_criteria}} \geq 0.85$ ).

## 5 Results

### 5.1 Proof validation and match of justifications (RQ 1)

For each group and each purported proof, descriptive results are given in Table 3, displaying (a) the percentage of participants providing a *correct judgment* regarding the acceptability of the proof, (b) the percentage of participants *providing a justification*, and (c) the percentage of participants providing a *matching justification* (relative to each group or the number of participants in each group who provided a justification). Overall, the three groups differ significantly with regard to the distribution of correct judgments for the four proofs [ $\chi^2(6) = 33.09$ ,  $p < .001$ ]. Post-hoc pairwise comparisons show significant differences between school students and mathematicians as well as between university students and mathematicians [ $\chi^2(3) > 15.10$ ,  $p < .002$ ] in favor of the mathematicians, yet school and university students do not differ significantly [ $\chi^2(3) = 3.23$ ,  $p = .358$ ].

For both student groups, solution rates differ significantly among the four purported proofs [ $\chi^2(3) > 35.51$ ,  $p < .001$ ]. Most school and university students identified the acceptable proof as acceptable (> 81.8%) and the proof with inductive reasoning as unacceptable (> 84.8%), whereas solution rates for the proof with an incorrect warrant (< 57.0%) and circular reasoning (< 42.1%) are considerably lower.

<sup>3</sup> All illustrative examples were translated from German to English.



**Table 3** Descriptive results for the purported proofs

	School students	University students	Mathematicians
Acceptable proof			
Correct judgment	82.5%	81.8%	97.8%
Justification given	50.9%	45.5%	100%
Proof with an incorrect warrant			
Correct judgment	57.0%	47.0%	88.3%
Justification given	70.2%	68.2%	86.1%
Matching justification	50.9% (72.5%)	30.3% (44.4%)	83.5% (97.0%)
Proof with circular reasoning			
Correct judgment	42.1%	24.2%	95.2%
Justification given	46.5%	28.8%	85.7%
Matching justification	19.3% (41.5%)	9.1% (31.6%)	76.6% (89.3%)
Proof with inductive reasoning			
Correct judgment	86.0%	84.8%	97.8%
Justification given	79.8%	75.8%	83.9%
Matching justification	64.0% (80.2%)	62.1% (82.0%)	74.0% (88.2%)

Numbers in brackets are relative to the number of participants in each group who provided a justification. All other numbers are relative to the respective group

Focusing on justifications, data show that the percentage of justifications given by school and university students varied significantly among the different purported proofs [28.8–79.8%;  $\chi^2(3) > 50.17$ ,  $p < .001$ ]. School and university students provided fewest justifications for the proof with circular reasoning (46.5% and 28.8% respectively) and most for the proof with inductive reasoning (79.8% and 75.8% respectively). For the latter, 80.2% and 82.0% respectively of the provided justifications matched the proof's error, whereas only 41.5% and 31.6% respectively of the justifications for the proof with circular reasoning could be seen as matching. Mathematicians frequently provided justifications (> 83.9%) and, in contrast to school and university students, gave fewest justifications for the inductive reasoning and most for the acceptable proof. The justifications given by mathematicians matched the purported proofs well (> 88.2%), although some provided only meta-comments on the task, notation, or other aspects that could not be classified as 'matching'.

