Students' beliefs about empirical arguments and mathematical proof in an introduction to proof class

We report findings from a longitudinal study of students' beliefs about empirical arguments and mathematical proof. We consider the influence of an 'Introduction to Proof (ITP)' course and the consequences of the observed changes in behaviour. Consistent with recent literature, our findings suggest that a majority of the thirty-eight undergraduate students in this study do not find empirical arguments convincing, even at the beginning of their ITP course. We use Sankey diagrams to show that, while many were unconvinced by these arguments at the start and end of the course, others began the course endorsing empirical arguments as similar to their own, shifting toward deductive-symbolic arguments by the end. Finally, we consider the value of Sankey diagrams for understanding changes in population behaviours, and the consequences of our work for future research on the role of empirical arguments in the classroom.

Keywords: introduction-to-proof; mathematical proof; proof transitions; empirical and deductive arguments, and Sankey diagrams

Introduction

Proof is central to the discipline of mathematics and is a well-known stumbling block for many students (Moore 1994; Selden & Selden 2003, Weber 2010). During their undergraduate studies, mathematics majors must make the transition, typically from calculation-based calculus courses, to proof-based mathematics. To combat the difficulty of this transition, many universities offer so-called ITPs, or 'Introduction to proof' courses. While these courses vary in mathematical content, they pursue fundamentally similar goals: developing students' ability to produce, comprehend, and validate written mathematical arguments.

In this paper, we focus on students' beliefs about acceptable forms of evidence in written mathematical arguments and the role of ITPs in the evolution of these beliefs. We attend to the types of evidence students find convincing at the beginning and end of an ITP class, and offer conclusions about the role of empirical and deductive evidence. We offer a novel approach to understanding students' epistemic positions, using Sankey Diagrams to visualize changes in the cohort as a whole.

Literature

We first review the literature on undergraduate students' epistemic positions regarding mathematical proof. We consider sources of conviction for students' in introduction-to-proof settings, and the forms of evidence students appear willing to accept in different contexts. We then introduce Raman's (2003) *procedural, heuristic* and *key ideas,* as a means to understand our own data on students' evaluation of mathematical arguments.

Empirical arguments and proof

Even after several proof-based courses, many students continue to produce mathematical arguments that would likely not be sanctioned as 'proofs' by the community of mathematicians (Stylianides & Stylianides 2009, Selden & Selden 2003, Weber 2010). In particular, students frequently produce arguments based on empirical evidence, rather than generalizable logical deductions. In their seminal work on proof schemes, Harel and Sowder (1998) concluded that many students based mathematical arguments on empirical evidence, and described these students as invoking an *empirical proof scheme.* This was understood to be indicative of students' belief that empirical evidence is a satisfactory basis upon which to make mathematical arguments. Following Harel and Sowder (1998), researchers have used students' over-reliance on empirical evidence to make inferences about the types of arguments that students find convincing (e.g., Küchemann & Hoyles, 2006; Sowder & Harel, 2003; Recio & Godino, 2001).

Others, however, have been critical of these inferential leaps, seeking to distinguish between the types of arguments students produce, and the types of arguments they find convincing. Weber (2010) was critical of this earlier body of work, finding that the majority of his 28 undergraduate students were unconvinced by empirical arguments, even if they themselves would produce one in certain contexts. Stylianides and Stylianides (2009) and Iannone and Inglis (2010) reported similar findings, concluding that students were aware of the limitations of the evidential forms upon which many of their arguments were based. Moreover, Healy and Hoyles (2000) found that even high school students nominate empirical arguments as most similar to their own, but explicitly demonstrate awareness that these arguments would not be sanctioned by the mathematical community. Finally, Stylianou, Blanton, and Rotou (2015) found that students often preferred deductive arguments over empirical arguments when given a variety of proofs to choose from for a particular statement, though their actual constructed proofs did not seem to reflect this preference. The empirical study presented in this paper is based upon tools introduced to the literature by Stylianou et al. In particular, we use their survey focused on eliciting students' beliefs about particular mathematical arguments, to gain empirical insights into the role and value of an introduction to proof module. Before presenting this study, we explore two more theoretical constructs upon which to build our understanding of the data presented.

Beyond a dichotomous view of conviction

We join Brown (2014) in the observation that 'much of the research on empiricism has employed a dichotomous view of conviction and… that an alternative approach may be warranted' (p. 332). In moving toward an alternative approach, we note the pivotal role of mathematical context in the findings of, for example, Knuth, Choppin, and Bieda (2009). Knuth et al. found that amongst 400 middle school students' justification for claims in elementary number theory, 81% of the arguments produced relied on empirical evidence for one conjecture, while another conjecture yielded only 36% empirically-based responses. The authors hypothesize that the relative difficulty of the task plays a substantive role in students' inclusion of various evidential forms. In particular, as in other reports discussed earlier (Weber 2010; Stylianides & Stylianides 2009, Iannone & Inglis 2010), many students may understand that a deductive argument would be desirable but cannot produce one in the moment.

In an attempt to develop an account for the apparent contradiction between students' espoused beliefs and their proof seeking behaviours, Weber, Lew, and Mejia-Ramos (2020) invoked Expectancy Value Theory, most frequently seen in educational psychology. Weber et al. noted that 'notions of value, cost, and likelihood of success could explain why [students] settle for empirical justifications' (p. 27) instead of seeking arguments they understand to be considered more rigorous. In many cases, the (time-)cost of continuing the search for a deductive proof, combined with their perception of their chances of success lead students to 'settle for an empirical justification, even if they are aware that a proof, in principle, bestow[s] more conviction' (p. 32). Important to Weber et al.'s account of students' behaviour is the explicit distinction between the arguments students produce and the types of evidence they believe form the basis of rigorous mathematical proofs.

