ENGAGING MATHEMATICAL REASONING-AND-PROVING: A TASK, A METHOD, AND A TAXONOMY

Unni Wathne and Cornelia Brodahl
University of Agder, Norway

Abstract: *This article is the second paper in a series of papers on studies focusing on teaching mathematical reasoning-and-proving in elementary mathematics classroom. Participants are in-service teachers enrolled in a continuing university education program in mathematics. Results from the first paper suggested the method of imaginary dialogues to have the potential to support in-service teachers in engaging their students in mathematical reasoning-and-proving, and Balacheff's taxonomy of proofs to support in-service teachers in identifying students' argumentation. This study is on the following years' in-service teachers in the program. It examines their perceptions of the usefulness of two constituent parts of this approach, and what insights students' written dialogues might provide. The study draws on G. J. Stylianides' analytic framework for reasoning-and-proving. Main data were obtained from a questionnaire taken by 32 in-service teachers and follow-up interviews with four of them. The study reveals engaging students to reason, argue, and prove, while making students' argumentation visible for teachers was perceived the most useful with imaginary dialogues. The teachers' increasing awareness of levels of argumentation, was perceived to be the most useful with getting exposed to Balacheff's distinctions.*

Keywords: Balacheff's four levels of proofs, mathematical reasoning-and-proving, written imaginary dialogues

Introduction

Students need exposure to reasoning-and-proving activities, and teachers need tasks with which they can engage their students. While teacher educators may recognize proof as a fundamental activity in mathematical practice and the importance of reasoning-and-proving in elementary school, there is a shortage of research and resources they can draw on in preparing pre- or in-service teachers in engaging their students in proving activities in primary and lower secondary classroom (Stylianides, 2016).

This study is a continuation of a series of ongoing studies seeking to contribute to the limited research and need of resources for teacher educators' instructional support. Focus is on a combination of a mathematical task, a method to approach students' mathematical thinking processes, and a taxonomy for analysis. While the previous part of the study (Brodahl & Wathne, 2018) explored perceptions of first experiences with the complex combination as a whole, the current effort is a case study with new teachers under education that narrows the research perspective to the usefulness of the particular method and the particular taxonomy, as perceived by the teachers. It also deals with insights teachers gained in their students' process of reasoning-and-proving, when applying method and taxonomy. The theoretical framework for analysis for the current study, with its research focus on the need of resources for teacher educators' instructional support for teachers' reasoning-and-proving in the classroom, is from Stylianides (2016). Supporting aspects to this framework, Balacheff's (1988) taxonomy of four levels of proofs, constitute the conceptual and theoretical frame provided to the in-service teachers.

Mathematics teacher educators may be well acquainted with the shaking-hand-problem, as it might involve students in shifting from arithmetic reasoning to algebraic reasoning-and-proving when making a conjecture and justifying it. It deals with finding the number of handshakes needed if a group of people shook hands with each other. In order to give teachers more detailed insight into how students in class develop, the teacher-educator authors started to explore the potential of letting students write dialogues, and in-service teachers use a taxonomy to identify their students' levels of reasoning-and-proving. They assigned in-service teachers to engage their class, in pairs, to continue writing a given dialogue between two imaginary students having started discussing the shaking-hand-problem.

In this paper, we present in-service teachers' experiences with implementing so-called "imaginary dialogues" in their classroom and with their analysis of their students' written reasoning-and-proving. Writing in the form of dialogues was inspired by the method of imaginary dialogues used by Wille (2017), where a single student composes a written dialogue between two protagonists who discuss a mathematical task or question. Wille found the method to initiate reflection processes and argumentation. However, working collaboratively, not individually, makes it a modification of the method. In-service teachers identifying any reasoning-and-proving in their students' dialogue was based on Balacheff's four proof levels (1988).

Research Questions

This paper examined in-service teachers' perceptions on working with the mission of provoking and analyzing their students' reasoning-and-proving. Our research questions were

1. How do in-service teachers perceive the usefulness of introducing imaginary dialogues as a means to engage students in reasoning-and-proving in their classroom?
2. What types of insight in students' processes of reasoning-and-proving do in-service teachers perceive gaining in students' written imaginary dialogues?
3. How do in-service teachers perceive the usefulness of Balacheff's taxonomy in their process of identifying students' levels of reasoning-and-proving in students' written imaginary dialogues?