## 5.2 Acceptance criteria referred to during proof validation (RQ 2)

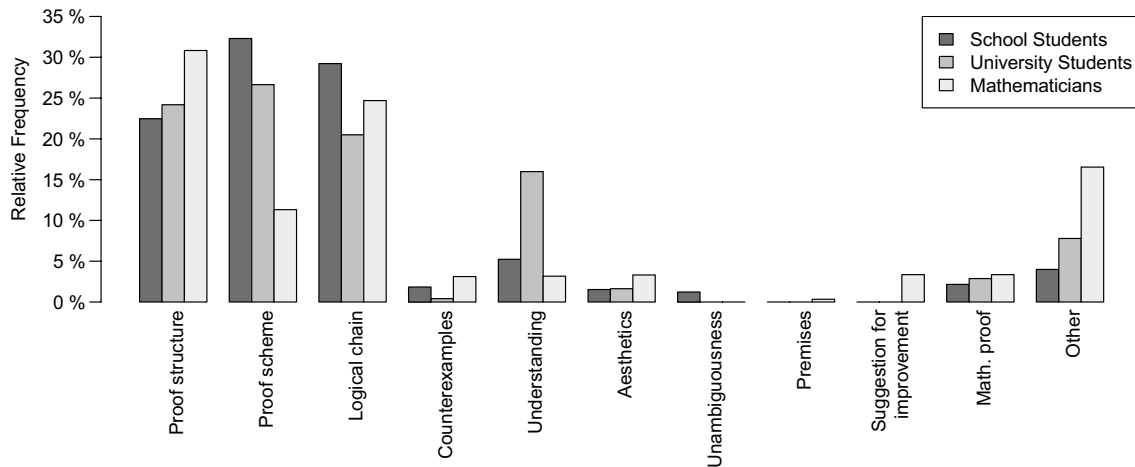
To capture the used acceptance criteria, each justification was coded separately. Overall, justifications contained more than 2500 references to acceptance criteria. Combining the justifications for all four purported proofs, school students on average mentioned 2.8, university students 3.5, and mathematicians 7.6 acceptance criteria. Yet, the number of acceptance criteria varied from proof to proof as the number of justifications also varied considerably.

School students, university students, and mathematicians referred to all deductively derived categories, except for *consistency with prior knowledge* (Fig. 3). Furthermore, *unambiguousness* (e.g., "I'm not sure if the proof is unambiguous enough") was used only by school students, whereas *all premises were used* (e.g., "Were all premises used? If not, the proof is either more general than required or wrong") and *suggestions for improvements* (e.g., "the student should do a proof by cases mod 3") were only used by mathematicians. All other categories were used by all three groups. In contrast to prior studies in research contexts, no group referred to social criteria.

Overall, groups significantly differed in the frequency distribution of the acceptance criteria they referred to [ $\chi^2(14) = 242.12$ ,  $p < .001$ ]<sup>4</sup>. Post-hoc pairwise tests also indicated significant pairwise differences [ $\chi^2(7) > 28.35$ ,  $p < .001$ ].

Out of the thirteen categories, *logical chain* (e.g., "the theorem he refers to is both wrong and unknown"), *proof scheme* (e.g., "the so-called proof consists only of examples"), and *proof structure* (e.g., "the assertion to be proved is assumed at the beginning") were used most often (Fig. 3). Mathematicians referred to proof structure more frequently and to proof scheme less frequently than school and university students.

<sup>4</sup> The categories *unambiguousness*, *all premises were used*, and *suggestions for improvements* were excluded for the test as cell frequencies were zero.



**Fig. 3** Acceptance criteria referred to by the three groups (based on the data for all tasks, relative to the number of criteria referred to within each group)

Interestingly, university students referred to *understanding* (16.0% of the coded criteria; e.g., “I can’t follow the proof”) significantly more often than both other groups [5.2% for school students, 3.2% for mathematicians;  $\chi^2(3)=68.06$ ,  $p < .001$ ]. Mathematicians’ high number of *non-criterial* codes is due to answers that included references to their educational practice or other aspects (e.g., “I would not use this proof in my teaching”), which could not be coded as a reference to an acceptance criterion.

Examining differences in the use of acceptance criteria between groups, five of the categories (*proof structure*, *proof scheme*, *aesthetics*, *math. proof*, and *other*) show a monotonic increase or decrease in the percentage of references to the corresponding acceptance criteria from school to university students and from university students to mathematicians. Three categories (*logical chain*, *counterexamples*, and *understanding*) show an increase–decrease or a decrease–increase pattern. Contrary to our hypothesis, this result indicates that there is no universally increasing alignment with mathematicians’ use of acceptance criteria, but more nuanced changes can be observed.

Overall, the deductively derived acceptance criteria *proof structure*, *proof scheme*, and *logical chain* also empirically represent main acceptance criteria during proof validation as they are the criteria most referred to by each group. Additionally, *understanding* appears to be a main acceptance criterion, at least for university students.