Raman's (2003) key ideas

In this work, we utilize Raman's (2003) notion of the *key idea* as a theoretical frame for approaching the data. This concept posits that expert proof activity is predicated on the ability to isolate and comprehend, at a broad level, the most important detail(s) of an argument. The key idea then serves as a pivot from our individual heuristics toward a more rigorous argument to be shared with the wider mathematical community. 'The heuristic idea is essentially private, the procedural idea is essentially public, and the key idea provides the link between the two' (p. 324). Raman (2003) suggests that one's ability to engage with key ideas is not only a sign of expert proving activity but also an indicator of one's epistemological maturity: "They [students] do not see the essential connection between their privately held idea and what they expect to produce as a formal, public proof" (p. 321).

A study by Weber and Mejia-Ramos (2014), investigating mathematics majors' and mathematicians' beliefs concerning proof reading, found that "most mathematics majors believe they understand a proof completely if they can justify each step in the proof, while mathematicians generally believe understanding a proof consists of more than this" (Weber & Mejia-Ramos, 2014, p. 99), including understandings which we interpret as mapping onto Raman's key and heuristic ideas.

The questions we ask, then, using Raman's (2003) framework, is how close should we expect students' arguments and endorsed arguments to be to the key idea and should our instruction be designed so that they are given adequate opportunity to identify and develop key ideas? Should introductory proof courses offer students a realistic glimpse of what a mature proof epistemology looks like? And if so, how? Using Raman's (2003) framing allows us to better understand how introductory proof courses provoke students (or not) toward a kind of mean in which "thinking carefully and critically about mathematical ideas" (Thurston, p. 34, 1995) is emphasized over simple appeals to rigor or empirical examples.

Our research questions and motivations

We build on Weber et al. (2020) and Stylianou et al. (2015), intending to further understand students' enculturation into the community of proof-based mathematics. It may be natural to think that ITP courses play a pivotal role in the maturing of mathematical beliefs. However, to our knowledge, this question of timing with regard to the development of proof-related beliefs has gone largely unexamined. In this light, the research we present here is guided by the following research questions.

RQ1: Upon entering an ITP course, what views do students hold regarding the role and nature of empirical and deductive mathematical arguments?

RQ2: How are students' views about empirical and deductive arguments impacted by an ITP course?

Methods

Building on an existing instrument

The data we present is based on a survey developed by Stylianou, Blanton, and Rotou (2015). This survey was designed to probe students thinking about mathematical arguments for each of four given mathematical conjectures, see Table 1.

[Insert Table 1]

For each conjecture, students were shown two empirical and two deductive arguments. Both pairs varied in presentation format to generate a variety of potential responses. The empirical arguments were either visual or numeric, while the deductive arguments were either narrative or symbolic. Figure 1 shows examples of the four arguments forms with respect to Conjecture 1.

[Insert Figure 1]

We adopted the following classification from Stylianou et al.'s (2015) paper (see p. 99)):

(a) an argument characterized as empirical (e.g., Argument A in Fig. 1),

(b) an argument that relied on common properties presented in a narrative but with insufficient explanation (e.g., Argument B in Fig. 1). This argument can, by some standards, be considered deductive,

(c) an argument that could be characterized as a deductive proof written in simple formal style using mathematical symbols (e.g., Argument C in Fig. 1), and

(d) a visual argument that relied on common properties of a generic case (e.g., Argument D in Fig. 1).

Not all deductive narrative arguments were fully correct arguments; most contained a narrative argument that included the key idea(s) of the argument while others had some flaws.

Our survey

The survey we use in this report is a subset of the multiple-choice instrument used by Stylianou et al. (2015). This survey was designed to probe students thinking about mathematical arguments for each of four given mathematical conjectures[1], see Table 1. We focus on the students' perceptions of proof and their changes over time, rather than their alignment with a preordained notion of correctness.

For each conjecture, participants were asked to respond to the following three prompts:

---

[1] While the original authors refer to this instrument as a 'test', we prefer the term 'survey' to distance ourselves from the implicit claim that the instrument's items feature objectively correct (and incorrect) responses.

(1) Choose the argument that is *closest to what you would produce* if you were asked to produce one

(2) Choose the argument that is *most rigorous* (mathematically correct)

(3) Look again at 'Argument A/B/C/D' and determine which of the following best fits your thinking:

   (a) It is logically flawed

   (b) It is a correct argument but it is not a rigorous one [2]

   (c) It only shows that the statement is true for some [cases]

   (d) It is a strong mathematical argument[3]

Participants and procedures

This study involved thirty-eight undergraduate students, all enrolled in an Introduction to Proof (ITP) course at a large US research university. Instruction was carried out via a traditional lecture format, with accompanying notes provided to the students. Participants were typically in their second or third year of tertiary study and had completed both differential and integral calculus with grades that ranged from low C's to high A's.

The data we present here is a subset of a larger study with the same cohort of students. For details of the larger study, we refer to the reader to Miller and CadwalladerOlsker (2020).

---

[2] We note the nuances of statement b and recognize this may be problematic but we take the phrasing that Stylianou et al. use in their paper.

[3] We understand that statements (a) through (d) are not mutually exclusive, nor do they need to be. The participants are asked to choose between statements (a) through (d) that best fits their thinking.

The full survey was given at the beginning and end of the semester. On both occasions, participants were given 50 minutes to complete the survey. Thirty-eight students voluntarily completed the survey at the beginning and end of the course of instruction.

