Review of Education
An International Journal of Major Studies in Education
BERA
British Educational Research Association

Review of Education
Vol. ●●, No. ●●, ●● 2018, pp. ●●–●●

DOI: 10.1002/rev3.3157

Secondary students' proof constructions in mathematics: The role of written versus oral mode of argument representation

Andreas J. Stylianides*
University of Cambridge, UK

Prior research showed that many secondary students fail to construct arguments that meet the standard of proof in mathematics. However, this research tended to use survey methods and only consider students presenting their perceived proofs in written form. The limited use of observation methods and the lack of consideration of students presenting their perceived proofs orally—in tandem with their written proofs for the same claims—might have resulted in a skewed picture of the potential of students' constructed proofs, and this raises concern about the validity of research findings. The research reported in this article substantiates this concern. Using data from a design experiment in two secondary mathematics classrooms (14–15-year-olds), I explored the role of the written versus the oral mode of argument representation in students' proof constructions. Findings from the comparison between the written arguments (perceived proofs) that the students produced during small group work and the oral arguments that the students presented in front of the class for the same claims showed that the oral mode of representation is more likely than the written mode to be associated with the construction of arguments that meet the standard of proof. Thus if a study had analysed students' written arguments only (as in survey research), it would have reported a less favourable picture of the potential of students' constructed proofs than another study that would focus only on students' oral arguments (as in observational research). Implications for methodology, research and practice are discussed in light of these findings.

Introduction

The concept of 'proof' is fundamental to deep learning in mathematics and in various countries it is considered to be important for students' mathematical experiences across all levels of education, as early as the primary school (e.g. Ball & Bass, 2003; Yackel & Hanna, 2003; National Governors Association Center for Best Practices & Council of Chief State School Officers [NGA & CCSSO], 2010; Department for Education, 2013). For example, one of the three core aims that the national mathematics curriculum in England sets for students of all ages relates to proof: '[Students should] reason mathematically by following a line of enquiry, conjecturing relationships and generalisations, and developing an argument, justification or proof using mathematical language'

*University of Cambridge, Faculty of Education, 184 Hills Road, Cambridge CB2 8PQ, UK. Email: as899@cam.ac.uk

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

(Department for Education, 2013, p. 3). The concept of proof is also hard-to-teach and hard-to-learn, and thus over the past few decades it has attracted significant attention internationally by researchers in the field of mathematics education (for reviews of the state of the art in this area, see Harel & Sowder, 2007; Mariotti, 2006; Stylianides *et al.*, 2016, 2017).

A main research strand has focused on secondary (i.e. post-primary) students' constructions of mathematical arguments, showing that many secondary students fail to produce arguments that meet the standard of proof (e.g. Senk, 1989; Healy & Hoyles, 2000; Küchemann & Hoyles, 2001–03; Knuth *et al.*, 2009). For example, in a large-scale study Senk (1989) asked 1520 secondary students taking a geometry class in the USA to prove four theorems, two of which required only a single deduction beyond the hypotheses. Senk found that only 30% of the students were able to prove at least three theorems, while 29% were unable to construct a single proof.

However, the studies in this research strand have tended to only consider secondary students presenting their perceived proofs in written form, primarily in the context of survey studies. The lack of consideration by these studies of secondary students presenting their perceived proofs orally—in tandem with students' written proofs for the same claims—might have resulted in an incomplete or a skewed picture of the potential of students' constructed proofs, and this raises concern about the validity of research findings. In this article I explore the possible role of the *mode of representation* in secondary students' argument constructions by addressing the following research question: How do the degrees to which a student's argument (perceived proof) for a claim approximate the standard of proof compare when the bulk of the argument is communicated using a written versus an oral mode of representation?

The research question was purposefully phrased so as to leave open the possibility for any kind of relation to emerge between the degrees to which written and oral arguments by the same student and for the same claim approximate the standard of proof. Indeed there is no systematic research evidence to support a hypothesis for a unidirectional relation, such as that the oral mode of representation is more likely, compared with the written mode, to be associated with arguments that approximate the standard of proof. The empirical investigation of the research question took place in the natural context of secondary students' engagement with communicating arguments that they perceive meet the standard of proof, namely, the classroom. This parallels research on mathematicians' engagement with communicating their finished proofs via both written and oral modes of representation in various contexts relevant to their work, including journals, conferences and lectures (e.g. Burton & Morgan, 2000; Weber, 2004; Artemeva & Fox, 2011; Greiffenhagen, 2014; Lai & Weber, 2014).

Background

The concept of proof

The construction of arguments that meet the standard of proof is often the last step in a complex and multifaceted activity typically referred to as *proving* that includes also other processes such as the following: work with examples or exploration of particular cases, identification of patterns and generation or refinement of conjectures or other

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

kind of mathematical claims, and attempts to develop informal arguments for these mathematical claims that may offer insight, or ultimately translate, into a proof (e.g. Mason, 1982; Boero *et al.*, 1996; Weber & Alcock, 2004; Mariotti, 2006; Stylianides, 2007, 2008; Alcock & Inglis, 2008; Zazkis *et al.*, 2008, 2016). In this article, I focus on the presentation of mathematical arguments whose constructors perceive they meet the standard of proof and who might have engaged previously in some other processes within the broader activity of proving. In other words, my focus is on the presentation of the *final* product of an individual's proving activity.

This raises the question of what counts as a proof. There is a multiplicity of perspectives on and usages of the term 'proof' in the mathematics education research literature (for summaries of some of these perspectives, see Balacheff, 2002; Weber, 2008; Stylianides *et al.*, 2017). At the same time, there is recognition that it is neither sensible nor desirable for all researchers to adopt a common definition of proof, for specific research goals may be served better by different definitions of proof (Reid, 2005; Stylianides *et al.*, 2017). Indeed, the important point here seems to be that researchers specify clearly the definition of proof that has underpinned their research, as well as their reasons for their choice, so as to facilitate understanding of their findings and comparison of findings across studies (Balacheff, 2002).

Being consistent with this call for specificity, I state the definition of proof that has underpinned the classroom-based research at the secondary level that I report in this article:

> *Proof* is a *mathematical argument*, a connected sequence of assertions for or against a mathematical claim, with the following characteristics:
>
> 1 It uses statements accepted by the classroom community (*set of accepted statements*) that are true and available without further justification;
> 2 It employs forms of reasoning (*modes of argumentation*) that are valid and known to, or within the conceptual reach of, the classroom community; and
> 3 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. (Stylianides, 2007, p. 291; italics in original)

This definition breaks down a mathematical argument into three components—the set of accepted statements, the modes of argumentation and the modes of argument representation—and imposes on each of them certain requirements before the argument can be said to meet the standard of proof. In describing the requirements that the three components should satisfy for the argument to be a proof, the definition seeks to achieve a defensible balance between two important considerations: the discipline of mathematics and the classroom community that engages with the production of a proof. For example, the definition requires that proofs not only use *valid* modes of argumentation (cf. the first consideration) but also that these modes be *known* or *conceptually accessible* to the classroom community where the proofs are produced (cf. the second consideration). The merits of seeking to achieve a balance between these two considerations, both in relation to proof and more generally in relation to the teaching and learning of mathematics, have been discussed in Stylianides (2016).

An important implication of the definition, and a main reason for its choice in this research, is that it makes clear that one's ability to construct a proof cannot be

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

dissociated from one's readiness to communicate the respective argument using appropriate modes of representation. Other available definitions of proof (see, e.g. Griffiths, 2000; Mariotti, 2000; Healy & Hoyles, 2001; Knuth, 2002) tend to view proof as a logical deduction linking premises with conclusions, thus considering the use of the set of accepted statements and modes of argumentation in Stylianides' definition but not explicitly the modes of argument representation, which is the key construct of the research reported herein. Another reason for the choice of Stylianides' definition is that, unlike other available definitions of proof that attend mostly to mathematics as a discipline by emphasising the logico-deductive aspects of arguments that meet the standard of proof (ibid.), Stylianides' definition attends also to the characteristics of the classroom community that engages with the production of a proof. This is important for a classroom-based study like the one reported in this article.

Examples of modes of representation that can be used when communicating a mathematical argument or a proof include written language, oral language, diagrams or drawings, concrete materials, and gestures. For example, some mathematicians who present previously completed proofs in front of an audience (such as their students in an undergraduate class) write words and create diagrams or drawings on a board while simultaneously gesturing and describing orally the proofs as the presentation unfolds (Artemeva & Fox, 2011; Greiffenhagen, 2014). Two modes of representation—the written and the oral—are main communication means through which students describe their thinking processes in mathematics (Pugalee, 2004) and through which the bulk of a proof is often communicated to an audience and are thus the focus of this research. I return to modes of argumentation in the next section.