Theoretical Framework

There are several approaches to what is meant by the terms conjecture, argument, and proof, and the processes of explanation, justification, and proof-related-reasoning, in different research communities within mathematics education (see Reid, 2005, for a review; Stylianides, 2016). Stylianides (2007) offered a definition to proof in the context of a classroom community that includes three criteria:

> *Proof* is a *mathematical argument*, a connected sequence of assertions for and against a mathematical claim, with the following characteristics: (a) it uses statements accepted by the classroom community (*set of accepted statements*) that are true and available without further justifications; (b) it employs forms of reasoning (*modes of argumentation*) that are valid and known to, or within the conceptual reach of, the classroom community; and (c) it is communicated with forms of expression (*modes of argument representation*) that are appropriate and known to, or within the conceptual reach of, the classroom community. (p. 291, emphasis in original)

We adopted this definition for its balance between two considerations, mathematics as a discipline, focusing on what is accepted by mathematicians as proofs, and students as mathematical learners, focusing on what is within the conceptual reach of the classroom community. It underpins the notion of a sufficient argument in class (Brodahl & Wathne, 2018). For this study, we followed Stylianides' (2008) notion of reasoning-and-proving as describing the overall process of "making sense of and establishing mathematics knowledge" (p. 9) and used the analytic framework for reasoning-and-proving he presented for studying such processes (p. 10).

The mathematics subject curriculum for primary and secondary education (1–13) in Norway (Ministry of Education and Research, 2013), where this study has been conducted, expects students to engage in reasoning-and-proving in all four main activities in accordance with Stylianides' (2008) analytic framework of activities. They started with making mathematical generalizations (identify a pattern then make a conjecture) and end with providing support to mathematical claims (proof-arguments or non-proof arguments). This framework consists of mathematical, psychological, and pedagogical components. The mathematical component distinguishes four constituent main activities, two of them under the notion of mathematical generalization: (a) identifying a pattern and (b) making a conjecture, and two of them under the notion of providing support to mathematical claims: (c) providing a proof and (d) providing a non-proof argument. The framework also offers a further breakdown of these main activities that together comprise reasoning-and-proving to seven subcategories. Five of them are most central for the research focus of this study, as is the support of teachers in engaging their students in making mathematical generalizations and providing a proof: plausible pattern and definite pattern for main activity (a); conjecture for main activity (b); generic example and demonstration for main activity (c).

Balacheff (1988) suggested four levels of proof that differ in the degree of generality required and conceptualization involved, as described in our previous study (see Brodahl & Wathne, 2018, p. 32–33 for more detailed review):

1. Naive empiricism: The learner concludes based on only a small number of cases that are practically convenient to check.
2. Crucial experiment: The learner tests the conjecture with an example well outside the range so far considered, to explore the extent of its validity.
3. Generic examples: The learner concludes on a prototypical case, where an object is chosen not on its own, but as a characteristic representative of its class.
4. Thought experiment: Detached from any examples, the learner arrives at structured logical formulations and formalized symbolic expressions.

Balacheff (1988) identified the first three proof levels as pragmatic, being dependent on actions or visual representations. The third level, though, constitutes a transition from the specific to the general and from pragmatic justification to conceptual. The fourth level Balacheff distinguished as theoretical proof. These four levels constitute a taxonomy of proofs he used to classify proving tasks in school mathematics.

Stylianides (2008) acknowledged Balacheff's terms naive empiricism and crucial experiment as special kinds of empirical arguments for or against a mathematical claim, not qualifying as general evidence. Stylianides' framework separates providing a proof into two categories, generic examples and demonstrations (p. 10). Stylianides suggested a generic example to be a proof that uses a particular case seen as representative of the general case, in accordance to Balacheff (1988), while a demonstration to be a proof that uses formally established modes of mathematical proof, as is similar to Balacheff's thought experiment.