Sankey Diagrams

Sankey diagrams are used to emphasize the major transfers or flows within a system ("Sankey Diagrams", 2022). In terms of this study, Sankey diagrams highlight transitions from one belief state for mathematical proofs to another as measured pre- and post-instruction. The Sankey diagrams illustrate the most important contributions in the flow on how subsets of the participants transition from pre- to post-instruction for the four different types of arguments (empirical-numeric, empirical-visual, deductive-narrative, and deductive symbolic). In terms of the transitions that participants make across the course, Sankey diagrams can help visualize if the transitions are stable and that most of the students are transitioning to one or two particular belief states or if the transitions are unstable and there is a mixture of transitions (sometimes very chaotic) that are occurring to multiple belief states.

Results

We focus on two aspects of the results of our investigation. First, we use Sankey diagrams to highlight the changes in students' responses to prompt 1 (selecting the argument closest to their own approach) and prompt 2 (selecting the argument they perceive as most rigorous). By tracking changes in students' responses pre- and post-ITP instruction, we can consider the impact of this instruction on students' beliefs about the role and nature of proof in mathematics. Second, we present frequency statistics documenting changes in students' responses to prompt 3, in which students were asked to make specific evaluations of each of the 16 arguments (four arguments for each of four conjectures). In the interests of space, we omit several Sankey diagrams and pay attention only to aspects of the data we deem most interesting.

Sankey diagrams for Prompt 1 - 'Closest to your own approach'

Figure 2 shows the Sankey diagram for Conjecture 1. Pre-instruction, the majority of the participants choose deductive-symbolic (50%) and empirical-numeric arguments (24%), while fewer choose deductive-narrative (21%) and empirical-visual (5%). At post-instruction, participants almost exclusively choose deductive-symbolic (84%). The transition across the semester reflects the expected (or at least, hoped for) transition through an ITP course. At the beginning of the course, students reported a variety of approaches to prove this conjecture. By the end of the semester, the majority had converged toward a deductive-symbolic approach, consistent with the goals of instruction. The diagram for Conjecture 1 was similar to Conjecture 2 and is not shown here. We did not ask students to actually produce their own arguments, so we interpret their responses as indicative of their intentions, rather than the results of their labour.

[Insert Figure 2]

Figure 3 shows the equivalent Sankey diagrams for Conjecture 4. Pre-instruction, the majority of participants choose deductive-symbolic (39%) and empirical-visual (34%), while fewer choose deductive-narrative (11%) and empirical-numeric (13%). At post-instruction, almost all participants choose deductive-symbolic (74%) and deductive-narrative (16%), with just a few participants choosing empirical-visual (8%) and empirical-numeric (2%). In contrast to Conjectures 1 and 2, the transitions for Conjecture 4 were more complex. And in particular, note that a small number of students began the course by asserting they would invoke a deductive-symbolic argument, and ended asserting that they would invoke either a deductive narrative or empirical-visual report. That said, these students were in the minority, with the majority of participants transitioning *toward* the expected deductive-symbolic approach.

[Insert Figure 3]

Conjecture 3 yielded the most complex, and hence most interesting Sankey diagram, see Figure 4. Pre-instruction, the majority of participants chose deductive-narrative (42%) and deductive-symbolic (32%), while a number of participants chose empirical-narrative (18%) and fewer chose empirical-visual (8%). The percentages did not change substantially at post-instruction, with the majority still chose deductive-narrative (47%) or deductive-symbolic (45%), and only a few chose empirical-numeric (5%) and empirical-visual (3%).

[Insert Figure 4]

Our findings with respect to prompt 1 suggest, as one might expect, that students' view of their own approaches varies by mathematical context and that their ITP instruction did not uniformly prompt them to report using deductive-symbolic arguments.

As was discussed earlier in our literature review, there are limited epistemic inferences we can draw based explicitly on the types of arguments students produce, or purport to produce. Hence, we turn to prompt 2 and perform a similar analysis based on students' perceptions of the most rigorous mathematical argument for each conjecture.

Sankey diagrams for Prompt 2: 'Choose the most rigorous argument'

The Sankey diagrams for Conjectures 1, 2, and 4 for the most rigorous response were similar and we present the diagram for Conjecture 1 as a representative, see Figure 5. Pre-instruction, the majority of participants chose deductive-symbolic (74%) and empirical-numeric (16%), while fewer chose deductive-narrative (8%) and empirical-visual (2%). At post-instruction the percentages were very similar, the majority of participants choose deductive-symbolic (68%) and empirical-numeric (18%), while fewer chose deductive-narrative (8%). The transitions across the semester were somewhat mixed, with a small gain for deductive-narrative and loss for the empirical ones.

[Insert Figure 5]

As was the case for Prompt 1, the Sankey diagram for Conjecture 3 is more complex than its Prompt 2-counterpart, see Figure 6. Both before and after instruction, the majority of participants chose deductive-symbolic and deductive-narrative arguments, while fewer chose empirical-numeric (11% pre- instruction and 8% post-instruction). However, again, the transitions across the semester show that many participants have changed their answers. Of particular interest are the 24% who have moved away from valuing deductive-symbolic arguments as most rigorous on conjecture 3. For these students, their ITP instruction has convinced them that, at least in this context, the deductive symbolic argument was not the 'most rigorous' or 'most correct' response.

[Insert Figure 6]

Again, from the differences between Conjecture 3 and the others, we conclude that mathematical context has a role to play in students' perceptions of the value of various arguments. We discuss the consequences of this conclusion in the discussion section to follow.

Prompt 3: Evaluate each of the 16 arguments

For each of the 16 arguments (four for each conjecture), participants were asked to examine each argument and choose one of the following responses: a) "It is logically flawed", b) "It is a correct argument but it is not a rigorous one", c) "It only shows that the statement is true for some [cases]", or d) "It is a strong mathematical argument." Table 2 summarizes the responses pre- and post-instruction for each conjecture.