### 5.3 Mathematicians’ acceptance criteria in the general context of teaching (RQ 3)

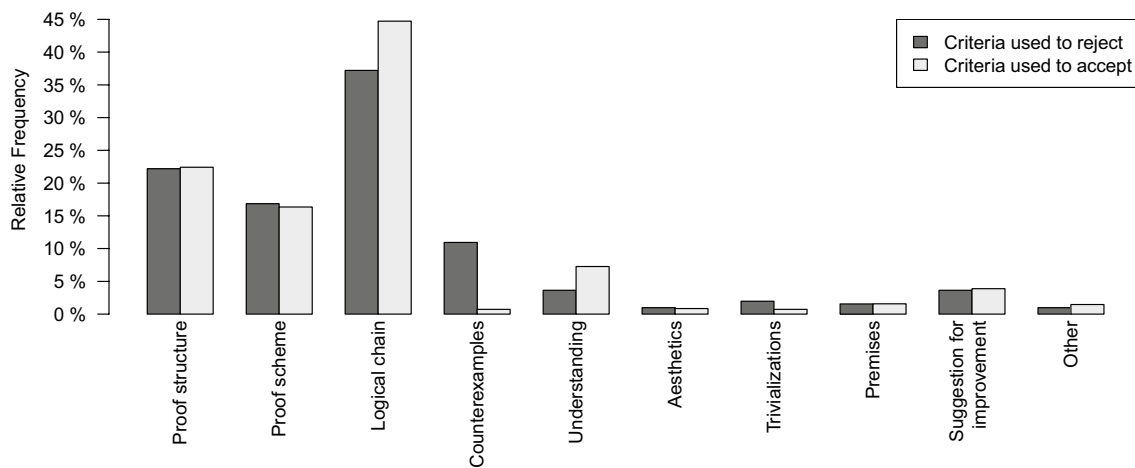
To complement the acceptance criteria derived from the validation of specific proofs, mathematicians were asked to

provide criteria that a tutor could use to accept or reject purported proofs. Two items were used, so that acceptance criteria mentioned for accepting and rejecting could be separated. Mathematicians’ answers were detailed and 1537 codes for acceptance criteria were assigned. On average, mathematicians referred to 2.6 criteria for accepting and 3.0 for rejecting a proof.

The criteria contained in the answers and their relative frequency (Fig. 4) show a pattern similar to the results from the proof validation tasks. Only the category *mathematical proof* was no longer found in mathematicians’ answers, whereas a new category *trivializations* had to be introduced, as the use of words such as “obviously” were mentioned as a criterion to reject proofs. Vice versa, not using such words was mentioned as a criterion to accept a proof, often labeled as a necessary, yet not sufficient criterion. Furthermore, criteria relating to the *proof structure* were mentioned less frequently in this section as compared to the validation tasks.

The results from this more general section (without a specific proposition and purported proof) mirror the result from the proof validation section, in which the acceptance criteria *logical chain*, *proof structure*, and *proof scheme* are referred to most.

The distinction between criteria for accepting and rejecting a proof further shows significant differences [ $\chi^2(9)=72.72$ ,  $p < .001$ ]. *Counterexamples* were used more often when rejecting a proof [ $\chi^2(1)=67.18$ ,  $p < .001$ ], whereas *understanding* was referred to more often when accepting a purported proof [ $\chi^2(1)=7.87$ ,  $p = .005$ ]. Thus, in the context of accepting a proof, there appears to be a shift towards meaning-oriented criteria, although structure-oriented criteria are still mentioned more frequently.



**Fig. 4** Mathematicians' acceptance criteria for and against the acceptance of a student proof (relative to the number of criteria referred to for/against the acceptance)

Finally, it should be mentioned that 8.3% of mathematicians' answers contained an explicit comment that asserting the correctness of a proof may be complicated or impossible.

## 6 Discussion

The presented study includes two major contributions for research on mathematical argumentation, proof, and evidence: First, it systematically compares acceptance criteria between different groups of persons. Although there is related prior research (e.g., Harel and Sowder 1998; Hemmi 2006; Weber 2003), it has so far been hard to disentangle task effects from population effects as different tasks were used for different populations. Second, it examines mathematicians' acceptance criteria in the context of teaching and not in the context of research, which has been the focus so far. The results show a considerable discrepancy between both contexts, which has yet to be addressed and explained.