Stylianides' (2007) definition of proof does not impose variable requirements on the standard of proof by the mode of representation that is used in an argument that is perceived to be a proof in a classroom community, and so, within this perspective, an oral argument should be evaluated against the requirements for a proof the same way as the argument's respective written (transcribed) version. The situation is more complex when it is unclear what the perceived proof is, such as in the case when an oral proof is co-constructed by many individuals like the teacher and the students during a lesson. In cases such as this one it is methodologically challenging to identify which contributions in the verbal discourse of the classroom are part of the proof and how the relevant contributions fit in together to actually constitute the proof. González and Herbst (2013) discussed how using linguistic tools can allow researchers to deal with these methodological challenges to identify and map the oral proof onto a written version. However, in the research I report in this article these methodological challenges are being bypassed by the fact that in the observed lessons the student participants presented orally, and in the most part undistracted by other classroom members, the *final* products of their *own* proving activity.

Representations in proofs

Despite the wide appreciation of the importance of representations and mathematical language more broadly as tools for communication linked both to mathematicians' and to students' mathematical thinking (e.g. Burton & Morgan, 2000; Lamon, 2001; Pugalee, 2001, 2004; Sfard, 2001; Greiffenhagen, 2014), the modes of representation

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

in secondary students' proof constructions have not received much attention by research thus far. Also, in those studies where the modes of representation did receive attention, including relevant studies with university students, the specific role played by a written versus an oral mode of representation in students' proof constructions has not been an explicit research focus.

Research on secondary students' engagement with proof has tended to focus on students' difficulties with logical ideas underpinning mathematical arguments or proofs, such as the inadequacy of a few confirming examples to establish the truth of a general claim (e.g. Healy & Hoyles, 2000; Knuth *et al.*, 2009) or the meanings of logical implication and logical rules of inference (e.g. Hoyles & Küchemann, 2002; Yu, Chin, & Lin, 2004). This is a sensible research focus, for, according to Stylianides' (2007) definition of proof, a key expectation from mathematical arguments that meet the standard of proof, which is also one that many students have difficulty grasping, is that the arguments use *valid* modes of argumentation (*modus ponens*, *modus tollens*, etc.) to derive conclusions from accepted statements (definitions, axioms, etc.). Some studies explored the use of concrete materials or drawings as ways to reduce abstraction and support young children in argument construction (e.g. Morris, 2009; Schifter, 2009), while a few studies with older students explored the role of diagrams in mathematical reasoning and found little relationship between students' preference for visualisation and their achievement in mathematics (e.g. Wheatley & Brown, 1994; Presmeg, 2006). These studies have offered useful insights into the role that concrete materials, drawings and diagrams play, or might play, in students' argument constructions but not into how the written or verbal modes of representation might be associated with arguments that meet the standard of proof.

The most relevant piece of research with secondary students that I was able to find was an analysis of video records of geometry lessons when secondary students presented in front of their class their previously completed (written) arguments that they perceived to be proofs (Dimmel, 2015; cited in Dimmel & Herbst, 2017). A communication practice in the observed lessons was students presenting their proofs by creating on the board mark-for-mark reproductions of their previously completed proofs, an act that Dimmel and Herbst (2017) called 'proof transcription' and found secondary mathematics teachers in the USA perceived to be a routine communication practice when students presented their proofs in front of their class. The students' presented arguments were, in this case, the same as their transcribed written arguments, with no verbally described version of these arguments to compare with, and so again no light is cast on the particular issue of interest in this article. Dimmel and Herbst offered a fair criticism of 'proof transcription' as being out of line with proof presentation done by disciplinary experts in front of an audience, where an oral commentary would normally be a key part of the presentation. In the classroom-based research that I report in this article, the secondary students were expected to engage with proof presentation in front of their class in a way that approximated more disciplinary experts' presentations than the practice of 'proof transcription' considered to be routine in secondary school geometry classes in the USA.

Research on university students' engagement with proof has paid considerably more attention to the role of modes of representation in proof construction, especially in relation to the use of diagrams, and so it is pertinent to look at that literature too despite my focus in this article on the secondary school level. Some researchers have

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

asked university students to explain their reasoning orally as they wrote their proofs (e.g. Alcock & Simpson, 2004, 2005; Weber & Alcock, 2004; Weber, 2005; Zazkis *et al.*, 2016; Mejía-Ramos & Weber, in press), but these researchers did not compare or explore the relationship between students' oral and written arguments (K. Weber, personal communication, 6 July 2018). Consider, for example, a study by Mejía-Ramos and Weber (in press) where 73 mathematics majors were asked to think aloud while they were proving seven calculus theorems, and then write their proofs up as if they were doing so for an examination. In their analysis of the available data, Mejía-Ramos and Weber explored (among other things) the possible correlation between students' diagram usage during the broader activity of proving (i.e. while the students were attempting to construct a proof) and students' success at the end of this activity as reflected in their final written products (i.e. students' perceived proofs). Here, like in other similar studies at the university level that used think-aloud research designs, students' verbal descriptions of their thinking processes during their search for a proof might have included oral arguments that students perceived to be proofs, but such arguments, *if* present in students' verbal descriptions, were not singled out by the researchers or the students themselves for comparison with students' written proofs.

Of course the fact that this comparison was not made is by no means a criticism of these studies, which were designed to make other important contributions in the area of proof. For example, Mejía-Ramos and Weber's (in press) study has extended prior research knowledge about the role of diagrams in university students' and mathematicians' argument constructions, including the process of 'translating' between informal arguments with diagrams and written proofs (e.g. Alcock & Simpson, 2004; Samkoff *et al.*, 2012; Zazkis *et al.*, 2016), by showing that there was little correlation between participants' propensity to use diagrams and their success in writing a proof in response to a proving task.

Finally, I searched the mathematics education research literature beyond the area of proof for any comparisons between students' use of the written and oral modes of representation. My search returned a relevant study by Pugalee (2004) who compared secondary students' oral and written descriptions of their thinking while solving mathematical problems. The study's sample comprised 20 secondary students (ninth graders) from an introductory algebra class who worked individually to each solve six problems of different levels of difficulty. For half of the problems the students were asked to write everything that came to their mind while solving the problems, whereas for the other half the students were asked to think out aloud by telling everything that came to their mind as they solved the problems. The assignment of students to problems and conditions (written vs. oral) controlled for the students' mathematical and linguistic abilities and for the problems' level of difficulty. Pugalee found that students who wrote descriptions of their thinking while solving the problems were statistically more successful in the problem-solving tasks than students who verbalised their thinking.

On the surface, these findings extrapolated to the area of secondary students' proof constructions might appear to lend support to the hypothesis that secondary students' written arguments (perceived proofs) are more likely to be successful (i.e. meet the standard of proof) than students' oral arguments for the same claims. However, such an extrapolation would be unwarranted owing to some fundamental differences between Pugalee's study and the study I report herein. One major difference concerns

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

the objects of students' written and oral descriptions in the two studies: *everything* that came to students' mind *while* solving the set problem-solving tasks in Pugalee's study versus the *final product* of students' work from solving the set proving tasks in my study. Another major difference is that in Pugalee's study each student described either in writing or orally his or her thinking while solving a given problem, whereas in my study each student presented both in writing and orally his or her perceived proof for a given claim thus allowing direct comparison between the two. Although Pugalee's (2004) study is not directly relevant to the issue I explore in this article, the study has made an important contribution to research knowledge in the intersection of writing mathematically and metacognition. Specifically, the study's further analysis of students' descriptions of their thinking during problem solving using a metacognitive framework (see also Pugalee, 2001) has lent support to the premise that writing can be an effective tool in supporting metacognitive behaviours and thus it can be associated with better success in solving problem-solving tasks.

The broader research study

The findings of a classroom-based investigation of the set research question naturally depend on students' prior learning experiences with proof. Thus it is necessary for me to offer sufficient background information about the context where the research was conducted so as to establish, for example, that the standard of proof used by myself as the researcher in analysing students' perceived proofs was indeed appropriate from a pedagogical standpoint. This necessity derives also from the definition of proof used in the research (Stylianides, 2007), which describes proof not only as a mathematically sound argument but also as an argument that uses resources that are known or conceptually accessible to the classroom community where the argument was developed.

The investigation reported in the article was part of a classroom-based *design experiment* (see, e.g. Cobb *et al.*, 2003) that aimed to engineer classroom instruction to support secondary students' learning of proof and to theorise the emerging relationship between learning and instruction. The study, which followed up on a university-based design experiment in the USA that had similar aims (e.g. Stylianides & Stylianides, 2009a,b, 2014), was carried out for 2 years in two Year 10 (and then Year 11) classes (14–15-year-olds and then 15–16-year-olds) in a state school in England. All 61 students from the two highest attaining Year 10 classes in the school (out of a total of seven classes) and their teachers participated in the research over the 2-year period (each teacher taught the same class in Years 10 and 11). The focus on high-attaining students was motivated by the findings of a prior large-scale longitudinal study in England (Küchemann & Hoyles, 2001–03) that showed (1) weak understanding of proof among a national sample of high-attaining Year 8–10 students and (2) modest (if any) improvements in students' understanding from Year 8 to Year 10. These findings raised concerns about English high-attaining secondary students' learning of proof and suggested an even more pessimistic prospect of learning proof for lower-attaining or younger students.