[Insert Table 2]

Table 2 presents a static picture at pre- and post-survey and does not capture the variety of transitions students made from pre- to post-survey. To better visualize these changes, we present selected Sankey diagrams for participants' views of the arguments with some observations. We organize these results in the different argument types to bring out the salient features of the student responses for each conjecture. In the conclusion, we will synthesize the information to state corroborating findings and major results.

Empirical-Numeric Arguments

The students' responses for empirical-numeric arguments were very similar for conjectures 1 and 2. Figure 7 shows the archetype Sankey diagram for conjecture 1. Before instruction, the majority of the participants chose 'only true for a limited number of cases' (66%), while less participants chose 'correct not rigorous' (32%) and only a few chose 'strong mathematical argument' (3%). At post-instruction, participants almost exclusively chose 'only true for a limited number of cases' (82%), while a small minority of participants selected the response 'correct not rigorous' (13%) and only a few chose the other two responses (6%). The transition across the semester reflects that the majority of students understand that empirical-numeric arguments have limitations since they only show the statement is true for a small number of cases, especially at post-instruction. That is, the majority of participants chose the 'only true for a limited number of cases' response at both pre- and post-instruction while a large percentage of participants that choose one of the other views ('correct not rigorous' or 'strong mathematical argument') at pre-instruction changed their view to 'only true for a limited number of cases' at post-instruction.

[Insert Figure 7]

There were a sizeable number of students that believed that an empirical-numeric argument was 'correct not rigorous' at pre-instruction, but this decreased considerably by post-instruction signifying that they understood that these types of arguments are not rigorous. Overall, we see few students transitioned to the choice of ‘correct not rigorous’ by post-instruction. We note that although one student chose ‘only true for a limited number of cases’ at pre-instruction, they responded that it was a ‘strong mathematical argument’ at the end of the ITP course.

The transitions for the empirical-numeric argument for conjecture 3 and 4 from pre- to post-instruction were more complex and we present the Sankey diagram for Conjecture 3 as an illustration of these types of transitions. At pre-instruction, the majority of the participants chose ‘only true for a limited number of cases’ (71%), while a minority chose a ‘strong mathematical argument’ (16%) and ‘correct not rigorous’ (13%). At post-instruction, participants chose ‘only true for a limited number of cases’ at a slightly lower level as pre-instruction (61%), while a minority chose the response ‘correct not rigorous’ (18%) and ‘logically flawed’ (13%), and fewer chose ‘strong mathematical argument’ (8%). We see in Figure 8 that it is concerning that some participants chose ‘only true for a limited number of cases’ at pre-instruction but chose other responses at post-instruction. Contrast this with the more appropriate transitions where a small number of participants changed their view at pre-instruction from a response other than ‘only true for a limited number of cases’ to ‘only true for a limited number of cases' at post-instruction.

[Insert Figure 8]

Therefore, in Table 2, there was a small 10% post-instruction decrease for ‘only true for a limited number of cases’. We cannot see the variety of transitions that are occurring. However, from the Sankey diagram, we can clearly see the different transitions occurring regarding students’ choice of ‘only true for a limited number of cases’, along with other choices at pre- and post-instruction.

Deductive-Symbolic Arguments

Students' responses for deductive-symbolic arguments were similar for conjectures 1, 2, and 4 and Figure 9 embodies this with the Sankey diagram for conjecture 1.

[Insert Figure 9]

At pre-instruction, the majority of the participants chose a 'strong mathematical argument' (84%), while fewer participants chose 'only true for a limited number of cases' (11%), 'correct not rigorous' (3%) and 'logically flawed' (3%). At post-instruction, almost an equal number of participants chose a 'strong mathematical argument' (87%), while fewer chose 'correct not rigorous' (11%) and 'logically flawed' (3%). The transition across the semester reflects that the majority of students understand that deductive-symbolic arguments are strong mathematical arguments. We observe that all the participants that chose 'only true for a limited number of cases' at pre-instruction, chose a 'strong mathematical argument' at post-instruction. However, there were a close to equal number of students that transitioned away from a 'strong mathematical argument' at pre-instruction to either 'logically flawed' or 'correct not rigorous' at post-instruction. Finally, a small number of students believed that the deductive-symbolic was 'correct but not rigorous'.

The transitions for the deductive-symbolic argument for conjecture 3 from pre- to post-instruction were more complicated as shown in Figure 10.

[Insert Figure 10]

At pre-instruction, close to half of the participants chose a 'strong mathematical argument' (45%) and about one-fourth chose 'correct not rigorous' (26%). While a minority chose 'only true for a limited number of cases' (16%) and 'logically flawed' (13%). At post-instruction, again about half of the participants chose a 'strong mathematical argument' (42%) and about one-third chose 'correct not rigorous' (32%). One quarter of participants chose either 'logically flawed' (13%) or 'only true for limited number of cases' (13%). It is troubling that a number of participants moved away from choosing a 'strong mathematical argument' to either 'correct not rigorous' or 'logically flawed'. However, we do note that a similar number of participants transitioned toward this more desired 'strong mathematical argument' response at post-instruction.

Empirical-Visual Argument

The transitions that students make from pre- to post-instruction are very mixed for all the conjectures. Conjectures 1, 2 and 4 are similar in that half or more of the participants choose the response of 'correct not rigorous' and their transitions are similar. We choose Conjecture 2 as a representative of the types of responses shown in Figure 11. At pre-instruction, half of the participants chose 'logically flawed' (50%) and a little over a quarter (29%) chose 'correct not rigorous', while a little less than one-quarter of the participants chose 'only for a limited number of cases' (18%) and only 3% chose a 'strong mathematical argument'. At post-instruction, about half of the participants again chose 'logically flawed' (47%), while almost half of the responses were 'correct not rigorous' (26%) and 'only true for a limited number of cases' (21%). The transitions from pre- to post-instruction were quite complex. For example, approximately two-thirds of the participants transitioned away from the most popular response of 'logically flawed' at pre-instruction to one of the other responses at post-instruction, even though the raw percentage for 'logically flawed' did not change substantially from pre-to post-instruction (see Table 2). In addition, nearly three quarters of the participants transitioned away from 'correct not rigorous' response at pre-instruction to one of the other choices and similar transitions for the other two choices.