Stylianides' (2016) review of mathematics education research literature justified the importance of reasoning-and-proving as early as elementary school. From both philosophical and pedagogical standpoint, it can be argued that reasoning-and-proving deserves a central place throughout the school mathematics curriculum and is necessary for deep learning in mathematics. Nevertheless, in the body of research literature, numerous factors are found to have contributed to a rather marginal place of reasoning-and-proving in the elementary mathematics classroom. Stylianides (2016) singled out four factors for attention:

1. *Teachers' knowledge*: the weak knowledge that many elementary teachers have about proof
2. *Teacher's belief*: their presumed beliefs that proving is an advanced mathematical topic beyond the reach of elementary students
3. *Pedagogical demands*: the high pedagogical demands placed on elementary teachers who strive to engage their students in proving
4. *Instructional support*: the inadequate instructional support offered or available to elementary teachers about how to achieve that goal in their classrooms. (pp. 21-24)

The interdependence and multiplicity of factors hampering imply no easy solution to elevating the place of reasoning-and-proving in elementary mathematics classroom. In this study, these factors will serve as a frame for analysis of in-service teachers' applying the given task, method, and taxonomy.

Method

Case study method was used for this study. The study involved both qualitative and quantitative research methods with data from in-service teachers' project reports, questionnaires, and from interviews conducted by the researchers. This section describes the participants, setting, instruments, and procedures for analysis used for the case study.

Participants

As in the first study (Brodahl & Wathne, 2018), subjects are in-service teachers in upper primary and lower secondary school enrolled in Year 1 of a national program of continuing university mathematics education, called "Competence for Quality", delivered entirely online. The program focuses on teachers with general teaching certificates who already work as teachers and teach mathematics. It provides scholarships for further training to increase the teachers' formal competence in mathematics and mathematics education to meet new qualification requirements for teaching mathematics. Different from the first study with data from autumn 2016, this study drew on data from the following year's program in autumn 2017.

All 52 in-service teachers who attended the course constituted the new purposive sample. Of those fifty-two, 32 teachers gave their consent to participate in the research (61.5 % of the sample with 15 male and 17 female) from across the country ranging in age from 28 to 56 (mean 44.5, median 45). The classes they taught ranged from grade four to 11 with 10 participants teaching upper primary, 21 participants teaching lower secondary, and one teaching upper secondary level.

Setting for the Study

Like the setting in the first study (Brodahl & Wathne, 2018), in-service teachers were assigned to plan and accomplish a teaching session where they should apply the method of imaginary dialogues in their classroom, presenting the same "shaking hands" dialogue between Knut and Idunn (p. 34) and letting students continue working on the mathematical problem in pairs or groups of three.

In preparation, in-service teachers were introduced to Balacheff's levels by a video providing characteristics for each level as well as exemplifying how students may argue for the sum of two odd numbers to be even on the respective level. In-service teachers were also introduced to the idea of imaginary dialogues as a method to get students started and working with reasoning-and-proving in the classroom. They were offered six examples of dialogues, called “start dialogues”, among them the one on the handshake problem to be used in the task. After the teaching session, in-service teachers reported on their experience.

Data Collection

In-service teachers’ reports after their session in class was a task in two parts. The first part was to briefly describe the planning and implementation of the session, what was expected and what was experienced. The second was to pick up and present two of their students’ dialogues and identify any reasoning-and-proving based on Balacheff’s hierarchy of proof levels in school mathematics. The project reports were a required pass/fail assignment for the course to be submitted in the learning management system, Canvas, by the deadline.

The questionnaire opened with three close-ended elements targeting participants’ experience with reasoning-and-proving before project start. Data were mainly drawn on 17 elements from the following two parts, both close and open-ended. Part 1 covered in-service teachers’ experiences from their lesson with imaginary dialogue and reflected research questions 1 and 2. Along four topics, part 1 asked to describe experiences with the implementation of imaginary dialogues in classroom:

1. Specific aspects of the session with imaginary dialogues that went well.
2. Specific aspects of the session with imaginary dialogues that could be improved.
3. Perceived usefulness of the session with imaginary dialogues.
4. Insights into students’ mathematical reasoning-and-proving gained in the session.

One statement concerned with the perceived usefulness of imaginary dialogues was to be rated on a 1-10 scale. Eight statements were posed, and respondents asked to indicate on a five-point Likert scale what best represented their experiences with imaginary dialogues in teaching.