The study involved the design, implementation, and analysis of six lesson sequences related to proof (ranging from one to five 45-minute periods each), which

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

were taught by the two teachers at different points in time during the 2-year period. Of relevance to this article are the first two lesson sequences: in lesson sequence 1, the students were introduced to the concept of proof; lesson sequence 2 capitalised on the previous lesson sequence and engaged the students in proof constructions that formed the data for this article. In what follows I outline students' introduction to proof in lesson sequence 1 so as to place in context the students' proof constructions that I analyse in this article, and I discuss the standard of proof against which students' perceived proofs were judged.

Lesson sequence 1 lasted two 45-minute periods and was taught 1 month into Year 10 (i.e., in October). The lesson sequence was an adapted version of the instructional engineering discussed in Stylianides and Stylianides (2009b, 2014) and had two main goals: (1) to help students begin to recognise the limitations of empirical arguments as methods for validating mathematical generalisations and see an 'intellectual need' (Harel, 1998) to learn about more secure validation methods (i.e. proofs); and (2) to engage students in a discussion about what counts as a proof, including a list of five criteria for deciding whether a mathematical argument meets the standard of proof. The criteria were as follows (this is an excerpt from a PowerPoint slide used in lesson sequence 1).

An argument that counts as *proof* [in our class] should satisfy the following criteria:

1 It can be used to convince not only myself or a friend but also a *sceptic*.

- It should not require someone to make a leap of faith (e.g. 'This is how it is' or 'You need to believe me that this [pattern] will go on forever.')

2 It should help someone *understand why* a statement is true (e.g. why a pattern works the way it does).
3 It should use *ideas that our class knows already or is able to understand* (e.g. equations, pictures, diagrams).
4 It should contain *no errors* (e.g. in calculations).
5 It should be *clearly presented.*

These criteria, phrased to be suitable for secondary students, are consistent with Stylianides' (2007) definition of proof I presented earlier, thus ensuring coherence with the researcher's perspective on proof. Criteria 1 and 4 correspond to the requirement in the definition for valid modes of argumentation; criterion 5 to the requirement for appropriate modes of argument representation; and criterion 3 to the requirement that all components of a mathematical argument (set of accepted statements, modes of argumentation and modes of argument representation) be readily accepted, known to or within the conceptual reach of the class. Furthermore, criteria 1 and 2 reflect, respectively, two important functions that arguments and proofs served in the two classes: to promote conviction (e.g. Mason, 1982) and understanding (e.g. Hanna, 1995). These functions are in line with a broader framework of engaging students in mathematics as a sense-making activity whereby knowledge is established through reason and mathematical argument rather than by appeal to the authority of the teacher or the textbook (e.g. Ball & Bass, 2003; Harel & Sowder, 2007).

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Overall, students' engagement with proof in the two classes sought, by instructional design, to be consistent with Lampert's (1992) notion of 'authentic mathematics'. According to Lampert (1992), '[c]lassroom discourse in "authentic mathematics" has to bounce back and forth between being authentic (that is, meaningful and important) to the immediate participants and being authentic in its reflection of a wider mathematical culture [where proof has a pivotal role]' (p. 310). In this sense, communication practices such as 'proof transcription' (cf. Dimmel & Herbst, 2017) in presenting proofs in front of an audience were unwelcomed in the discourse of the two classes.

Methods

Data

The data for the article derived from lesson sequence 2, which lasted three 45-minute periods in one class and two such periods in the other class (the pace of implementation of a lesson sequence was up to each teacher and this resulted in variations in the duration of some lesson sequences in the two classes). Lesson sequence 2 was implemented 1.5 months after lesson sequence 1 and had two main goals: (1) to help students further understand the criteria for a proof introduced during lesson sequence 1 and (2) to offer to students opportunities to apply these criteria in three proving tasks. The three tasks (Figure 1) were of varying levels of difficulty, but they were all considered by the teachers to be appropriate for their students following their recent introduction to proof. Furthermore, the three tasks were mathematically similar: each of them involved making and proving a generalisation by reference to an underlying mathematical structure. The mathematical similarity of the tasks ensured some uniformity in the data and respective analysis to address the research question. The choice of this kind of task was partly motivated by their abundance in the secondary school curriculum (at least in England) and their mathematical importance, for embedded in them is the notion of 'structural generalisation' (e.g. Bills & Rowland, 1999).

Lesson sequence 2 started with a review of the five criteria for a proof (cf. the Background section), followed by the teacher launching Task 1. There was individual or small group work on the task, during which the teacher asked the students to write down their 'best' arguments with reference to the five criteria for a proof. By specifying to the students the frame of reference against which their arguments would be evaluated (i.e. the five criteria for a proof), the teacher (following the lesson plan) clarified essentially to the students the issue of *audience* for their arguments. This was an important clarification not only from a pedagogical standpoint but also from a research standpoint, for prior research showed that students' sense of audience might influence their proof-related work (e.g. Healy & Hoyles, 2000). The students were given ample time to engage in proving activity related to the set task and ultimately write down the final products of this activity. They were free to work in pairs or larger groups during the proving activity, but they were instructed specifically to write down their arguments individually. However, a few of them ended up writing their arguments in pairs. There was then a whole-class discussion during which the teacher

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

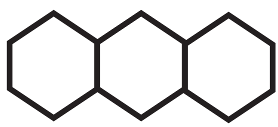
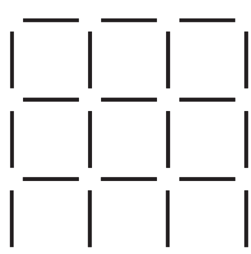
Task 1	**The Toothpicks Problem**  1. How many toothpicks make up this 1-by-4 rectangle? 2. How many toothpicks make up a 1-by-60 rectangle? **Prove your answer**. 3. Can you find an expression that would give the number of toothpicks that make up an 1-by-N rectangle? **Prove your answer**.
Task 2	**The Hexagons Problem** Mark has some hexagons. The sides of the hexagons are all the same length and equal to **1cm**. Mark puts hexagons in rows to create trains of different sizes. Here's Train 3, which is made up of 3 hexagons in a row:  1. What is the perimeter of Train 3? 2. What is the perimeter of Train 100, which is made up of 100 hexagons in a row? **Prove your answer**. 3. Can you find an expression that would give the perimeter of Train N, which is made up of N hexagons in a row? **Prove your answer**.
Task 3	**The New Toothpicks Problem**  1. How many toothpicks make up the 3-by-3 square above? 2. How many toothpicks make up a 60-by-60 square? **Prove your answer**. 3. Can you find an expression that would give the number of toothpicks that make up an N-by-N square? **Prove your answer**.

Figure 1. The three proving tasks

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

made an open call for students to come up on the board and present their proofs, with the presented arguments being considered against the five criteria for a proof. The use of the criteria as a frame of reference for evaluating the quality of the oral arguments in the public domain supported a coherent sense of audience for students' written and oral arguments, thus creating a common ground for the comparison of the two kinds of argument in this research. A similar procedure as in Task 1 was followed for the implementation of the other two tasks. The lesson sequence was videotaped using a camera that was placed at the back of the room, and the video records were fully transcribed.

The specific data for the article are the transcripts of the oral presentations of 17 students who presented their arguments in front of the class and copies of the corresponding written arguments that the same students had produced during their individual or small group work prior to the presentations. These were all the students who presented their arguments, which they perceived to be proofs, to the rest of the class for any of the three tasks. Thirteen students wrote and presented their arguments individually; two students wrote and presented their arguments together; and two students wrote their arguments individually but presented them together. Overall there were 16 distinct written arguments (one of which was co-authored by two students) and 15 distinct oral arguments (two of which were each presented jointly by a pair of students). Also 13 students (or pairs of students) presented arguments for only one task while two students each presented arguments for two tasks. The distribution of student presenters across the three tasks was 10 students for the Task 1, four for Task 2, and three for Task 3.

The data collection and broader study were scrutinised according to the procedures for ethical approval of the University of Oxford (my home institution at the time of the data collection), which met the British Educational Research Association and British Psychological Society standards. School, teacher and student participation in the study was voluntary and was solicited through consent forms to the school's head teacher, the school's head of mathematics, the two teachers, the students' parents or guardians and the students themselves.

Analysis

A research assistant and I coded independently all written and oral (transcribed) arguments of the 17 students. We compared codes and discussed disagreements, reaching a consensus code as needed. Examples of all codes used in the analysis will be given in the Results section.

The unit of analysis was a student's entire (written or oral) argument rather than selected parts thereof. This choice of a unit of analysis was meant to respect the integrity of students' arguments and avoid any selection bias.

In conducting the analysis, first we used the coding scheme described in Stylianides and Stylianides (2009a) to classify each written or oral argument into one of the following five categories (M1–M5) ordered according to the degree to which the argument approximated the standard of proof (M1 indicated a proof). Two main reasons for the choice of this coding scheme were (1) its alignment with Stylianides' (2007) definition of proof that had underpinned the research reported herein and (2) its way

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

of operationalising this definition for use in coding students' proof constructions. Codes M1–M4 refer to (genuine) mathematical arguments and can be clustered into two broader categories: *strong* (M1 and M2) and *weak* (M3 and M4).