### 6.1 Problems in validating proofs

The findings are consistent with previous results (e.g., Alcock and Weber 2005; Selden and Selden 2003), in which school and university students show mixed performances when validating proofs, and extends them by showing that success significantly depends on the purported proof and the error contained in it. In particular, university students' solution rate when validating a proof with circular reasoning was merely 24% and only about 10% gave a matching justification, whereas 68% of the justifications did not match. In contrast, school and university students were quite aware that inductive reasoning is not acceptable in proofs (> 80%) and many students (> 60%) provided a matching justification.

The results obtained show that university students, although having finished at least their first semester, do not differ significantly from school students in their solution rates. University students' performance is even slightly worse, both in providing correct judgments and matching justifications. Although this may partially be an effect of the selected sample, there appears to be a substantial number of university students who even after studying one semester do not show a robust understanding of criteria to judge the acceptability of proofs. Since understanding proof as a concept and handling proofs successfully are among the most important learning goals in early undergraduate mathematics (see Selden 2011), results appear to question the current *definition-theorem-proof-teaching* style that often only implicitly addresses the concept of proof and according norms and values (see also Lakatos 1963).

### 6.2 Acceptance criteria and enculturation

Despite school and university students' problems when validating proofs, the results obtained show that they refer to multiple acceptance criteria and use mostly the same ones as mathematicians. Thus, school and university students appear to have (at least some) knowledge about different acceptance criteria for mathematical proofs. Still, the combination of school and university students' judgments, justifications, and their acceptance criteria show that they have severe problems *employing* acceptance criteria effectively, as they often fail to give matching justifications. This extends the exploratory results by Selden and Selden (2003).

Examining the acceptance criteria referred to by all three groups (see Fig. 3), *proof structure*, *proof scheme*, *logical chain*, and *understanding* appear to be most important in

the context of teaching, showing an emphasis on structure-oriented criteria. However, there are significant differences: Whereas mathematicians focus on *proof structure* and *logical chain*, implying a local and global perspective on proof, school and university students refer more often to the *proof scheme*, focusing on a local perspective. This aspect is reflected in students' low solution rates when validating the proof with circular reasoning (see Table 3), which is an error relating to a global aspect of the proof. Results thus underpin qualitative results that students focus on surface features or individual inferences and that experts pay more attention to the overall structure (Inglis and Alcock 2012; Weber 2008; Weber and Mejía-Ramos 2011).

In contrast to mathematicians, school and university students seem to refer to *understanding* more often. This may reflect a focus on understanding and explanation as functions of proof within the school community. From an enculturation perspective, it is particularly interesting to see that university students used *understanding* more often than school students, as one might have expected the opposite as university students should be more familiar with validity as a function of proof. This result might be interpreted as failed enculturation, as university students do not align more with mathematicians' use of the acceptance criterion than school students do. However, this could be interpreted more favorably: School students transitioning from school to university enter a new mathematical community with new (at least for them) socio-mathematical norms and acceptance criteria. It may be normal that new university students are insecure in applying the new acceptance criteria properly (matching this study's and prior results), and thus may refer to meaning-oriented criteria such as understanding, as they perceive the proof's error, but cannot frame it using the new structure-oriented criteria.

Overall, the results show that *structure-oriented criteria* appear to be the most important acceptance criteria for all three groups in this context, with university students also using meaning-oriented criteria from time to time.

### 6.3 Mathematicians' double role: teaching and research

Prior research on acceptance criteria in *the context of research* (Hanna 1989; Heinze 2010; Weber 2008; Weber and Mejía-Ramos 2011) have shown that meaning-oriented criteria are used frequently by mathematicians, which could not be replicated in our study for *the context of teaching*. This can possibly be interpreted in the sense of weak and strong problem-solving strategies (e.g., Chinnappan and Lawson 1996): proofs in the context of teaching can be assumed to be relatively easy for mathematicians. They may thus be able to apply structure-oriented

criteria (as strong problem-solving strategies, which can be used to ensure absolute conviction) more easily than in the context of research, where proofs are generally more difficult. There, they may not always be able to use these strong criteria, but have to use weaker social or meaning-oriented criteria. For students, on the other hand, proofs in the context of teaching are mostly difficult, thus using more meaning-oriented criteria as reflected in the data.