Part 2 covered in-service teachers’ experiences from their analysis of students’ written dialogues and reflected research question 3. Along two topics, it asked to describe experiences with identifying students’ reasoning-and-proving based on Balacheff’s theory of proof levels:

1. Specific aspects according to Balacheff’s level classification that were helpful.
2. Specific aspects according to Balacheff’s level classification that were challenging.

It contained one five-point Likert scale statement and one 1-10-point scale statement to indicate their experience with and perceived usefulness of Balacheff’s levels of proof.

Different from the previous study, semi-structured interviews were chosen as the primary source of data for this study because usefulness and insights identified in the questionnaires were not directly observable. The participants received the interview guide, structured in four parts, prior to the interviews. Part A targeted in-service teachers’ background and experience with reasoning-and-proving before project start. For parts B and C, the point of departure was in-service teachers’ questionnaire response to the respective 1-10-point scale statement, respectively in parts 1 and 2, followed by additional questions taking the form of statements (e.g., “In the questionnaire, you answered…” or “... you described…”), seeking in-service teachers’ explanations and clarifications. Part D targeted their project report and asked to comment their findings (e.g. "You responded… Could you elaborate?").

The study was announced in Canvas where in-service teachers could give their informed consent to complete an online questionnaire (using SurveyXact.com) and allow their project report to be used in the research, in addition to a possible follow-up phone interview. "Reflecting the variety of experiences represented in questionnaires" was announced as the main criterion for selecting interviewees from the pool of volunteer candidates.

Thirty-two agreed to provide their reports and reply to questionnaire; ten of them agreed to be interviewed. Seven participants completed the interviews: four males and three females of varying ages (32.5-56.4 years, average 46.3). Interviews were conducted by phone after the exam and transcribed. Four of them are presented in the study: two males and two females, 32.5-56.4 years, average 47.6, who resided in different parts of the country. The criterion of data saturation was used to determine whose data were used. The three remaining did not yield considerably new information. The written material was anonymized before analysis.

Data Analysis

In-service teachers' reflections in the concluding part of their reports were analyzed by both researchers, and dominant themes and codes were identified and subsequently applied to all reports. They were discussed and structured, then used to refine the research questions and to build the questionnaire.

Fixed-choice responses in questionnaires were organized in Microsoft Excel and the open-ended descriptions in Word. All data were encrypted and shared between the researchers. Coding was mainly guided by the research questions and questionnaires' themes. Both researchers analyzed and coded the descriptions independently, then together organized themes and codes in a multifaceted codebook in an iterative process using inductive and deductive approaches (Bryman, 2012). Independently coding and recoding the data set, they compared and discussed coding until consensus was established. In a corresponding procedure as used for questionnaire responses, researchers' interview transcriptions were coded with the codebook as a sound basis to build on and analyzed using a content analysis approach (Bryman, 2012).

Results

Thirty-two in-service teachers in the class (61.5 %), hereafter called participants, agreed to provide their reports and reply to a questionnaire. Interviews with four of them established data for analysis (7.7 %). In the questionnaires, 44 % of the participants declared to have received an introduction into proof and argumentation prior to the project. Two-fifths stated not having experience with designing mathematical claims and arguments for and against; 38% asserted to have included very little argumentation and proof in their teaching.

Questionnaires and Reports

According to participants' ranking (Table 1), 71.9 % of the in-service teachers perceived imaginary dialogues useful (7-10) as a tool for engaging students in reasoning-and-proving in the mathematics classroom, 12.5 % not useful (1-4), while 15.7 % responded neutrally (5-6) about their usefulness. As for the perceived usefulness of Balacheff's levels of proof, 68.8 % thought they were useful, 12.5% not useful, and 18.7% rated the question neutrally.

Table 1
Rankings of Perceived Usefulness

Statement	10	9	8	7	6	5	4	3	2	1
Imaginary dialogues	**4** *(12.5)*	**4** *(12.5)*	**6** *(18.8)*	**9** *(28.1)*	**2** *(6.3)*	**3** *(9.4)*	**3** *(9.4)*	**1** *(3.1)*	**0** *(0.0)*	**0** *(0.0)*
Balacheff's levels of proofs	**4** *(12.5)*	**3** *(9.4)*	**7** *(21.9)*	**8** *(25.0)*	**5** *(15.6)*	**1** *(3.1)*	**1** *(3.1)*	**2** *(6.3)*	**1** *(3.1)*	**0** *(0.0)*

Note: *n*=32. Response frequencies in bold, percentage italicized. Participants were asked to rate on a scale of 1 to 10 with 10 being very useful and 1 not useful.