- Strong mathematical arguments:
  - code M1—a proof;
  - code M2—a valid general argument but not a proof.
- Weak mathematical arguments:
  - code M3—an unsuccessful attempt for a valid general argument;
  - code M4—an empirical argument.
- Non-mathematical arguments:
  - code M5—a non-genuine argument.

All three tasks involved proving a generalisation, and so the definitions of the codes were tailored to classifying arguments that aimed to establish the truth of a general claim. The definition of code M1 was consistent with Stylianides' (2007) definition of proof, which had underpinned the criteria for a proof used in the two classes. Specifically, code M1 was used for arguments that were *general* (i.e. they covered all the cases in the domain of the respective generalisation), used *valid* modes of argumentation (i.e. they offered conclusive evidence for the truth of the generalisation) and were accessible to the students in the class (i.e. they used statements that were readily acceptable by the class as well as modes of argumentation and modes of argument representation that were known to the students or were judged by the teachers to be within the students' conceptual reach at the particular point in time). Code M2 was used for arguments that approximated but not quite met the standard of proof, because, for example, of a missing or an inadequate explanation of an assertion that could not have been considered readily acceptable by the class. Code M3 was used for arguments that reflected an attempt to justify the respective generalisation for all the cases in its domain, but the arguments were either incomplete or used *invalid* modes of argumentation (i.e. they included a logical flaw). Code M4 was used for arguments that verified the truth of the generalisation only in a proper subset of all the cases in its domain but nevertheless concluded that the generalisation was true for all cases. Code M5 was used for non-genuine arguments, i.e. responses to the proving tasks that showed minimal engagement, were irrelevant to what was being asked, or were potentially relevant but the relevance was not made evident to the coder.

The research question was addressed primarily by comparison of the codes given to the written and oral versions of students' arguments for the generalisation in each task. The varying (albeit appropriate) levels of difficulty of the three tasks promised a good distribution of students' written arguments across the categories of strong and weak mathematical arguments, thus promising further an appropriate setting to explore the research question. Did strong written arguments retain or (if relevant) improve their status during their oral presentations, or were they downgraded to weak

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

arguments? Conversely: Did weak written arguments improve their status thus being upgraded to strong arguments, or did they retain their weak status?

In addition to the arguments' degree of proximity to a proof, we coded the following two factors so as to offer more insight into the outcomes of the comparison:

- *Who* wrote or orally presented each argument: an *individual student* or a *pair of students*; and
- The kind of *verbal input* from the teacher or the rest of the class during the oral presentation of the argument: *no input*; *some but not substantial input* (i.e. input that simply reiterated or briefly clarified a point mentioned by the student without an apparent influence on the quality of the presented argument); or *substantial input* (i.e. input that influenced the presented argument and possibly altered its degree of proximity to a proof).

Other factors

In what follows I discuss four other factors that could have influenced potentially the results of the comparison between students' written and oral arguments for the same claims. Regarding the first factor, I argue that an influence was possible but nevertheless inevitable given the nature of classroom work. Regarding the other three factors, I argue that, despite their potential to do so, most likely they did not influence the results.

The first factor concerns the temporal sequencing of students' arguments: first written, then oral. It is possible that a student's experience of producing a written proof helped the student build familiarity with the task and underlying concepts, thus placing the student in a better position later on to orally present an argument that approximated the standard of proof. While from a research standpoint the lack of control over the temporal sequencing of students' arguments might be viewed as a methodological limitation, the particular sequencing (first written, then oral) is legitimate from a pedagogical standpoint. Specifically, that sequencing was an inevitable implication of how lessons were organised in the two classes during the implementation of high-level tasks (proving tasks being a case in point). According to Stein *et al.* (2008), lessons involving high-level tasks in reform-oriented classes often start with the teacher launching one such task; they continue with students working on the task individually or in small groups, with the expectation being usually that students will keep a written record of their work; and they conclude with a whole-class discussion where students' solutions are orally presented and discussed. I revisit this factor at the end of the article.

Given the particular temporal sequencing, the comparison between students' arguments is essentially between the 'written mode' and the 'oral mode in the context of a class presentation following a written proof construction'. Thus it is reasonable to ask whether and to what extent students' oral arguments were influenced by factors pertaining to the class presentations (cf. the second and third factors discussed below) or the prior written work (cf. the fourth factor).

The second factor was whether a student's oral presentation of an argument for a proving task influenced subsequent student presentations for the same task. If that happened, the overall quality of oral arguments might have been enhanced. Yet I

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

found no indication of such an influence: students' oral presentations were rather distinct from one another; students' orally presented arguments matched closely their written arguments (an indication of this was that the oral presentations were based on the same diagrams as in students' written work); and students looked at or referred to their own written work during their oral presentations.

The third factor was whether the teacher or the rest of the class offered any *non-verbal* input to students' oral presentations that influenced the course or content of those presentations thus influencing also the degree to which the respective oral arguments approximated the standard of proof. Indeed, non-verbal ways of expression, notably gestures, can reveal implicit knowledge and provide the imagery of language (McNeill, 1992; Roth, 2001). Data constraints did not allow me to examine systematically the possible influence of gestures: the lessons were videotaped by only one camera at the back of the room that focused primarily on the presenter and was thus inadequate to capture the gestures (if any) of the teacher and of all the students (approximately 30) in each class. However, even if gestures played a role in students' presentations, their particular influence on the quality of the presented arguments could be both negative and positive depending on whose gestures appealed to the presenter at the time and the presenter's interpretation of those gestures (which might have differed from their intended meaning). Also, it is possible that any gestures with a substantial influence on an oral argument would be accompanied with a relevant utterance, which would then be captured by the analysis of *verbal input* as explained in the previous section.

The fourth factor was whether the teacher offered any substantial input during students' written work in small groups (the teacher's input during students' oral presentations was coded as described earlier). According to the plan that I had agreed with both teachers prior to the lessons, during small group work they would ask students probing questions, but they would not directly influence students' argument constructions. Indeed, my observations during the lessons and the videotapes of the lessons (reviewed afterwards) offered no indication of deviation from the agreed plan. But even if the teachers had offered substantial input during small group work, the result of that input would have been better written arguments and, presumably, better oral presentations of those arguments, too. In other words, there would likely be limited if any impact on the *comparison* between the degrees to which students' written and oral arguments approximated the standard of proof, which is the issue examined in this article.

Results

General trends

The results are summarised in Figures 2a–c and show the degrees to which every written argument and its respective oral argument approximated the standard of proof for the following groups of students: (1) those who wrote and presented their arguments individually, with no verbal input from the teacher or the rest of the class during the presentation of their oral arguments (N = 9, Figure 2a); (2) those who wrote their arguments individually or in pairs and presented the