Mathematicians' answers to the open questions regarding acceptance criteria for proof may provide further evidence for this interpretation. The meaning-oriented acceptance criterion *understanding* is used more often when showing the acceptability of a proof, a generally more difficult task than rejecting it, which is done rather with structure-oriented criteria (see Fig. 4).

Finally, the difference between research and teaching contexts is underlined by the fact that not a single social acceptance criterion was used, whereas these are mentioned as important in the context of research (Hanna 1989; Heinze 2010).

### 6.4 Limitations

The results from this study align with prior results in multiple ways and new results appear plausible. Still, they have to be handled with care. Firstly, the school student sample consists of future mathematics students and it could be argued that it is not representative for *all school students*. However, the emphasis of this paper was a cross-sectional comparison in a quasi-longitudinal way, which makes this choice more reasonable than a comparison with a *general school student sample*. Secondly, the student sample may be selective, as it was based on a voluntary course. Still, the size of the course equals a third of all students enrolled in the first semester, so that a large proportion of the students was included in this study.

Furthermore, two methodological limitations should be mentioned: In their online survey, mathematicians received proof validation tasks first and the general questions regarding acceptance criteria afterwards. As both parts were clearly separated and put in different contexts, severe priming effects were neither assumed nor observed. Still, the sequence should be counterbalanced in future research. Moreover, the study used one proposition with four purported proofs. Data from other propositions and proofs, including those from different content areas, would be beneficial to underpin the results. In particular, propositions and proofs of varying difficulty could be used to examine the dependency of the use of structure- and meaning-oriented acceptance criteria on the difficulty.

Finally, all three samples are from Germany. Although it seems likely that results generalize to other countries,

the generalizability of results in the context of proofs are always questionable, as they rely on local socio-mathematical norms. International replication studies would thus be advantageous.

## 6.5 Implications

The presented study gives essential insights into school students', university students', and mathematicians' use of acceptance criteria. Results show that school and university students have difficulties correctly validating proofs and, in particular, in giving matching justifications underpinned by appropriate acceptance criteria. Accordingly, school and university students appear to have problems implementing known acceptance criteria and thus may benefit from interventions focusing on the use of acceptance criteria. Here, future research is needed to analyze the effectiveness of such approaches and to find out more about school and university students' difficulties with implementing acceptance criteria.

Importantly, results confirm significant differences between the use of acceptance criteria in different communities. Not only could differences between all three groups be found, but results also highlight differences for mathematicians in teaching as contrasted with research. Apparently, there is not only a shift in the acceptance of proof from school to university, but a second shift from university teaching to research that has not yet received attention in the mathematics education community. In particular, mathematicians may intentionally focus on an idealized, structure-oriented picture of mathematics and according structure-oriented criteria in their teaching. However, it is unclear if this seeming detour has a positive effect on students' learning and enculturation, and if this emphasis is actually made deliberately. Educationally, the focus on structure-oriented criteria may be acceptable, but it appears unclear how students can learn to use acceptance criteria, if they are at least partially unauthentic, and how students can learn to use the criteria from mathematicians' research practice if they are not enculturated accordingly.

Based on the data and discussion, three main goals are seen for further research: to find effective ways to support university students in implementing the appropriate local acceptance criteria; to analyze the shift between acceptance criteria in teaching and research; and to use the data on acceptance criteria to inform ongoing discussions about the acceptance of proof.

Concluding, acceptance criteria for proofs are important both in understanding the concept of proof in mathematical practice as well as in supporting school and university students in handling proof, and thus also in their enculturation into local mathematical communities. Although there is little consensus about general acceptance criteria

for mathematical proofs so far, studying local criteria may prove helpful for mathematics, philosophy of mathematical practice, and mathematics education, and may help in gaining further insights into what is counted as evidence in mathematical arguments.