[Insert Figure 11]

At pre-instruction in Figure 12, around five-eighths of the participants chose ‘correct not rigorous’ (61%) and about a quarter chose ‘only true for limited number of cases’ (24%), while a small percentage of participants chose a ‘strong mathematical argument’ (8%) and ‘logically flawed’ (8%).

[Insert Figure 12]

At post-instruction, a large majority of the participants choices were mostly evenly distributed between ‘only true for a limited number of cases’ (34%), ‘correct not rigorous’ (29%), and ‘logically flawed’ (26%), while only (11%) chose a ‘strong mathematical argument’. We note that there was a significant decrease from ‘correct not rigorous’ from pre- to post-instruction. At post-instruction, the most popular response was ‘only true for a limited number of cases’ and the transitions to this response came from pre-instruction responses of ‘correct but not rigorous’ or ‘only true for limit of cases’. Similarly, to the transitions for conjecture 2 for Figure 11, close to three-fourths of the participants that chose ‘correct not rigorous’ transitioned to one of the other choices with the majority transitioned to ‘logically flawed’ and ‘only true for a limited number of cases’. Also, about half of the participants that chose ‘only true for a limited number of cases’ transitioned to one of the other choices.

Deductive-Narrative Arguments

The transitions for deductive-narrative arguments are mixed for all conjectures. We see similar transitions for deductive-narrative arguments on conjectures 2 and 4 where about half of the participants chose ‘correct not rigorous’ response at pre-instruction and post-instruction, with not much change in the percentages from pre- to post-instruction for the other responses. Although conjecture 1 has similar responses to conjectures 2 and 4 at pre-instruction, there is nearly a 20% increase in ‘correct not rigorous’ response at post-instruction. We illustrate these types of transitions for conjecture 2 and 4 in Figure 13 with the Sankey diagram for conjecture 2.

[Insert Figure 13]

At pre-instruction, a majority of the participants either chose 'correct not rigorous' (47%) or 'logically flawed' (29%), while about one-quarter of the participants chose a 'strong mathematical argument' (21%) or 'only true for a limited number of cases' (3%). At post-instruction, the majority of participants chose 'correct not rigorous' (50%) and 'logically flawed' (39%), while a small percentage chose a 'strong mathematical argument' (5%) or 'only true for a limited number of cases' (5%). About half of the participants that chose 'correct not rigorous' at pre-instruction transitioned to a different choice at post-instruction. Similarly, about half of the participants that chose 'logically flawed' transitioned away from 'logically flawed' to the choice of 'correct not rigorous'. It is important to note that all of the participants that choose a 'strong mathematical argument' at pre-instruction transitioned to one of the other choices at post-instruction. We see that at pre- and post-instruction a majority of the participants viewed deductive narrative as a correct argument that does not have sufficient amount of rigor to be a formal argument compared to what they may be used to seeing in lecture or in the textbook, or they judge these arguments not up to the same standard of rigor that they normally see in mathematics. Although the numbers for 'correct not rigorous' at pre- and post-instruction increase by a small amount (3%), we see that approximately half of participants changed their view from pre- to post-instruction. In addition, nearly half of those who chose 'logically flawed' and a 'strong mathematical argument' at pre-instruction transitioned to 'correct not rigorous' response at post-instruction.

A different picture forms for conjecture 3, shown in Figure 14. About half of the participants chose a 'strong mathematical argument' for conjecture 3 both pre- and post-instruction. At pre-instruction, about half of the participants chose a ‘strong mathematical argument’ (47%) and about an equal percentage of participants chose either ‘logically flawed’ (26%) or ‘correct not rigorous’ (21%), while very few chose ‘only true for a limited number of cases’ (5%).

[Insert Figure 14]

At post-instruction, again half of the participants chose a ‘strong mathematical argument’ (50%), about one-quarter chose ‘correct not rigorous’ (29%), and less than one-quarter chose ‘logically flawed’ (18%), while very few again choose ‘only true for a limited number of cases’ (3%). For conjecture 3, about half of the participants abandoned the most common view that deductive-narrative arguments are strong mathematical arguments and changed to a response of ‘correct not rigorous’ or ‘logically flawed’. There were quite a few different transitions occurring form pre- to post-instruction. For example, roughly about half of the ‘strong mathematical argument’ response at pre-instruction transitioned to ‘logically flawed’ or ‘correct not rigorous’. While around half of the responses at pre-instruction for ‘logically flawed’ and ‘correct not rigorous’ transitioned to a ‘strong mathematical argument’ at post-instruction. We will discuss this in the conclusion where context and difficulty play a role in students epistemic views of mathematical arguments.

Discussion

Having presented a variety of data and a series of Sankey diagrams, we now return to our research questions focused on students’ beliefs about empirical and deductive arguments, and the role of ITP courses in changing these beliefs. We organize the discussion in three different sections around three main findings. Finally, we turn to implications for practice, both in terms of future research and classroom practice.