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

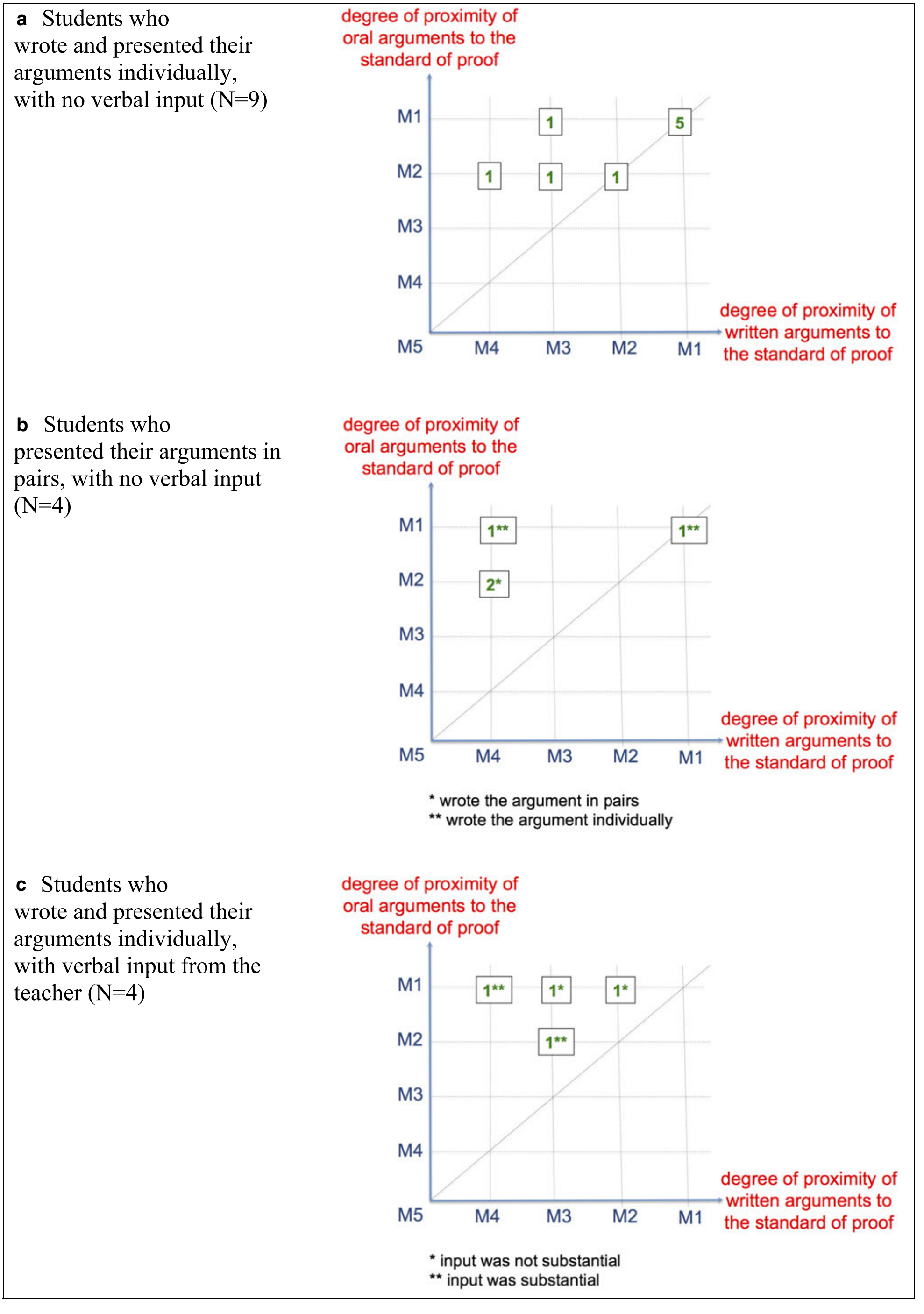


Figure 2. Summary of results

arguments in pairs, again with no verbal input (N = 4, Figure 2b); and (3) those who wrote and presented their arguments individually, with some verbal input from the teacher during the presentation of their oral arguments (N = 4,

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Figure 2c). With regard to the last group, in two cases only was the teacher's verbal input considered to be substantial and thus potentially influential of the quality of the presented arguments.

No arguments were coded as non-genuine (code M5), which was unsurprising given the generally high mathematical competence of the students and their prior learning experiences with proof. The distribution of students' *written* arguments across the categories of strong and weak mathematical arguments (codes M1–M2 and M3–M4 respectively) was rather uniform: eight students wrote strong arguments and nine wrote weak arguments. This distribution created an appropriate setting to explore the research question, as it allowed investigation of whether the arguments in each category retained their status or crossed categories. Consideration also of students' *oral* arguments revealed a clear pattern: all strong written arguments retained their status during their oral presentations, while all weak written arguments were upgraded to strong arguments during their oral presentations. In particular, all of the orally presented arguments had the same or better quality than their written counterparts. This pattern applied across the board, i.e. it was not restricted to the two cases where the teacher offered substantial verbal input during the students' oral presentations.

Exemplification and further insights

Next I present some student arguments for the three proving tasks to exemplify the general trends and offer further insights into the relationship between students' written and orally presented arguments. All student names are pseudonyms.

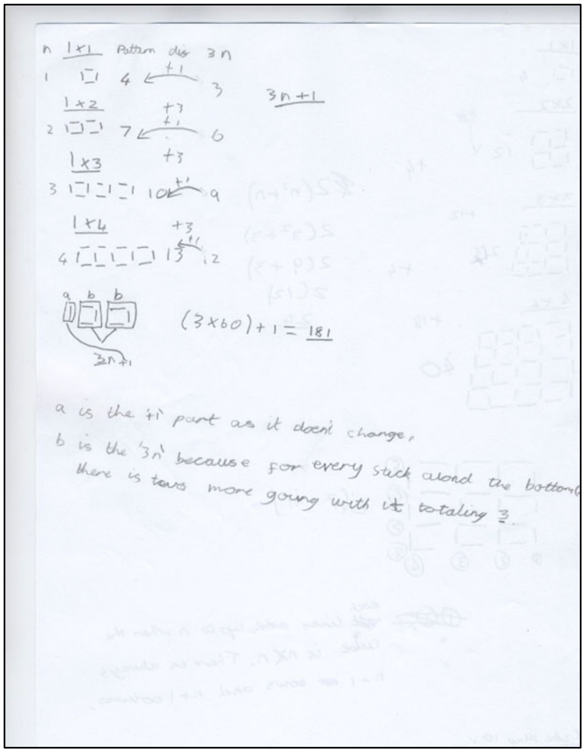
All six students who wrote proofs (M1) ended up presenting proofs. There was a close correspondence between students' written and orally presented proofs as illustrated by Larry's work for Task 1 (Figures 3a,b). Larry's oral presentation was uninterrupted by the teacher or the rest of the class, and so his argument was coded as having received 'no input'.

Mac's work for Task 3 (Figures 4a,b) illustrates the upgrade of an M2 written argument to a proof (M1) during its oral presentation. Mac's written argument did not quite meet the standard of a proof, as it offered no explicit explanation about the derivation of different algebraic expressions (n, $n + 1$, $2n + 1$, etc.); Mac offered these explanations during his oral presentation of the argument. Indeed, in their oral presentations students tended to fill in gaps of their written arguments, or articulate explanations that were absent or implicit in the written arguments, thus elevating their status. This is illustrated further by the work of Blaze for Task 2 (Figures 5a,b); he wrote an M3 argument that was elevated to a proof during its oral presentation.

The oral presentation of Mac's argument (Figure 4b) illustrates also what counted as 'some but not substantial input': The teacher simply clarified briefly a point mentioned by Mac during his presentation, and it seems plausible that the teacher's comment did not influence the quality of the presented argument. The oral presentation of Sylvia's argument (Figure 6b) illustrates what counted as 'substantial input': The teacher made a comment at a point when Silvia paused and seemed to have difficulty articulating the general case in the context of the specific diagram she had drawn on

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

a



b

Larry is up on the board and says the following:

"The formula is 3n+1. Say for 1-by-3, you've got a pattern like that. [He draws a 1-by-3 rectangle on the board similar to the one that appears in his paper; see last diagram in Figure 3a.] And I split it up because this bit [he shows the first vertical line of the rectangle] is the '+1' part of the equation and then these being separate... [he puts into a square outline the three sticks that made up a square – the top, bottom, and right one – like in his paper, Figure 3a] and then I've put this bit 'a' and these bits all 'b.' And then I've written 'a' is the '+1' part as it doesn't change and 'b' is the '3n,' because every stick along the bottom, which is n, there is two more going with it, therefore here is 2, 2 more 3… [inaudible words]. So it is 3 times of how many are in the bottom plus the 'a'."

Figure 3. (a) Larry's written argument for Task 1 (code M1); (b) Larry's orally presented argument (transcribed) for Task 1 (code M1)

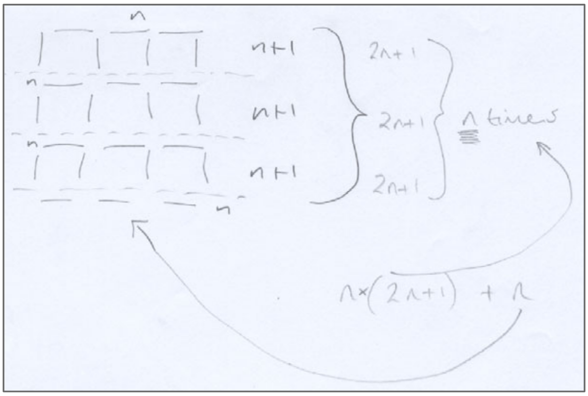
the board, and it seems plausible that Silvia ended up presenting an argument that better approximated the standard of proof than she would have presented without the teacher's input.

Discussion

All of the oral arguments that the secondary students in this study presented in front of their class and perceived to be proofs approximated the standard of proof to the

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

a



b

Mac is up on the broad to present his proof. He works on a 3-by-3 square that is already on the board and is essentially the same as the diagram he had in his paper (see Figure 4a). Mac says the following:

"In this example, this group of three here [shows the top horizontal line of toothpicks] could be called 'n' [writes 'n' next to the line]; the same for this group here [shows the second horizontal line of toothpicks and writes 'n' next to the line]; this group [shows the third horizontal line of toothpicks and writes 'n' next to it]; and this group [shows the bottom horizontal line of toothpicks and writes 'n' next to it]. Now along there [shows the top row of squares] it has four [vertical toothpicks; shows the toothpicks], which in this case here is 'n+1' [writes 'n+1' on the right corner of the diagram]. So combine that with the 'n' and the 'n+1,' we get '2n+1' [writes '2n+1' on the board]. Now this is duplicated three times down here [shows the three rows of squares in the diagram and writes '2n+1' two more times next to each of the remaining rows]."

The teacher interrupts Mac to say: "…because this is a 3-by-3 square."

Mac repeats what the teacher said and continues:

"Assume there is a formula, we've got '2n+1' and that would be multiplied by how many times we've gone down with 'n' [shows the rows of squares]. This end line [shows the bottom line of toothpicks] … [inaudible] … would just be added on [writes on the board 'n x (2n+1) + n']."