## References

- Aberdein, A. (2009). Mathematics and argumentation. *Foundations of Science*, 14(1), 1–8.
- Alama, J., & Kahle, R. (2013). Checking proofs. In A. Aberdein & I. J. Dove (Eds.), *The argument of mathematics* (pp. 147–170). Dordrecht: Springer.
- Alcock, L., & Weber, K. (2005). Proof validation in real analysis: Inferring and checking warrants. *The Journal of Mathematical Behavior*, 24(2), 125–134.
- Azzouni, J. (2004). The derivation-indicator view of mathematical practice. *Philosophia Mathematica*, 12(2), 81–106.
- Balacheff, N. (2008). The role of the researcher's epistemology in mathematics education: An essay on the case of proof. *ZDM - The International Journal on Mathematics Education*, 40(3), 501–512.
- Balacheff, N. (2010). *Bridging knowing and proving in mathematics: A didactical perspective*. Dordrecht: Springer.
- Bieda, K., Holden, C., & Knuth, E. (2006). Does proof prove? Students' emerging beliefs about generality and proof in middle school. In S. Alatorre, J. L. Cortina, M. Sáiz, & A. Méndez (Eds.), *Proceedings of the 28th annual meeting of the North American chapter of the international group for the psychology of mathematics education* (Vol. 2, pp. 395–402). Mérida: Universidad Pedagógica Nacional.
- Bieda, K., & Lepak, J. (2014). Are you convinced? Middle-grade students' evaluations of mathematical arguments. *School Science and Mathematics*, 114(4), 166–177.
- CCSSI. (2010). *Common core state standards for mathematics*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.
- Chinnappan, M., & Lawson, M. J. (1996). The effects of training in the use of executive strategies in geometry problem solving. *Learning and Instruction*, 6(1), 1–17.
- Clark, M., & Lovric, M. (2009). Understanding secondary-tertiary transition in mathematics. *International Journal of Mathematical Education in Science and Technology*, 40(6), 755–776.
- Corriveau, C., & Bednarz, N. (2017). The secondary-tertiary transition viewed as a change in mathematical cultures: An exploration concerning symbolism and its use. *Educational Studies in Mathematics*, 95(1), 1–19.
- Dawkins, P. C., & Weber, K. (2017). Values and norms of proof for mathematicians and students. *Educational Studies in Mathematics*, 95(2), 123–142.
- de Villiers, M. (1990). The role and function of proof in mathematics. *Pythagoras*, 24, 17–24.
- Devlin, K. (2003). When is a proof? Retrieved from [https://www.maa.org/external\\_archive/devlin/devlin\\_06\\_03.html](https://www.maa.org/external_archive/devlin/devlin_06_03.html). Accessed 2 Mar 2018.
- Hanna, G. (1989). More than formal proof. *For the Learning of Mathematics*, 9(1), 20–23.
- Hanna, G. (2000). Proof, explanation and exploration: An overview. *Educational Studies in Mathematics*, 44(1–2), 5–23.
- Hanna, G., & Jahnke, H. N. (1996). Proof and proving. In A. Bishop, M. A. K. Clements, C. Keitel-Kreidt, J. Kilpatrick & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 877–908). Dordrecht: Springer.



- Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. In A. H. Schoenfeld, J. Kaput, E. Dubinsky & T. Dick (Eds.), *Research in collegiate mathematics education. III* (pp. 234–283). Providence: AMS.
- Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education*, 21(4), 396–428.
- Heinze, A. (2010). Mathematicians' individual criteria for accepting theorems and proofs: An empirical approach. In G. Hanna, H. N. Jahnke & H. Pulte (Eds.), *Explanation and proof in mathematics: Philosophical and educational perspectives* (pp. 101–111). Boston: Springer.
- Heinze, A., & Reiss, K. (2003). Reasoning and proof: Methodological knowledge as a component of proof competence. In M. A. Mariotti (Ed.), *Proceedings of the third conference of the European society for research in mathematics education* (Vol. 4, pp. 1–10). Bellaria, Italy.
- Heinze, A., & Reiss, K. (2009). Developing argumentation and proof competencies in the mathematics classroom. In D. A. Stylianou, M. L. Blanton & E. J. Knuth (Eds.), *Teaching and learning of proof across the grades: A K-16 perspective* (pp. 191–203). New York: Routledge.
- Hemmi, K. (2006). *Approaching proof in a community of mathematical practice*. Stockholm: Department of Mathematics, Stockholm University.
- Hilbert, D. (1931). Die Grundlegung der elementaren Zahlenlehre. *Mathematische Annalen*, 104(1), 485–494.
- Inglis, M., & Alcock, L. (2012). Expert and novice approaches to reading mathematical proofs. *Journal for Research in Mathematics Education*, 43(4), 358–390.
- Inglis, M., Mejía-Ramos, J. P., Weber, K., & Alcock, L. (2013). On mathematicians' different standards when evaluating elementary proofs. *Topics in Cognitive Science*, 5(2), 270–282.
- Lakatos, I. (1963). Proofs and refutations (I). *The British Journal for the Philosophy of Science*, 14(53), 1–25.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.
- Manin, Y. (2010). *A course in mathematical logic for mathematicians*. New York: Springer.
- Mariotti, M. A. (2006). Proof and proving in mathematics education. In Á Gutiérrez & P. Boero (Eds.), *Handbook of research on the psychology of mathematics education: Past, present and future* (pp. 173–204). Rotterdam: Sense Publishers.
- Mariotti, M. A., & Balacheff, N. (2008). Introduction to the special issue on didactical and epistemological perspectives on mathematical proof. *ZDM - The International Journal on Mathematics Education*, 40(3), 341–344.
- Mayring, P. (2014). Qualitative content analysis: Theoretical foundation, basic procedures and software solution. Klagenfurt. <http://nbn-resolving.de/urn:nbn:de:0168-ssaoar-395173>. Accessed 13 Apr 2018.
- Moore, R. C. (1994). Making the transition to formal proof. *Educational Studies in Mathematics*, 27(3), 249–266.
- Perrenet, J., & Taconis, R. (2009). Mathematical enculturation from the students' perspective: Shifts in problem-solving beliefs and behaviour during the bachelor programme. *Educational Studies in Mathematics*, 71(2), 181–198.
- Reid, D. A., & Knipping, C. (2010). *Proof in mathematics education. Research, learning and teaching*. Rotterdam: Sense Publishers.
- Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. In D. Grouws (Ed.), *The handbook for research on mathematics teaching and learning* (pp. 334–370). New York: MacMillan.
- Selden, A. (2011). Transitions and proof and proving at tertiary level. In G. Hanna & M. de Villiers (Eds.), *Proof and proving in mathematics education* (pp. 391–420). Dordrecht: Springer.
- Selden, A., & Selden, J. (2003). Validations of proofs considered as texts: Can undergraduates tell whether an argument proves a theorem? *Journal for Research in Mathematics Education*, 34(1), 4–36.
- Tall, D. (2008). The transition to formal thinking in mathematics. *Mathematics Education Research Journal*, 20(2), 5–24.
- Tanswell, F. (2015). A problem with the dependence of informal proofs on formal proofs. *Philosophia Mathematica*, 23(3), 295–310.
- Ufer, S., Heinze, A., Kuntze, S., & Rudolph-Albert, F. (2009). Beweisen und Begründen im Mathematikunterricht: Die Rolle von Methodenwissen für das Beweisen in der Geometrie. *Journal für Mathematik-Didaktik*, 30(1), 30–54.
- Weber, K. (2003). Students' difficulties with proof. In A. Selden & J. Selden (Eds.), *In The Mathematical Association of America online: Research sampler* (Vol. 8). <https://www.maa.org/programs/faculty-anddepartments/curriculum-department-guidelines-recommendations/teaching-and-learning/research-sampler-8-students-difficulties-with-proof>.
- 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., & Mejía-Ramos, J. P. (2011). Why and how mathematicians read proofs: An exploratory study. *Educational Studies in Mathematics*, 76(3), 329–344.
- Weber, K., & Mejía-Ramos, J. P. (2015). On relative and absolute conviction in mathematics. *For the Learning of Mathematics*, 35(2), 15–21.
- Weber, K., Mejía-Ramos, J. P., Inglis, M., & Alcock, L. (2013). On mathematicians' proof skimming: A reply to Inglis and Alcock / Skimming: A response to Weber and Mejía-Ramos. *Journal for Research in Mathematics Education*, 44(2), 464–475.
- Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education*, 27(4), 458–477.

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