The role of context on participants transitions and views of mathematical arguments

The first main finding is that mathematical context has a substantive impact on students' views and behaviours. We note that students' responses to prompts associated with conjecture 3 and/or conjecture 4 differed from the norm in multiple ways for participants' closest to their own approach and most rigorous arguments responses, and for their views on each of the four argument types. First, while the majority of participants transitioned to deductive-symbolic on the argument closest to their own approach on conjectures 1, 2, and 4, this was not the case for conjecture 3 where a similar number transitioned to deductive-symbolic and deductive-narrative at post-instruction. Second, while the majority of participants transitioned to deductive-symbolic argument as the most rigorous argument on conjectures 1 and 2, where most of these transitions started with a deductive-symbolic, this was not the case for conjectures 3 and 4. Third, for each argument type, conjecture 3 and/or conjecture 4 elicited different transitions than the other conjectures which can be seen by looking at the representative Sankey diagrams for the types of arguments. Therefore, the type of arguments that students would produce (closest to their own), the arguments that they view as the most rigorous, and their views of each argument type are not stable when the context is changed.

There may be a variety of reasons for the change in students' view of proofs based on the context. First, there is a difference in mathematical context for conjectures 1, 2 and 3 (number theory) and conjecture 4 (geometry). This might account for some of the differences we see in conjecture 4. Secondly, some of the arguments (conjectures 1, 2 and 4) may have been more familiar to the students than conjecture 3, especially at post-instruction since students are asked to prove statements similar to conjectures 1 and 2 in the ITP course. On one hand, for the more familiar conjectures where students have been exposed to the material and statements, students have a better idea of what type of argument they would produce and feel would be most rigorous. On the other hand, when students are faced with unfamiliar statements, students will first need to get acclimated before figuring out how to approach and develop an argument or proof. Therefore, we expect that participants responses will be less aligned and more mixed for the unfamiliar conjectures than for the familiar ones. Third, the conjectures were expressed in more natural language (words) in conjectures 1, 3, and 4 and more symbolic in conjecture 2. We suspect that the views of participants and their choices were mainly due to the fact that conjectures 1 and 2 and their proofs were familiar to them from their ITP course, but the proofs of conjectures 3 and 4 were less familiar from material in the ITP course even though the conjectures themselves may be more familiar from previous mathematics classes. We agree with Knuth et al. (2009) that as mathematical statements become more complex (abstract), students are more likely to produce less precise corresponding arguments, especially in more abstract upper-level mathematics courses. In these less familiar and more complex contexts (like conjecture 3 and 4), students may first utilize mathematical tools to become familiar with both the context and statement and need to think more deeply about the mathematics. Then they can develop an informal argument that emphasizes the key idea that will lead to the production of a formal proof.

Students' transitions and views for deductive-symbolic and empirical-numeric arguments

The second main finding for the task of choosing the closest to their own approach and most rigorous arguments are that students' transitions are aligning almost exclusively with deductive-symbolic and almost always eschew empirical-numeric arguments after completing an introduction-to-proof course, except for more complex statements like conjecture 3. In addition, a majority of the participants' chose the deductive-symbolic and empirical-numeric arguments as a strong mathematical argument and only true for a limited number of cases at pre-and post-instruction, respectively, (except for the context discussed above). For example, as can be seen in Figures 2 through 6 and holds true for Sankey diagrams that we did not show, only a few (24% or less) students favoured the empirical-numeric argument as the argument they would have produced or advocated as the most rigorous argument at pre-instruction and even fewer (8% or less) at post-instruction. Prior studies have documented that different populations can be convinced by empirical-numeric arguments (similar to our data at pre-instruction), including pre- or in-service teachers (Knuth, 2002; Martin & Harel, 1989), and middle school (Bieda & Lepak, 2014; Knuth, 2002; Knuth et al., 2009), high school (Chazan, 1993; Healy & Hoyles, 2000), and college (Recio & Godino, 2001; Sowder & Harel, 2003; Stylianou et al, 2015) students.

However, at post-instruction, we find that very few students agree that the empirical-numeric argument is the closest to their own approach or the most rigorous argument and this aligns well with past research that more experienced college math majors do not advocate that an empirical-numeric argument is a valid proof (Iannone &. Inglis, 2010; Weber, 2010). We find that at post-instruction for closest to their own approach and most rigorous, except for the contextual cases mentioned above, the majority of students align with the deductive-symbolic arguments that are espoused by mathematicians and eschew empirical-numeric arguments. In addition, students knew that those types of arguments were strong mathematical arguments and only true for a limited number of cases. Raman's key idea (2003) may be more implicit for these students that choose deductive-symbolic arguments where they see the key idea and are able to map it to a formal argument. Our data corroborate Weber's (2010) finding that most students understand the types of arguments valued by the mathematical community at large. Moreover, Weber, Lew, and Mejia-Ramos (2020) concluded that students produce arguments based on problematic evidential forms not because they believe these arguments to be desirable but because they are the best product available to them in the moment.

Students' transition and views for deductive-narrative and empirical-visual arguments

A third main finding is that students' views for deductive-narrative and empirical-visual arguments are very mixed. For the views of students on the empirical-visual and deductive-narrative arguments, students' responses changed dramatically from pre- to post-instruction. Figure 11 illustrated these mixed views, where a substantial proportion of students changed their view at post-instruction. For example, on the empirical-visual argument on conjecture 2, 61% of students chose "correct not rigorous" at pre-instruction and around two-thirds of these students changed their view to one of the other choices and this was similar to how students' views changed for other responses of 'strong mathematical argument', 'only true for limited number of cases', and 'logically flawed'. This shows that students' views of visual arguments are not well formed when students are being introduced to proofs. This is not completely surprising since diagrams are seen as ancillary to the proving process (Samkoff, Lai, & Weber, 2012) and not all mathematicians agree about visual proofs (Weber and Czocher, 2019).