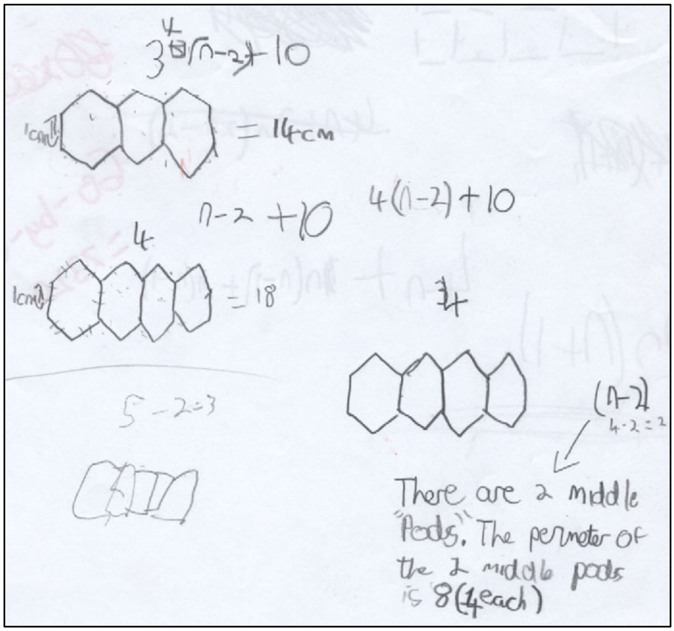
Figure 4. (a) Mac's written argument for Task 3 (code M2); (b) Mac's orally presented argument (transcribed) for Task 3 (code M1)

same or higher degree than the corresponding written arguments that the students had produced previously for the same claims. These findings suggest that, in a classroom context at the secondary school level, the use of the oral mode of representation is more likely, compared with the written mode, to be associated with the construction of student arguments (perceived proofs) that meet the standard of proof.

As discussed in the Methods section, the temporal sequencing of students' arguments might have played a role in the findings: it is possible that students' work of first

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

a



b

Blaze is up on the board and draws a train comprising three hexagons. He says the following:

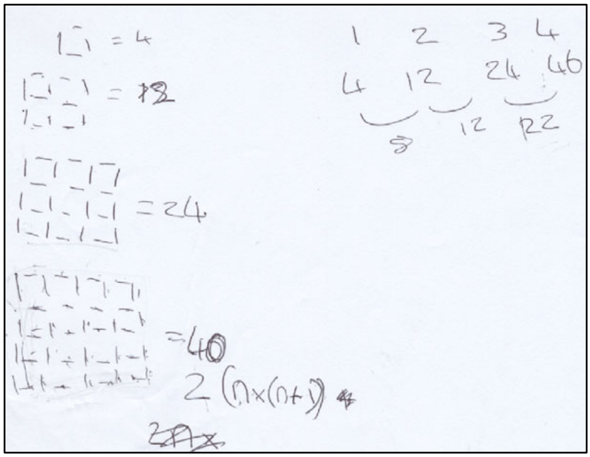
"No matter how long the chain [train] is, you're always going to have these two end pods [he draws a circle around the first and last hexagons in the train]. On each end pod you've got five sides so you get '+10,' cause they're 10 [he indicates that he counted the sides of the first and last hexagons that contribute to the perimeter of the train]. And this is a train of three, so to get the one middle pod you have to do 'n – 2,' so on each middle pod there is 4 sides so you times that by 4. [As he talks he writes '4(n – 2) + 10' on the board.]"

Figure 5. (a) Blaze's written argument for Task 2 (code M3); (b) Blaze's orally presented argument (transcribed) for Task 2 (code M1)

writing their perceived proofs before orally presenting them in front of their class had afforded them familiarity with the task and underlying concepts thus positioning them to present better oral arguments later on. According to Pugalee (2001, 2004), writing can support metacognitive behaviour including a level of reflection that promotes students' attention to their thinking about mathematical processes, which in turn can support students' selection of appropriate strategies while solving mathematical problems. Even if temporal sequencing did play a role, however, there might not be a compelling reason to try to design a classroom-based study to investigate the effect of 'first written, then oral' versus 'first oral, then written' on the quality of students' written and respective oral arguments for the same claims. This is because one of these conditions (i.e. 'first oral, then written') would go against the common classroom practice during the implementation of high-level tasks whereby students' written

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

a



b

Silvia is up on the board to present her proof. She works on a 3-by-3 square that is already on the board and is the same as one of the diagrams she had in her paper (see Figure 6a). She says the following:

"In the diagram, you have certain vertical lines [she shows the vertical lines of the 3-by-3 square] and certain horizontal lines [she shows the horizontal lines of the 3-by-3 square]. And the number of… squares you have is 3 [she shows the three squares along the bottom row of the 3-by-3 square], so n is 3. And… one second [she has a look at her notes] … you have… if we look at… the horizontal lines, you have 4 times 3… but n [she shows all horizontal lines in the 3-by-3 square, beginning from the bottom of the square]."

Silvia pauses and the teacher comments: "If that [showing the bottom horizontal line] was n…?"

Silvia continues the teacher's sentence: "…that's n plus 1 [she shows the collection of all horizontal lines]. So [for the total number of toothpicks in the horizontal lines] you've got n times n plus 1 [she writes on the board 'n x n + 1'] and it's same for the vertical lines… So… if you do that [she puts into brackets 'n+1' in the previous expression and then the whole expression 'n x (n+1)' into brackets] and have 2 there [she completes the expression to read '2(n x (n+1))'], so you do it twice and that's all."

Figure 6. (a) Silvia's written argument for Task 3 (code M4); (b) Silvia's orally presented argument (transcribed) for Task 3 (code M1)

work precedes the oral presentations of students' produced solutions (Stein *et al.*, 2008). Thus, the findings reported herein might be viewed as documenting a specific variation of the relation between students' written and oral arguments (perceived proofs) under the particular variation of the temporal sequencing typically found in actual classroom settings. Having said that, it would be interesting, and not incompatible with common classroom practice, to explore a more complex structure of

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

temporal sequencing whereby students first write their arguments (perceived proofs), then they orally present their arguments, and finally they write their arguments again (i.e. 'first written, then oral, then written again'). This exploration would offer some insight into whether the oral presentation of an argument, which the present study suggests might be influenced by the prior work of writing that argument, can influence in turn the subsequent re-writing of the argument for the same claim. If it turned out that the written and re-written arguments were identical or approximated the standard of proof to the same degree, this finding would create some doubt over the role played by temporal sequencing in the quality of students' written and oral arguments for the same claims in the context of classroom work.

A possible setting to explore the particular variation of temporal sequencing 'first oral, then written' would be outside of the classroom, by adaptation of the task-based interviews that are conducted usually with university students in proof construction studies. For example, consider again the study of Mejía-Ramos and Weber (in press) with 73 mathematics majors that I discussed in the Background section. These researchers asked the student participants to think aloud while proving seven calculus theorems, and then write their proofs up as if they were doing so for an examination. The data yielded by these task-based interviews allowed the clear identification of the written arguments that the students perceived to be proofs but not the corresponding oral arguments. If the students were asked to orally present the final argument (perceived proof) emerging from their proving activity around a set task before they wrote that argument up, there would be a record of both kind of arguments for comparison. The findings from such a comparison would enhance the field's understanding of the relationship between students' oral and written arguments (perceived proofs) for the same claims, albeit under a different variation of temporal sequencing, with a different student population, and in a substantially different setting than in the study I reported in this article.

The possible role played by temporal sequencing left aside, what else might explain the generally lower degree of proximity of students' written arguments to the standard of proof compared with their respective oral arguments? A possible reason might be sought in the difficulty coming from writing mathematics. According to Vygotsky (1987; cited in Pugalee, 2004), writing involves 'deliberate analytical action on the part of the writer requiring the writer to maximally compact inner speech so that it is fully understandable; thus, written words require a deliberate structuring of a web of meaning' (Pugalee, 2004, p. 28). Proof writing must place increased demands on the writer for deliberate analytical action and structuring of a web of meaning owing to the expectations from a proof to construct a logical argument from accepted statements and offer clarity about what statements are considered to belong to the set of accepted statements and how premises are linked with conclusions through logical deductions. Verbal expression might relieve the speaker from the requirement 'to maximally compact inner speech' thus allowing the speaker more freedom to convey the intended meaning. Indeed, in their presentations of their oral arguments, which were more extended in length than the respective written arguments, students tended to fill in gaps of their written arguments, or articulate explanations that were absent or implicit in these arguments, thus elevating their status. Students' prior learning experiences with proof are pertinent here: If students' presentations were mere mark-for-

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

mark reproductions of their written proofs (cf. Dimmel & Herbst, 2017), thus departing from the perspective of proof as a vehicle to mathematical sense-making that had underpinned instruction in my design experiment, the students' presentations would be unlikely to fill in any gaps of their written proofs.