The count of responses (see Table 2) to statements 1-5 concerned the use of imaginary dialogues in reasoning-and-proving tasks. Nearly half of the participants (46.9 %) agreed to have experienced the lectures challenging to prepare or implement. While half (50.1 %) perceived that students did not immediately understand the task or start writing the imaginary dialogue, 53.1 % perceived that students enthusiastically continue writing, and 71.9 % found students explaining their thoughts and putting their ideas into words – building mathematical arguments. As to statement 6, 81.3 % found Balacheff's levels of proof useful in identifying students' reasoning-and-proving. Statements 7-9 on future directions (7-9) revealed that most (81.3 %) anticipated imaginary dialogues useful in teaching when the teacher and the class have more experience. Likewise, 81.3 % expressed that they want to continue using imaginary dialogues. Finally, 87.5 % indicated the task revealed the importance of providing their students with exploring and explaining opportunities.

Table 2
Participant Experiences

	Responses					
Statements	SA	SLA	N	SLD	SD	TA
My lecture on imaginary dialogue was challenging to prepare or carry out.	**5** *(15.6)*	**10** *(31.3)*	**6** *(18.8)*	**9** *(28.1)*	**2** *(6.3)*	**15** *(46.9)*
The students understood the task and continued writing the imaginary dialogue.	**3** *(9.4)*	**8** *(25.0)*	**5** *(15.6)*	**14** *(43.8)*	**2** *(6.3)*	**11** *(34.4)*
The students were enthusiastic when they continued to write an imaginary dialogue.	**7** *(21.9)*	**10** *(31.3)*	**3** *(9.4)*	**10** *(31.3)*	**2** *(6.3)*	**17** *(53.1)*
The students explained their thoughts and put their ideas into words when they continued to write an imaginary dialogue.	**10** *(31.3)*	**13** *(40.6)*	**4** *(12.5)*	**2** *(6.3)*	**3** *(9.4)*	**23** *(71.9)*
My students built an argument when they continued writing an imaginary dialogue.	**6** *(19.8)*	**17** *(53.1)*	**4** *(12.5)*	**4** *(12.5)*	**1** *(3.1)*	**23** *(71.9)*
Balacheff´s levels of proof were useful in identifying my students' reasoning-and-proving.	**10** *(31.3)*	**16** *(50.0)*	**3** *(9.4)*	**2** *(6.3)*	**1** *(3.1)*	**26** *(81.3)*
Imaginary dialogues will be useful in teaching when I and the class get more experience.	**14** *(43.8)*	**12** *(37.5)*	**5** *(15.6)*	**1** *(3.1)*	**0** *(0.0)*	**26** *(81.3)*
I will continue to use imaginary dialogues in my teaching.	**12** *(37.5)*	**14** *(43.8)*	**4** *(12.5)*	**2** *(6.3)*	**0** *(0.0)*	**26** *(81.3)*
Working with imaginary dialogues has shown how important it is that the students are allowed to explore and explain.	**11** *(34.4)*	**17** *(53.1)*	**3** *(9.4)*	**1** *(3.1)*	**0** *(0.0)*	**28** *(87.5)*

N=32 *Note*. Response frequencies in bold; percentages italicized. Key for Table 2: strongly agree (SA); slightly agree (SLA); neither agree or disagree (N); slightly disagree (SLD); strongly disagree (SD); and total agreement (TA) adds SA and SLA together.

