MATHEMATICIANS' CRITERIA FOR ACCEPTING THEOREMS AND PROOFS – AN INTERNATIONAL STUDY

Anika Dreher[1] and Aiso Heinze[2]

[1] Freiburg University of Education, Germany
[2] IPN – Leibniz-Institute for Science and Mathematics Education, Kiel, Germany

Argumentation and proof are crucial for the mathematics discipline and should thus permeate mathematics education. In particular proof validation lately became a focus of attention in mathematics education research. Since expert practice is seen as an important frame of reference for instructional goals regarding proof validation, it was emphasized that empirical research on mathematicians' criteria for accepting proofs is needed. However, such empirical research based on reliable quantitative data is still scarce. Consequently, this study analyzes criteria for accepting proofs in their daily work held by N = 243 highly respected mathematicians from all over the world. The results indicate three types of mathematicians who rely on certain criteria to various degrees as well as differences between status groups.

INTRODUCTION

As mathematics is a proving science, proof is widely seen as essential especially in secondary and tertiary mathematics education (e.g., Hanna, 1997; Marriotti, 2006). Consequently, there is a broad base of educational research on how students engage with mathematical proofs. Although this research focuses mainly on the construction of proofs (Sommerhoff, Ufer, & Kollar, 2015), there is a growing interest in practices of proof validation (e.g., Selden & Selden, 2003; Weber, 2008). Several scholars in mathematics education emphasized that corresponding goals for instruction should be informed by expert practice of mathematicians (e.g., Inglis & Alcock, 2012; Weber, Inglis, & Ramos, 2014). Since this requires a thorough understanding of such expert practice, research focusing on mathematicians' professional practices regarding proofs was called for (Weber et al., 2014). Against this background there were in particular some studies conducted exploring how mathematicians validate proofs (e.g., Weber, 2008) as well as their criteria for accepting mathematical theorems as being valid (Heinze, 2010; Mejía-Ramos, & Weber, 2014). Based on corresponding findings Weber and colleagues (2014) argued that current instructional recommendations in mathematics education are "oversimplified and not based on an accurate understanding of mathematical practice" (p. 54). However, as these studies were so far mostly explorative and based on relatively small sample sizes, further research with better-developed instruments and better-quality data is necessary to get more insight into expert's criteria for accepting mathematical theorems and proofs as being valid (Heinze, 2010). Consequently, this online survey study explores mathematicians' ac-

2018. In E. Bergqvist, M. Österholm, C. Granberg, & L. Sumpter (Eds.). *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 363-370). Umeå, Sweden: PME.

ceptance criteria in their daily work based on a high-class international sample by means of quantitative methods.