As I noted at the beginning of the article, prior research on secondary students' argument constructions tended to use survey methods and only consider students presenting their perceived proofs in written form. Based on students' written arguments, this research has painted a bleak picture of secondary students' ability to construct arguments that meet the standard of proof. The findings reported in this article suggest that the limited use of observation methods and the lack of consideration of students presenting their perceived proofs orally —in tandem with students' written proofs for the same claims—is a serious threat to the validity of research findings in this area. Indeed, if a study had analysed students' written arguments only (as in survey research), it would have reported a less favourable picture of the potential of students' constructed proofs than another study that would focus only on students' oral arguments (as in observational research). Also, by considering only one mode of representation and ignoring the other, each study individually would have reported an incomplete picture of students' constructed proofs, for apparently it matters whether students present their perceived proofs orally or in writing. Given that in England and other countries students' mathematical knowledge is assessed primarily, if not exclusively, through students' response to tasks in written tests, the findings reported in this article also have implications for practice. Specifically, the findings suggest that the lack of consideration by current assessment practices of students presenting their solutions to proving or other kind of mathematics tasks orally—in tandem with students' written solutions to the same tasks—might be yielding an inaccurate picture of students' mathematical potential.

Overall, the findings reported herein highlight, from a research standpoint, the importance of multiple research methods in studying complex cognitive abilities, secondary students' proof constructions being a case in point. Another complex cognitive ability in the area of proof that I believe would benefit from a study of this kind is secondary students' evaluations of the extent to which given (researcher generated) arguments meet the standard of proof. Despite the rather extensive body of research on secondary students' argument evaluations (e.g. Küchemann & Hoyles, 2001–03; Hoyles & Küchemann, 2002; Yu *et al.*, 2004), there is evidence to suggest that a number of factors, often uncontrolled for in relevant studies, influence the nature of secondary students' argument evaluations and thus the validity of researchers' conclusions based on students' work. For example, there is evidence to suggest that it is easier for secondary students to identify invalid arguments as invalid than it is for them to identify valid arguments as valid (Reiss *et al.*, 2002); thus a study would likely report a better picture of secondary students' potential to accurately evaluate given arguments if it asked students to evaluate more invalid arguments and fewer valid arguments. There is also evidence to suggest that secondary students evaluate given arguments from different perspectives, such as what would satisfy them personally or what would satisfy their teacher, with the latter perspective triggering evaluations that better align with the conventional meaning of proof (Healy & Hoyles, 2000); thus a

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

study would likely report a better picture of secondary students' potential to accurately evaluate given arguments if it asked students to do their evaluations based on what they thought would count as a proof for their teacher than for themselves. Similar to secondary students' proof constructions and the role that a written versus an oral mode of representation might play in the quality of these constructions, the aforementioned pieces of evidence spotlight the study of secondary students' argument evaluations as a methodologically complex territory and indicate the need for the use of multiple research methods to better understand the role that different factors play (or can play) in the quality of these evaluations. Research with university students has made more progress than research with secondary students in understanding what factors impact on such evaluations (see, e.g. Selden & Selden, 2003; Alcock & Weber, 2005; Stylianides & Stylianides, 2009a; Weber, 2010).

To conclude, in this article I have exemplified the threats to the validity of research findings arising from a singular methodological approach to the study of a multifaceted cognitive ability related to the fundamental concept of proof in mathematics at the secondary school level. Unless we, as a field, achieve an adequate understanding of the complex network of factors influencing our findings in the study of this and other abilities of a similar kind, it will be difficult to develop a broad-based methodological approach that will generate an accurate picture of students' potential with respect to these abilities.

Acknowledgements

The research reported in this article was supported by a grant from the Economic and Social Research Council in England (grant reference RES-000-22-2536). The opinions expressed in the article are mine and do not necessarily reflect the position, policy or endorsement of the Council. Some parts of earlier versions of the article were presented at the 2013 Annual Meeting of the American Educational Research Association (San Francisco, USA) and the 9th Congress of the European Society for Research in Mathematics Education (Prague, Czech Republic, 2015). I wish to thank Keith Weber and two anonymous reviewers for useful feedback on earlier versions of the article.

References

Alcock, L. & Inglis, M. (2008) Doctoral students' use of examples in evaluating and proving conjectures, *Educational Studies in Mathematics*, 69(2), 111–129. https://doi.org/10.1007/s10649-008-9149-x

Alcock, L. & Simpson, A. P. (2004) Convergence of sequences and series: Interactions between visual reasoning and the learner's beliefs about their own role, *Educational Studies in Mathematics*, 57(1), 1–32. https://doi.org/10.1023/b:educ.0000047051.07646.92

Alcock, L. & Simpson, A. P. (2005) Convergence of sequences and series 2: Interactions between nonvisual reasoning and the learners' beliefs about their role, *Educational Studies in Mathematics*, 58(1), 77–100. https://doi.org/10.1007/s10649-005-2813-5

Alcock, L. & Weber, K. (2005) Proof validation in real analysis: Inferring and checking warrants, *The Journal of Mathematical Behavior*, 24(2), 125–134. https://doi.org/10.1016/j.jmathb.2005.03.003

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Artemeva, N. & Fox, J. (2011) The writing's on the board: The global and the local in teaching undergraduate mathematics through chalk talk, *Written Communication*, 28(4), 345–379. https://doi.org/10.1177/0741088311419630

Balacheff, N. (2002) The researcher epistemology: A deadlock for educational research on proof, in: F. L. Lin (Ed.) *Proceedings of the 2002 International Conference on Mathematics: Understanding proving and proving to understand* (Taipei, NSC and NTNU).

Ball, D. L. & Bass, H. (2003) Making mathematics reasonable in school, in: J. Kilpatrick, W. G. Martin & D. Schifter (Eds) *A research companion to Principles and Standards for School Mathematics* (Reston, VA, National Council of Teachers of Mathematics).

Bills, L. & Rowland, T. (1999) Examples, generalisation and proof, in: L. Brown (Ed.) *Making meaning in mathematics: Advances in mathematics education*, (vol. 1) (York, UK, QED).

Boero, P., Garuti, R. & Mariotti, M. A. (1996) Some dynamic mental processes underlying producing and proving conjectures, in: L. Puig & A. Gutierrez (Eds) *Proceedings of the 20th Conference of the International Group for the Psychology of Mathematics Education* (Valencia, Valencia University).

Burton, L. & Morgan, C. (2000) Mathematicians writing, *Journal for Research in Mathematics Education*, 31(4), 429–453. https://doi.org/10.2307/749652

Cobb, P., Confrey, J., diSessa, A., Lehrer, R. & Schauble, L. (2003) Design experiments in educational research, *Educational Researcher*, 32(1), 9–13. https://doi.org/10.3102/0013189x032001009

Department for Education. (2013) *Mathematics: Programmes of study: Key Stages 1–2* (National Curriculum in England). Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/239129/PRIMARY_national_curriculum_-_Mathematics.pdf.

Dimmel, J. K. (2015) *How do teachers expect students to represent mathematical work? A study of teachers' recognition of routine ways that proofs are presented and checked in high school geometry*. Unpublished doctoral dissertation. (Ann Arbor, University of Michigan).

Dimmel, J. K. & Herbst, P. G. (2017) Secondary mathematics teachers' attitudes toward alternative communication practices when doing proofs in geometry, *Teaching and Teacher Education*, 68, 151–160. https://doi.org/10.1016/j.tate.2017.08.018

González, G. & Herbst, P. (2013) An oral proof in a geometry class: How linguistic tools can help map the content of a proof, *Cognition and Instruction*, 31(3), 271–313. https://doi.org/10.1080/07370008.2013.799166

Greiffenhagen, C. (2014) The materiality of mathematics: Presenting mathematics at the blackboard, *The British Journal of Sociology*, 65(3), 502–528. https://doi.org/10.1111/1468-4446.12037

Griffiths, P. A. (2000) Mathematics at the turn of the millennium, *American Mathematical Monthly*, 107(1), 1–14. https://doi.org/10.1080/00029890.2000.12005154

Hanna, G. (1995) Challenges to the importance of proof, *For the Learning of Mathematics*, 15(3), 42–49.

Harel, G. (1998) Two dual assertions: The first on learning and the second on teaching (or vice versa), *The American Mathematical Monthly*, 105(6), 497–507. https://doi.org/10.1080/00029890.1998.12004918

Harel, G. & Sowder, L. (2007) Toward comprehensive perspectives on the learning and teaching of proof, in: F. K. Lester (Ed) *Second handbook of research on mathematics teaching and learning* (Greenwich, CT, Information Age).