The most frequent experiences and perceptions quoted in open-ended statements from the questionnaire are grouped based on the codebook. The groups are listed in descending order of frequency:

- What-went-well with imaginary dialogues or was perceived useful?
    - Students became committed in a new way.
    - Students explored and expressed mathematics ideas.
    - Teachers increased their awareness of students' capability of argumentation and need of starting early to train.
    - Students engaged in mathematical discussion.
    - Students, usually not active in mathematics lessons, participated.
- Insights into students' mathematical reasoning-and-proving.
    - how they approached and coped, e.g. point of departure, angle of entry, path of thinking, conjectures and testing
    - how they used their knowledge and where they came to a halt
    - the large variation of reasoning-and-proving in class
- Perceived usefulness with Balacheff's level classification
    - helped identifying and distinguishing students' levels of reasoning-and-proving
    - provided a system of concepts and notions
    - arose teachers' awareness of own teaching and students' need to train reasoning skills
- Perceived challenges with Balacheff's level classification
    - to separate the levels and determine students' proficiency
    - to place the students' dialogue on right level
    - to transfer theory to practice

Other most frequently mentioned issues: Almost half of the participants brought up examples of students struggling with writing down the continuing dialogue. Almost one third perceived the method of imaginary dialogues less suitable for some of their groups, including low-achieving students, students with foreign background or behavioral difficulties, or immature students being most keen to fool about. One third emphasized, optimistically or apologetically, that this was both for them and students a new method to get familiar with so that they could succeed better. The most prevailing improvement suggestion was to spend more time on both introducing the method of imaginary dialogues, and next time applying it to students' writing imaginary dialogues or presenting their findings.

The reflections in the participants' reports deal with the same issues as their responses to the open-ended questionnaire statements do and substantiate these.

Interviews

In-service teacher A, sixth grade: She explained her reasons for ranking usefulness of imaginary dialogues, 10 of 10 points, by "it was amazing to get to know the students' way of thinking", "to get better acquainted with the students' ability to argue", and "in fact, to realize that [the students] need to formulate early and explain why." Teacher A was surprised by "the diversity in my class" and explained, "I got more insight into how [the students] think when I use [this] method. Having sufficient time to argue, students choose to look for possible approaches to solve the problem, and not just the right answer". She explained her reasons for ranking usefulness of Balacheffs' levels of proof, 9 of 10 points, "I might then be able to see when my students actually take the step away from a practical approach, but at the same time when they are at such a low level, [...] they may just get up and 'have a sniff' [towards next level]". She gained insight into her students' levels, as being pragmatic, confirmed that her students "have not moved on to a conceptual proof", and pointed out the importance of "the teacher being able to explain at that level as well".

In-service teacher B, eighth grade: She explained her reasons for ranking usefulness of imaginary dialogues, 8 of 10 points, by "[the students] in a way go into a role where they are other people", and "[the students] both thought, and they wrote". Teacher B assessed insight in students' reasoning-and-proving and their explanations on "why" and "how" they are using the different approaches. She explained giving 8 of 10 points for the usefulness of Balacheff's levels by the system being a "step by step" hierarchy and great "to put [students' arguments] in the system". Teacher B sought to facilitate for students "to become aware of where they are" and "try to step further".

In-service teacher C, ninth grade: He ranked usefulness of imaginary dialogues 7 of 10 points and strongly emphasized that when students "have to explain in their own words how and why they do it, they will learn in a better and deeper way". It was his clear experience that "the students only were concerned with determining the solving, not with the way to the solution", just like "they tend to be in the classroom". Teacher C observed that his students "talked much better together than they wrote down". Ranking the usefulness of Balacheff's levels with 8 out of 10 points, he explained that "using the Balacheff levels means that you get some shelves to sort on", however, students could slightly change level along their path and move on from one level towards the next. Also, a group may for a short while reach a higher level, but then fall down to the lower. "Covering the levels a student may take, [Balacheff] provides a systematics that is easy to deal with". The written dialogues, he assumed, "may also support formative evaluation", and continued, "Students reaching a higher level, show that they manage to develop their reasoning-and-proving."

In-service teacher D, tenth grade: He explained his high ranking, 10 of 10 points, for usefulness of imaginary dialogues, saying: "I clearly see the advantage of mathematical [formulations] entering into their language", and "Language and thoughts connect [in their] mathematical argumentation". Teacher D experienced "committed students", and his "insight gained was that students' argument develops within the taxonomy of Balacheff". Explaining his highest ranking for the usefulness of Balacheff's levels, he expressed "how fun it was to discover the preciseness of Balacheff's model and how easy to place students' argument into it".