THEORETICAL BACKGROUND

The naïve idea that the deductive nature of mathematics allows determining the correctness of mathematical proofs with absolute certainty turned out to be wrong (e.g., Hanna, 1983). Consequently, the question as to how new mathematics results get accepted by mathematicians gives rise to a complex field of research which is interesting not only from a mathematics education point of view, but also from the perspective of philosophy of mathematics (e.g., Geist, Loewe, & van Kerkhove, 2010). Reflecting upon this question, among others, Heinze (2010) looked at the situation in science, where a criterion for accepting a new result based on experiments is that this result must be replicated independently under the same conditions by different researchers. Considering proofs as thought experiments, he asked whether mathematicians rely on experiments and replications by others or whether they have to replicate it themselves in order to accept a corresponding result. The question raised, namely to what degree knowledge by testimony (Geist et al., 2010) or authoritarian evidence (Weber et al., 2014) can lead mathematicians to accept a new theorem in the sense of relying on journals or other mathematicians is central to the discussion of mathematicians' acceptance criteria. There are mathematicians for whom it is crucial that they check a proof of every mathematical result which they apply in their work (e.g., Geist et al., 2010). However, often this is hardly possible, since the proofs of some theorems are extremely long and the diversity of the mathematics discipline involves that many mathematicians are unable to follow the proofs of theorems that come from another area of research (Auslander, 2008). Hence, there is a consensus that social processes play a role for the acceptance of new mathematical results (Hanna, 1983). Moreover, findings from first empirical studies in this area indicate that some mathematicians rely on authoritarian evidence to accept theorems and proofs as black boxes in their own research (Heinze, 2010; Weber et al., 2014). In their online survey with 118 American mathematicians, Mejía-Ramos and Weber (2014) found for instance that 72% of the participants agreed with the statement "It is not uncommon for me to believe that a proof is correct because it is published in an academic journal" (p. 166). In view of such results Weber and colleagues (2014) argued that authors in mathematics education "who believe that students should not accept claims as true because an authority told them that this was the case and that one should not consider who wrote the argument while evaluating its validity" (p. 45) should rethink the grounds for their instructional suggestions. The findings of Heinze's (2010) exploratory online survey with 40 German mathematicians also indicate that there is a substantial amount of reliance on the mathematics community and peer-reviewed journals with respect to the acceptance of theorems and proofs. He found, for instance, that on average the participants stated to frequently accept a theorem in their daily work as being valid, if a proof was published long ago and there was no contradiction so far. However, the results also suggested that full professors relied less frequently on the mathematics community and peer-reviewed journals than PhD students and postdocs. This could be a result of the professors' experience, in particular with reviewing papers for journals, since the mathematical refereeing process is sometimes not as trustworthy as one may think (Geist et al., 2014; Nathanson, 2008). Thus, the participants' high reliance on authoritarian evidence in the study by Mejía-Ramos and Weber (2014) might be partly due to the small share of full professors in their sample (28% faculty members) and the fact that more than half of the participants had never refereed a paper for a journal.

If mathematicians lack time and other resources to check a proof step by step of every theorem they use in their daily work, but still do not rely on authoritarian evidence, another solution could be to gain conviction by "partly" checking a proof. Indeed, acceptance criteria for some mathematicians can be that they checked the key arguments of a proof, are convinced that the main ideas of a given proof are correct (Heinze, 2010), or checked the theorem for carefully chosen examples (Weber, 2008). In particular in view of the latter criterion, Weber and colleagues (2014) "challenge[d] instruction that aims for students to never seek conviction in this way".

It can be summarized that mathematicians' criteria for accepting theorems and proofs in their own daily work may be individual checking – where a distinction can be made between "step by step" and "partly" checking – but also authoritarian evidence which may refer to respected journals or the assumption that "enough" mathematicians in the community have verified a proof. However, there is still little evidence for answering the question to what extent mathematicians rely on these different kinds of acceptance criteria in their daily work. Empirical findings indicate that there is substantial heterogeneity among mathematicians regarding their acceptance criteria (e.g., Heinze, 2010; Mejía-Ramos & Weber, 2014). This heterogeneity could be a result of external factors (e.g., different status groups, areas of mathematics, culture) or of more individual characteristics in the sense of different types of mathematicians. This is not clear, yet. Thus, there is a need for research into questions whether there are differences between the acceptance criteria of certain groups of mathematicians, what types of mathematicians exist and how dominant these types are in the mathematics community. The quantitative studies on mathematicians' criteria for accepting theorems and proofs that exist so far focus mainly on the level of PhD students and postdocs instead of full professors who are experienced and esteemed members of the community and on national samples. In view of the assumed heterogeneity, it might however be crucial to focus on an adequate sample in order to get the full picture. Moreover, since so far this research was based on analysis regarding single items, a study using more developed instruments is needed.

RESEARCH QUESTIONS

According to the need for research pointed out in the previous sections the study presented here aims to provide evidence for the following research questions:

1. *To what extent do mathematicians rely on different criteria for accepting a theorem and proof as valid in their daily work?*

2. *Are differences with respect to external factors (status groups, refereeing experience) associated with differences regarding acceptance criteria?*
3. *How can different types of mathematicians regarding their acceptance criteria be characterized?*

SAMPLE AND METHODS

For answering these research questions an online survey was designed using the software “Unipark”. The questionnaire was completed by a sample of 243 research mathematicians (177 male, 30 female, 36 without data) who have been participants of workshops at the highly reputable Mathematical Research Institute of Oberwolfach during the last years and are thus esteemed members of the international mathematics research community. The sample is international and not even restricted to European countries (151 from Europe, 39 from North America, 11 from Asia, 3 from Australia, 3 from South America, 36 without data). Most of the participants are full professors (114 full professors, 30 associate professors, 28 assistant professors, 27 postdocs, 1 PhD student, 2 professors emeritus, 3 senior lecturers (UK), 38 without data). The large majority has been referee for a journal paper multiple times (at least three times: 191, once or twice: 11, not yet: 7, without data: 34).