Healy, L. & Hoyles, C. (2000) A study of proof conceptions in algebra, *Journal for Research in Mathematics Education*, 31(4), 396–428. https://doi.org/10.2307/749651

Healy, L. & Hoyles, C. (2001) Software tools for geometrical problem solving: Potentials and pitfalls, *International Journal of Computers for Mathematical Learning*, 6(3), 235–256. https://doi.org/10.1023/a:1013305627916

Hoyles, C. & Küchemann, D. (2002) Students' understanding of logical implication, *Educational Studies in Mathematics*, 51(3), 193–223. https://doi.org/10.1023/a:1023629608614

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Knuth, E. J. (2002) Teachers' conceptions of proof in the context of secondary school mathematics, *Journal of Mathematics Teacher Education*, 5(1), 61–88. https://doi.org/10.1023/a:1013838713648

Knuth, E. J., Choppin, J. & Bieda, K. (2009) Middle school students' productions of mathematical justification, in: M. Blanton, D. Stylianou & E. Knuth (Eds) *Teaching and learning proof across the grades: A K–16 perspective* (New York, Routledge).

Küchemann, D. & Hoyles, C. (2001–03) Longitudinal Proof Project. Technical reports for Year 8–10 surveys. Available online at: http://www.mathsmed.co.uk/ioe-proof/techreps.html

Lai, Y. & Weber, K. (2014) Factors mathematicians profess to consider when presenting pedagogical proofs, *Educational Studies in Mathematics*, 85(1), 93–108. https://doi.org/10.1007/s10649-013-9497-z

Lamon, S. J. (2001) Presenting and representing: From fractions to rational numbers, in: A. Cuoco (Ed.) *The roles of representation in school mathematics* (Reston, VA, National Council of Teachers of Mathematics).

Lampert, M. (1992) Practices and problems in teaching authentic mathematics, in: F. K. Oser, A. Dick & J. Patry (Eds) *Effective and responsible teaching: The new synthesis* (San Francisco, CA Jossey-Bass).

Mariotti, M. A. (2000) Introduction to proof: The mediation of a dynamic software environment, *Educational Studies in Mathematics*, 44(1–3), 25–53. https://doi.org/10.1023/a:1012733122556

Mariotti, M. A. (2006) Proof and proving in mathematics education, in: A. Gutiérrez & P. Boero (Eds) *Handbook of research on the PME: Past, present and future* (Rotterdam, Sense Publishers).

Mason, J. (with L. Burton & K. Stacey). (1982) *Thinking mathematically* (London, Addison-Wesley).

McNeill, D. (1992) *Hand and mind: What gestures reveal about thought* (Chicago, IL University of Chicago Press).

Mejía-Ramos, J. P. & Weber, K. (in press) Mathematics majors' diagram usage when writing proofs in calculus, *Journal for Research in Mathematics Education*.

Morris, A. K. (2009) Representations that enable children to engage in deductive reasoning, in: D. A. Stylianou, M. L. Blanton & E. J. Knuth (Eds) *Teaching and learning proof across the grades: A K-16 perspective* (New York, Routledge).

National Governors Association Center for Best Practices & Council of Chief State School Officers (NGA & CCSSO). (2010) *Common core state standards for mathematics* (Washington, DC, Author).

Presmeg, N. C. (2006) Research on visualization in learning and teaching mathematics, in: A. Gutierrez & P. Boero (Eds) *Handbook of research on the psychology of mathematics education* (Rotterdam, Sense Publishers).

Pugalee, D. K. (2001) Writing, mathematics, and metacognition: Looking for connections through students' work in mathematical problem solving, *School Science and Mathematics*, 101(5), 236–245. https://doi.org/10.1111/j.1949-8594.2001.tb18026.x

Pugalee, D. K. (2004) A comparison of verbal and written descriptions of students' problem solving processes, *Educational Studies in Mathematics*, 55(1), 27–47. https://doi.org/10.1023/b:educ.0000017666.11367.c7

Reid, D. (2005) The meaning of proof in mathematics education, in: M. Bosch (Ed.) *Proceedings of the 4th Conference of the European Society for Research in Mathematics Education*(Sant Feliu de Guixols, Spain, Universitat Ramon Llull & ERME).

Reiss, K., Hellmich, F. & Reiss, M. (2002) Reasoning and proof in geometry: Prerequisites of knowledge acquisition in secondary school students, in: A. D. Cockburn & E. Nardi (Eds) *Proceedings of the 26th Conference of the International Group for the Psychology of Mathematics Education* (University of East Anglia, School of Education and Professional Development, Norwich, England).

Roth, W. M. (2001) Gestures: Their role in teaching and learning, *Review of Educational Research*, 71(3), 365–392. https://doi.org/10.3102/00346543071003365

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Samkoff, A., Lai, Y. & Weber, K. (2012) On the different ways that mathematicians use diagrams in proof construction, *Research in Mathematics Education*, 14(1), 49–67. https://doi.org/10.1080/14794802.2012.657438

Schifter, D. (2009) Representation-based proof in the elementary grades, in: D. A. Stylianou, M. L. Blanton & E. J. Knuth (Eds) *Teaching and learning proof across the grades: A K-16 perspective* (New York, Routledge).

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. https://doi.org/10.2307/30034698

Senk, S. L. (1989) van Hiele levels and achievement in writing geometry proofs, *Journal for Research in Mathematics Education*, 20(3), 309–321. https://doi.org/10.2307/749519

Sfard, A. (2001) There is more to discourse than meets the ears: Looking at thinking as communicating to learn more about mathematical learning, *Educational Studies in Mathematics*, 46(1–3), 13–57. https://doi.org/10.1023/a:1014097416157

Stein, M. K., Engle, R. A., Smith, M. S. & Hughes, E. K. (2008) Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell, *Mathematical Thinking and Learning*, 10(4), 313–340. https://doi.org/10.1080/10986060802229675

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

Stylianides, A. J. (2016) *Proving in the elementary mathematics classroom* (Oxford, Oxford University Press). https://doi.org/10.1093/acprof:oso/9780198723066.001.0001

Stylianides, A. J. & Stylianides, G. J. (2009a) Proof constructions and evaluations, *Educational Studies in Mathematics*, 72(2), 237–253. https://doi.org/10.1007/s10649-009-9191-3

Stylianides, A. J., Bieda, K. N. & Morselli, F. (2016) Proof and argumentation in mathematics education research, in: A. Gutiérrez, G. C. Leder & P. Boero (Eds) *The second handbook of research on the Psychology of Mathematics Education* (Rotterdam, Sense Publishers).

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

Stylianides, G. J. & Stylianides, A. J. (2009b) Facilitating the transition from empirical arguments to proof, *Journal for Research in Mathematics Education*, 40(3), 314–352.

Stylianides, G. J. & Stylianides, A. J. (2014) The role of instructional engineering in reducing the uncertainties of ambitious teaching, *Cognition and Instruction*, 32(4), 374–415. https://doi.org/10.1080/07370008.2014.948682

Stylianides, G. J., Stylianides, A. J. & Weber, K. (2017) Research on the teaching and learning of proof: Taking stock and moving forward, in: J. Cai (Ed.) *Compendium for research in mathematics education* (Reston, VA National Council of Teachers of Mathematics).

Weber, K. (2004) Traditional instruction in advanced mathematics courses: A case study of one professor's lectures and proofs in an introductory real analysis course, *The Journal of Mathematical Behavior*, 23(2), 115–133. https://doi.org/10.1016/j.jmathb.2004.03.001

Weber, K. (2005) Problem-solving, proving, and learning: The relationship between problem-solving processes and learning opportunities in the activity of proof construction, *The Journal of Mathematical Behavior*, 24(2), 125–134.

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

Weber, K. (2010) Mathematics majors' perceptions of conviction, validity, and proof, *Mathematical Thinking and Learning*, 12(4), 306–336. https://doi.org/10.1080/10986065.2010.495468

Weber, K. & Alcock, L. (2004) Semantic and syntactic proof productions, *Educational Studies in Mathematics*, 56(2–3), 209–234. https://doi.org/10.1023/b:educ.0000040410.57253.a1

Wheatley, G. H. & Brown, D. (1994) The construction and representation of images in mathematical activity, in: J. P. da Ponte, J. F. Matos (Eds) *Proceedings of the 18th Conference of the International Group for the Psychology of Mathematics Education* (Lisbon, Portugal, University of Lisbon).

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Yackel, E. & Hanna, G. (2003) Reasoning and proof, in: J. Kilpatrick, W. G. Martin & D. Schifter (Eds) *A research companion to Principles and Standards for School Mathematics* (Reston, VA National Council of Teachers of Mathematics).

Yu, J. Y. W., Chin, E. T. & Lin, C. J. (2004) Taiwanese junior high school students' understanding about the validity of conditional statements, *International Journal of Science and Mathematics Education*, 2(2), 257–285. https://doi.org/10.1007/s10763-004-6483-z

Zazkis, R., Liljedahl, P. & Chernoff, E. J. (2008) The role of examples in forming and refuting generalizations, *ZDM – The International Journal on Mathematics Education*, 40(1), 131–141. https://doi.org/10.1007/s11858-007-0065-9

Zazkis, D., Weber, K. & Mejía-Ramos, J. P. (2016) Bridging the gap between graphical argument and verbal–symbolic proofs in calculus, *Educational Studies in Mathematics*, 92(3), 155–173. https://doi.org/10.1007/s10649-016-9698-3

© 2018 The Author. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.