Students' views on the deductive-narrative argument for each conjecture were also fairly mixed, but there were differences in their views on conjectures 1, 2, and 4 when compared to conjecture 3. For conjectures 1, 2, and 4 at pre-instruction, close to 50% of the students viewed the deductive-narrative argument as 'correct not rigorous', while 21% or less viewed the deductive-narrative argument a 'strong mathematical argument'. For conjecture 3, the roles for these two choices were reversed, while the other significant choice of 'logically flawed' was chosen at a similar level on all conjectures. In addition, a significant percentage of students transitioned away from their pre-instruction choice for their post-instruction choice. For example, on conjecture 2, over 60% of the students changed their view from pre- to post-instruction. Similarly, for the other conjectures.

There may be legitimate reasons for students' views of empirical-visual and deductive-narrative arguments. First, students may be choosing "only true for a limited number of cases" for the empirical-visual argument because the picture is static and only captures a representative case (i.e., conjecture 1 only shows visual that 6+4 is even). Secondly, students may see the picture as a generic proof (Rowland, 2002) or the key idea (Raman, 2003) of seeing through the example(s) or "pairing" when they choose "correct but not rigorous". Thirdly, students may be choosing 'strong mathematical argument' on the deductive-narrative argument because of their explanatory power for students (Healy & Hoyles, 2001; Stylianou et al., 2015).

More fundamentally, we believe the empirical-visual and deductive-narrative arguments capture the key idea of the proof for students and they would need to utilize the key idea to map from the informal argument presented to a formal argument that would convince the mathematical community. For example, the empirical-visual argument for conjecture 1 captures the key idea of "pairing" for each of the two even number counters. For the deductive-narrative, the key idea is that each even number has a common factor of 2 (pairing) and thus the sum will have a common factor of 2.

Teaching Implications

It has been well-documented that mathematicians' use a variety of tools to conjecture and investigate mathematical statements and researchers have made the case that students should be experiencing an environment in the classroom that promotes some (not necessary all) of these same practices of a mathematicians (Harel & Sowder, 2007; Healy & Hoyles, 2001; Sfard, 1998;

Weber, Inglis, & Mejia-Ramos, 2014). One of the tools that mathematicians' use is the key idea to map from an informal proof argument to formal proof argument (Raman, 2003). We advocate giving students opportunities to recognize and develop the key idea of the proof for a given mathematical statement and mapping from informal to formal proof. Some students in this study aligned with the empirical-visual or deductive-narrative argument and these arguments can be viewed as having the key idea that can be mapped to a more formal proof (see example in the discussion subsection on students' transition and views for deductive-narrative and empirical-visual arguments).

Research has documented instructors focus on key ideas in the proofs they produce during lecture (Weber, 2004) and we recommend that instructors clearly articulate to students that they are concentrating on the key idea in their lecture proofs and illustrate periodically how students should utilize the key idea to map to a formal argument. Students may interpret the moves that instructors make in their classrooms too literally and imitate an overly rigorous approach in their own proofs, thereby ignoring or obscuring the key ideas.

In this study we see that for at least two of the conjectures (1 and 2), students' transitioned to the types of arguments endorsed by mathematicians. On the other hand, transitions associated with conjectures 3 and 4 were mixed, and we believe, will continue refining over a more extended period of time. It seems that students' transitions are continually evolving and may take long periods of time to transition to the proof schemes of mathematicians. Mathematicians know the limitations and utility of visual, numeric, and narrative arguments, but also recognize the usefulness of these informal arguments as they construct formal deductive proofs. We should make sure that this message is being received loud and clear and that students are not overgeneralizing to the point where they completely dismiss the usefulness of these types of arguments in their proof productions.

Final remarks

In this paper, we used Sankey diagrams to visualize changes in students' epistemic position on the role of empirical evidence in mathematics. Having observed the nuanced changes discussed above, we conclude with two final remarks.

First, our data shows that the relationship between students' proof-related beliefs and the proofs they endorse is not as straightforward as earlier scholars have presented. That said, our data are well-aligned with the recent report of Weber, Lew, and Mejia-Ramos (2020). In particular, we conclude that the key to helping students lies not in leading them to value deductive arguments, but rather in improving their abilities to produce them. And finally, students' beliefs about the role and validity of various argument forms are not general. Given that some conjectures elicit different beliefs than others, it is of utmost importance that future work in this area is careful about its claims of generality, and to not treat either the behaviours or espoused beliefs of students as static, fixed entities independent of time or context.

References

Bieda, K.N., & Lepak, J. (2014). Are you convinced? Middle-grade students' evaluations of mathematical arguments. *School Science and Mathematics*, 114(4), 166-177. https://doi.org/10.1111/ssm.12066

Brown, S. (2014). On skepticism and its role in the development of proof in the classroom. *Educational Studies in Mathematics*, *86*(3), 311–335. https://doi.org/10.1007/s10649-014-9544-4

Chazan, D. (1993). High school geometry students' justifications for their views of empirical evidence and mathematical proof. *Educational Studies in Mathematics*, 24(4), 359-387. https://doi.org/10.1007/BF01273371

Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. *CBMS Research in Collegiate Mathematics Education. III*, Vol. 7, pp. 234–283.

Harel, G., & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. *Second handbook of research on mathematics teaching and learning*, *2*, 805-842.

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

Iannone, P., & Inglis, M. (2010). Undergraduate students' use of deductive arguments to solve "Prove That …" tasks. *Proceedings of the 7th Congress of the European Society for Research in Mathematics Education*.

Knuth, E.J. (2002). Secondary school mathematics teachers' conceptions of proof. *Journal for Research in Mathematics Education*, 33(5), 379-405. https://doi.org/10.2307/4149959

Knuth, E., 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* (pp. 153–170). New York, NY: Routledge.