Corresponding to the research questions for this study, the participants were asked under which conditions they accept a mathematical theorem as valid in their daily mathematical work. The mathematicians could express their approval or disagreement regarding statements of the form “In my mathematical work I assume that a mathematical theorem is valid, if...” on a six-point Likert scale with endpoints “entirely disagree” and “entirely agree”. The statements were completed by criteria as identified in the previous section (for details and sample items see Table 1).

RESULTS

We started the data analysis by conducting a factor analysis with oblique rotation. After excluding one item that could not be assigned to any factor, the Kaiser criterion yielded 4 factors, where each item loads with > 0.4 on one factor (51% explained variance). The clustering of items suggests that the four factors represent the four kinds of acceptance criteria identified in the theoretical background. Hence, four scales could be formed (see Table 1). For three of the scales the reliability is good and in view of only two items forming the remaining scale, its reliability is acceptable.

The means and standard errors for these scales displayed in Figure 1 show that on average the mathematicians reported high agreement with the acceptance criterion individual checking “step by step” and medium agreement with the other three kinds of criteria. Regarding these other three kinds of criteria on average authoritarian evidence in the sense of “enough” other mathematicians have checked the proof received most and authoritarian evidence in the sense of peer-reviewed journals received least agreement.

Scale	Sample items	# items	Cronbach's α
Individual checking "step by step"	… I verified a given proof step by step.	2	.63
Individual checking "part-ly"	... I am convinced that the main ideas of a given proof are correct. ... the theorem is valid for all examples that I know.	7	.82
authoritarian evidence "jour-nals"	... the theorem was published with a proof in a refereed journal.	5	.88
authoritarian evidence "enough" mathematicians	… I know that a proof has been available for a long time and has been checked by many mathematicians.	6	.86

Table 1: Scales regarding different acceptance criteria

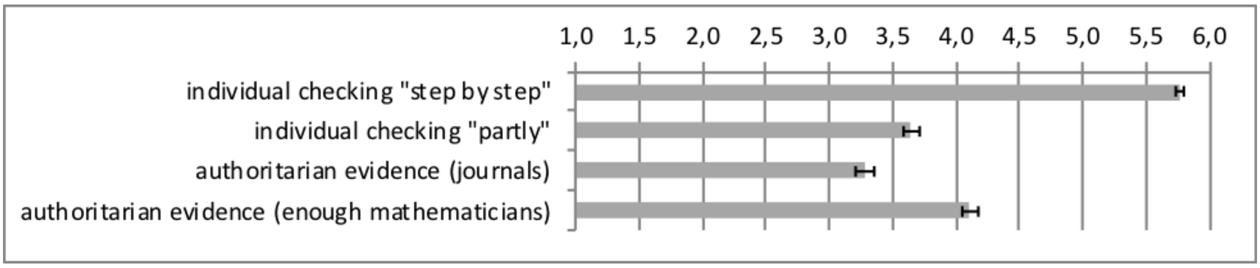

Figure 1: Agreement with acceptance criteria (means and their standard errors), six-point Likert scale from 1 = entirely disagree to 6 = entirely agree

In view of the second research question, we explored next whether mathematicians who differ regarding certain external factors also differ regarding their acceptance criteria. Due to space limitations, only selected results can be presented. For getting insight into whether the status group is a factor that accounts for heterogeneity among mathematicians regarding acceptance criteria, we consider the lowest and the highest status group in our sample with more than one representative: postdocs and full professor. Comparing these two status groups yields no significant differences regarding the two criteria referring to individual checking, but reveals that full professors agreed significantly less with authoritarian evidence "journals" ($t(138) = 2.96$, $p < .01$, $d = 0.73$) and with authoritarian evidence "enough mathematicians" ($t(53) = 2.72$, $p < .01$, $d = 0.52$) than postdocs. Both differences represent medium effect sizes. To investigate whether experience in refereeing for journals is associated with less reliance on peer-reviewed journals, those mathematicians with no refereeing experience where compared to those who had been refereeing for at least three times. Indeed, the former agreed significantly and substantially more with the criterion authoritarian evidence "journals" than the latter ($t(195) = 2.68$, $p < .01$, $d = 1.03$).