Discussion

A high percentage of in-service teachers reported to have been a little familiar to reasoning-and-proving before the project. This response is consistent with teachers' preconditions in the literature. According to Stylianides (2016), the weak knowledge about proof appears as *factor 1* of four challenges involved for non-specialist teachers of mathematics (see "Theoretical Framework" section). This weakness makes proof hard to teach and contributes to a rather marginal place of reasoning-and-proving in the elementary classroom.

Teacher educators offered in-service teachers the method of imaginary dialogues to promote their students' mathematical reasoning-and-proving, as well as Balacheff's taxonomy to identify students' argumentation. This combination should aim to remedy some of the hardships of teaching and learning proof, *factor 4*, inadequate instructional support (Stylianides, 2016) and hopefully enable elevating reasoning-and-proving in classroom. Still, teachers may rise both unique and similar first experiences, as their prior mathematical knowledge, their learning and teaching experiences, and context, differ.

Research Question 1

In answering the first research question – how they perceived the usefulness of imaginary dialogues as means to engage students in reasoning-and-proving – in-service teachers rated their experience as positive. This positivity also appears in the interviewees' open answers and interview. However, their reasons differ slightly. Teacher A, C, and D concurred that students' formulating, explaining, and reflecting are the greatest benefits. These teachers alluded to the following: The written dialogues helped the teacher realize students' need to learn to formulate and explain early on in their education; writing the dialogues supported language and connecting thoughts; by having students explain in their own words how and why they do it in a certain way, they learn better and deeper.

Teacher B saw the usefulness of students taking a different role: Writing the dialogues let students enter into a role where they, more fearless, act as other people. According to the questionnaires, some students took on the given roles in the start dialogue, while some others preferred arguing "as themselves". Understandably, one teacher asserted, "Knut and Idunn in the start dialogue appeared a bit too enthusiastic to relate to." It is valid to speculate that students' age may play a role. The older the students, the less likely they were to take the characters' roles.

Perceived usefulness may undergird imaginary dialogues to have the potential to support in-service teachers in engaging their students in mathematical reasoning (Wille, 2017). Teacher A's experience of the usefulness of the method in her six-grade class and perception of the possibility and necessity of beginning early is in line with Stylianides (2016) who stressed the importance of developing mathematical argumentation early in the elementary school.

Research Question 2

Analyzing all questionnaires to answer the second research question – what types of insight in students' process of reasoning-and-proving they perceived gaining in students' written dialogues – revealed that many teachers could detect both well-running approaches and where students came to a halt. This response was clearly valid for three of the interviewees. Teacher B left this answer blank. However, all four exemplified and highlighted the importance of support given to gain insight in students' way of thinking and different approaches. These insights then could create meaningful learning opportunities for their students to engage in reasoning-and-proving. The findings suggested imaginary dialogues can be cited as instructional support, a method available to elementary teachers to engage their students in powerful mathematical activity including reasoning-and-proving. Instructional support is rarely available, or it is inadequate (Stylianides, 2016); *factor 4* is a synergy of many factors relating to the marginal place of reasoning-and-proving. Teacher educators may welcome this contribution.

Teacher C expressed concern that only a small number of students focused on the approach to the handshake problem. Instead, they were more concerned with the solution, finding a number, a pattern, or a formula. The students resorted back to the way they used to do things in the whole-class teaching situation. Even the teachers who had experience with teaching reasoning-and-proving still encountered challenges getting students to shift their paradigm due to established pedagogical practices. Other researchers found similar high pedagogical demands, constituting *factor 3,* related to the marginal place of reasoning-and-proving (Stylianides, 2016).

Research Question 3

In answering the third research question – how in-service teachers perceived the usefulness of Balacheff's taxonomy to identify levels of reasoning-and-proving in students' written dialogues – they rated their experience as positive. Reasons for their positivity include its suitability and applicability to identify students' levels of reasoning-and-proving, provide a system of concepts and notions, and contribute to awareness of their own teaching. The interviewees all found Balacheff's taxonomy useful and, for the most part, easy to deal with. It assisted the teachers in recognizing their students' ability to make arguments and their progression in reasoning-and-proving.