For answering the third research question and exploring different answering patterns a hierarchical cluster analysis using Ward's method was conducted. The cluster analysis was based on the four scales representing different kinds of acceptance criteria. This analysis yielded three clusters showing distinct answering patterns which are presented in Figure 2.

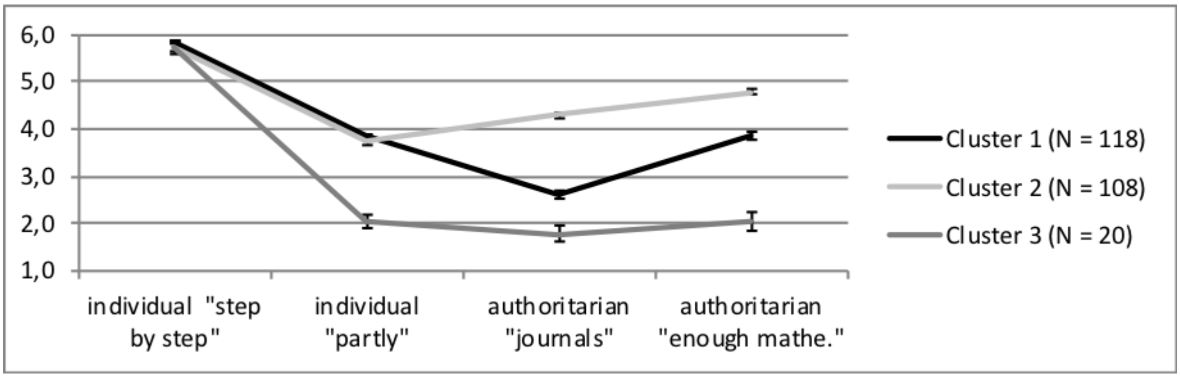

Figure 2: Means and their standard errors of the three clusters, six-point Likert scale from 1 = entirely disagree to 6 = entirely agree

The three clusters do not differ with respect to the criterion of individual checking "step by step", but there are substantial differences regarding the other three kinds of criteria. Cluster 3 consists of a minority of mathematicians who disagreed with any other acceptance criterion. Cluster 1 and 2 are almost of the same size and are characterized by the same level of medium agreement with the criterion individual checking "partly". However, the mathematicians of cluster 1 stated to rely less on journals than on their own partly checking, whereas this is vice versa for cluster 2. Moreover, the mathematicians of cluster 1 showed less conviction by authoritarian evidence "enough mathematicians" than their colleagues of cluster 2. Consistent with the results regarding the second research question, cluster 1 and 3 consist of relatively more full professors and cluster 2 consists of relatively more postdocs compared to the proportions in the full sample.

DISCUSSION AND CONCLUSIONS

Based on reliable scales and an international sample of highly esteemed members of the mathematics research community, this study provides further evidence that most mathematicians do not exclusively rely on individual step by step verification for accepting theorems and proofs as valid in their daily work. They may also use "partly" checking and authoritarian evidence referring to respected journals or the assumption that "enough" mathematicians have verified a proof as acceptance criteria. Concerning the heterogeneity within the mathematics research community with respect to the extent to which mathematicians rely on these different kinds of criteria, the findings of this study can give new insights: On the one hand, according to the second research question, it was explored whether external factors account for such heterogeneity. Indeed, full professors showed significantly less acceptance of both kinds of authoritarian evidence compared to postdocs. This could be a result of the full professors' experience in the mathematics research community. The assumption that in particular experience with the mathematical refereeing process could lead to less reliance on peer-reviewed journals is supported by the result that experience in refereeing was associated with significantly and substantially less agreement with the acceptance criterion referring to peer-reviewed journals. These results challenge Mejía-Ramos and Weber (2014), who concluded from their findings that there were no differences between the acceptance criteria of less experienced and more experienced mathematicians. However, their sample included only a small share of full professors and more than half of the participants had never refereed a paper for a journal. Consequently, the mathematicians' high reliance on authoritarian evidence reported from this study should be interpreted with care.

On the other hand, according to the third research question, different types of mathematicians regarding their acceptance criteria were explored by means of a cluster analysis. In line with Geist and colleagues (2010) the results indicate that there is a type of mathematician that clearly disagrees with every acceptance criterion except for individual step by step checking of a proof. However, this type appears to account for a minority in the mathematics research community. According to our findings the community is dominated by two types that do not strictly reject other acceptance criteria: one of them is characterized by relying more on individual "partly" checking than on peer-reviewed journals and the other one is characterized by relying more on authoritarian evidence than on "partly" checking. The fact that the three types represented the status groups by different proportions suggests that further analyses and research is necessary to investigate to what extent these types are based on external factors or more individual characteristics of mathematicians.

We would like to recall that the findings of this study should be interpreted with care, since the data is based on self-reports and thus social desirability may play a role. Hence, these results should be corroborated by means of other methods. Comments by some participants indicate that they often use a combination of acceptance criteria (see also Hanna, 1983). Thus, further research taking into account such combinations is necessary. Follow-up research should moreover focus on further external factors and also on the context of refereeing a paper complementing the context of mathematicians' daily work as considered in this study (e.g., Mejía-Ramos & Weber, 2014).

Acknowledgements

We would like to thank Benedikt Loewe for his support in initiating this study and in the item development. Moreover, we are thankful to Juan Pablo Mejía-Ramos for his helpful comments regarding the survey.

References

Auslander, J. (2008). On the roles of proof in mathematics. In B. Gold & R. A. Simons (Eds.), *Proofs and other dilemmas: Mathematics and philosophy* (pp. 61–77). Washington, DC: Mathematical Association of America.

Inglis, M., & Alcock, L. (2012). Expert and novice approaches to reading mathematical proofs. *Journal for Research in Mathematics Education, 43*, 358–390.

Inglis, M., & Mejia-Ramos, J. P. (2009). The effect of authority on the persuasiveness of mathematical arguments. *Cognition and Instruction, 27*, 25–50.

Geist, C. Löwe, B., & Van Kerkhove, B. (2010). Peer review and testimony in mathematics. In B. Löwe & T. Müller (Eds.), *Philosophy of Mathematics: Sociological Aspects and Mathematical Practice* (pp. 155–178). London: College Publications.

Hanna, G. (1983). *Rigorous proof in mathematics education*, OISE Press, Toronto.

Hanna, G. (1997). The ongoing value of proof in mathematics education. *Journal für Mathematik Didaktik, 97*(2/3), 171–185.

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). New York: Springer.

Marriotti, M. (2006). Proof and proving in mathematics education. In A. Guttierez & P. Boero (Eds.) *Handbook of research in mathematics education: Past, present, and future* [PME 1976-2006]. (pp. 173–204). Rotterdam: Sense Publishers.

Mejía-Ramos, J. P., & Weber, K. (2014). How and why mathematicians read proofs: further evidence from a survey study. *Educational Studies in Mathematics, 85*, 161–173.

Nathanson, M. B. (2008). Desperately seeking mathematical truth. *Notices of the American Mathematical Society, 55*(7):773.

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.

Sommerhoff, D., Ufer, S., & Kollar, I. (2015). Research on mathematical argumentation: A descriptive review of PME proceedings. In K. Beswick, T. Muir, & J. Wells (Eds.), *Proceedings of the 39th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 193–200). Hobart, Australia: PME.

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

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

Weber, K., Inglis, M., & Mejia-Ramos, J.P. (2014). How mathematicians obtain conviction: implications for mathematics instruction and research on epistemic cognition. *Educational Psychologist, 49*(1), 36–58.