While stating the usefulness of the taxonomy, in-service teachers experienced challenges with Balacheff's classification system. However, in-service teachers' perceived challenges may partly be due to the weak knowledge about reasoning-and-proving that many teachers reported in this study (cf. *factor 1*). Teacher C called for more levels to sort on, stating in the questionnaire, "Placing the students at the right level can be challenging, and there will be sliding transitions." As to preciseness and number of levels, there are later extensions on the base of Balacheff's taxonomy. Miyazaki (2000) added extensions with six levels of algebraic proof in lower secondary school mathematics, along with contents of proof, representation of proof, and students' thinking. Stylianides (2008) also added to the framework with the psychological and the pedagogical component to the mathematical one with four reasoning-and-proving activities. However, Teacher D found the taxonomy precise enough to place students' argument into. Notably, in the questionnaire, he pointed out the benefit of adopting Balacheff's terminology for gaining an even deeper insight into students' written argument, "I started to use these terms and, by this, came to understand [...] mediated through them. It helped me arguing and reflecting on my reflection in Balacheff's way of thinking." Balacheff's taxonomy contributing to a teacher's deeper insight in their students' reasoning-and-proving may encourage teacher educators who aim to provide required instructional support (cf. *factor 4*).

Conclusion

The problem addressed in this paper concerns teachers currently giving a marginal place to reasoning-and-proving activities. Many in-service teachers in our sample were inexperienced to reasoning-and-proving. They may not have learned about proofs themselves or are not aware of the importance of teaching reasoning. The goal for teacher educators is to develop teachers being prepared to lead to better student experiences of corresponding sense-making activities (Stylianides, 2008). In pursuing this aim, the study focused on assignment design in teacher education.

How did it work to provide teachers with the combination of a mathematical task, a method in classroom, and a taxonomy for analysis of the students' work? Is it worth further developing such instructional support? An important finding from the study is that it is possible to affect and improve teachers' engagement to help their students with learning to reason-and-prove. We interpret the case study results as an indication that planning and accomplishing this teaching session with the method of imaginary dialogues applied to the handshake task was of great help, ignited many teachers, and made them curious about reasoning-and-proving. For teachers, it is a journey not free from obstacles when first time implementing (Brodahl & Wathne, 2018), but most of them expected that both teachers and students will do better next time. The study revealed engaging students to reason, argue and prove, while making students' argumentation visible for teachers was perceived the most useful with imaginary dialogues. Further, in-service teachers were given a taxonomy for analysis. Their increasing awareness of levels of argumentation was perceived to be the most useful with getting exposed to Balacheff's distinctions. They got a little glimpse and want more. The combination of this mathematical task, this method to approach students' mathematical thinking processes, and this taxonomy for analysis may be part of what teacher educators can provide teachers with to help them generate a reasoning-and-proving activity and identifying students' argumentation. Thus, the study suggests this combination as a possible task design for teacher educators' instructional support.

References

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

Brodahl, C., & Wathne, U. (2018). Imaginary dialogues – In-service teachers' steps towards mathematical argumentation in classroom discourse. *Journal of the International Society for Teacher Education*, *22*(1), 30–42.

Bryman, A. (2012). *Social research methods*. Oxford: Oxford University Press.

Ministry of Education and Research (2013). *Curriculum for the common core subject of mathematics (MAT1-04).* Oslo: Utdanningsdirektoratet.

Miyazaki, M. (2000). Levels of proof in lower secondary school mathematics. *Educational Studies in Mathematics*, *41*(1), 47–68. doi: 10.1023/A:1003956532587

Reid, D. (2005). The meaning of proof in mathematics education. *European Research in Mathematics Education IV. Proceedings of CERME4*, San Feliu de Guíxols, Spain.

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

Stylianides, A. J. (2016). *Proving in the elementary mathematics classroom*. Oxford, UK: Oxford University Press.

Stylianides, G. J. (2008). An analytic framework of reasoning-and-proving. *For the Learning of Mathematics*, *28*(1), 9–16.

Wille, A. M. (2017). Imaginary dialogues in mathematics education. *Journal für Mathematik-Didaktik*, *38*(1), pp 29-55. doi: 10.1007/s13138-016-0111-7

Authors

Unni Wathne, Ph.D., is an Associate Professor in mathematics education at the University of Agder. She is interested in and does research on teaching and learning mathematics in all phases of education.

Cornelia Brodahl is an Associate Professor in ICT and learning at the University of Agder and holds a master's degree in mathematics. Her research interests include mathematics education with digital tools.