Küchemann, D. and Hoyles, C. (2006). Influences on students' mathematical reasoning and patterns in its development: Insights from a Longitudinal study with particular reference to geometry. *International Journal of Science and Mathematics Education. 4*(4), 591-608. https://10.1007/s10763-006-9039-6

Martin, W.G., & Harel, G. (1989). Proof frames of preservice elementary teachers. *Journal for Research in Mathematics Education*, 20(1), 41-51. https://10.5951/jresematheduc.20.1.0041

Miller, D., & CadwalladerOlsker, T. (2020). Investigating undergraduate students' view of and consistency in choosing empirical and deductive arguments. *Research in Mathematics Education*, *22*(3), 249-264. https://doi.org/10.1080/14794802.2019.1677489

Moore, R. (1994). Making the transition to formal proof. *Educational Studies in Mathematics*, *27*(3), 249–266. https://doi.org/10/1007/BF01273731

Raman, M. (2003). Key ideas: What are they and how can they help us understand how people view proof? *Educational Studies in Mathematics*, 52(3), 319-325. https://doi.org/10.1023/A:1024360204239

Recio, A., & Godino, J. (2001). Institutional and personal meanings of mathematical proof. *Educational Studies in Mathematics*, *48*(1), 83–99. https://doi.org/10/1023/A:1015553100103

Rowland, T. (2002). Generic proofs in number theory. *Learning and teaching number theory: Research in cognition and instruction*, 157-183.

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

Sankey Diagram. (2022, March 14). In *Wikipedia*. https://en.wikipedia.org/wiki/Sankey_diagram.

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher*, 27(2), 4-13. https://doi.org/10.3102/0013189X027002004

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

Sowder, L., & Harel, G. (2003). Case studies of mathematics majors' proof understanding, Production, and Appreciation. *Canadian Journal of Math, Science & Technology Education*, *3*(2), 251–267. https://doi.org/10/1080/14926150309556563

Stylianides, G. J., & Stylianides, A. J. (2009). Facilitating the transition from empirical arguments to proof. *Journal for Research in Mathematics Education*, *40*, 314–352. https://doi.org/10.5951/jresematheduc.40.3.0314

Stylianou, D., Blanton, M., & Rotou, O. (2015). Undergraduate students' understanding of proof: Relationships between proof conceptions, beliefs, and classroom experiences with learning proof. *International Journal of Research in Undergraduate Mathematics Education*, *1*, 91–134. https://doi.org/10.1007/s40753-015-0003-0

Thurston, W. P. (1995). On proof and progress in mathematics. *For the Learning of Mathematics*, *15*(1), 29-37.

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.mathb.2004.03.001

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., & Czocher, J. (2019). On mathematicians' disagreements on what constitutes a proof. *Research in Mathematics Education*, *21*(3), 251-270. https://doi.org/10/1080/14794802.2019.1585936

Weber, K., Inglis, M. & Mejia-Ramos, J.P. (2014). How mathematicians obtain conviction: Implications for mathematics instruction and research on epistemic cognition. *Educational Psychologist*, 49(1), 36-58. https://doi.org/10/1080/00461520.2013.865527

Weber, K., & Mejia-Ramos, P. (2014). Mathematics major's beliefs about proof reading. *International Journal of Mathematical Education in Science and Technology*, *45*(1), 89-103. https://doi.org/10.1080/0020739X.2013.790514

Weber, K., Lew, K. & Mejia-Ramos, P. (2020). Using expectancy value theory to account for individuals' mathematical justification. *Cognition and Instruction*, 38(1), 27-56. https://doi.org/10.1080/07370008.2019.1636796

Table 1. The Four Mathematical conjectures
Conjecture 1: *The sum of any two even numbers is always an even number.*
Conjecture 2: *For any integers* a, b, *and* c, *if* a *divides* b *with no remainder, then* a *divides* bc *with no remainder.*
Conjecture 3: *If the sum of digits of a three-digit integer is divisible by 3, then the number is also divisible by 3***.**
Conjecture 4: *The supplements of two congruent angles are congruent.*

Table 2. Participants' Views for the four arguments for each conjecture.

	Empirical-Visual		Empirical-Numeric		Deductive-Narrative		Deductive-Symbolic	
Conjecture 1	Pre	Post	Pre	Post	Pre	Post	Pre	Post
It is logically flawed	18%	37%	0%	3%	26%	13%	3%	3%
Correct argument but not rigorous	53%	37%	26%	13%	42%	68%	3%	11%
True for only a few cases	18%	26%	66%	82%	11%	8%	11%	0%
It is a strong mathematical argument	11%	0%	3%	3%	16%	11%	84%	87%
Conjecture 2								
It is logically flawed	8%	26%	11%	0%	29%	39%	13%	3%
Correct argument but not rigorous	61%	29%	18%	8%	47%	50%	5%	11%
True for only a few cases	24%	34%	63%	92%	3%	5%	8%	3%
It is a strong mathematical argument	8%	11%	8%	0%	21%	5%	74%	84%
Conjecture 3								
It is logically flawed	50%	47%	0%	5%	26%	18%	13%	13%
Correct argument but not rigorous	29%	26%	16%	13%	21%	29%	26%	32%
True for only a few cases	18%	21%	76%	79%	5%	3%	16%	13%
It is a strong mathematical argument	3%	5%	8%	3%	47%	50%	45%	42%
Conjecture 4								
It is logically flawed	21%	32%	0%	13%	24%	21%	8%	8%
Correct argument but not rigorous	50%	58%	13%	18%	47%	47%	5%	13%
True for only a few cases	3%	5%	71%	61%	8%	11%	18%	3%
It is a strong mathematical argument	26%	5%	16%	8%	21%	21%	68%	76%

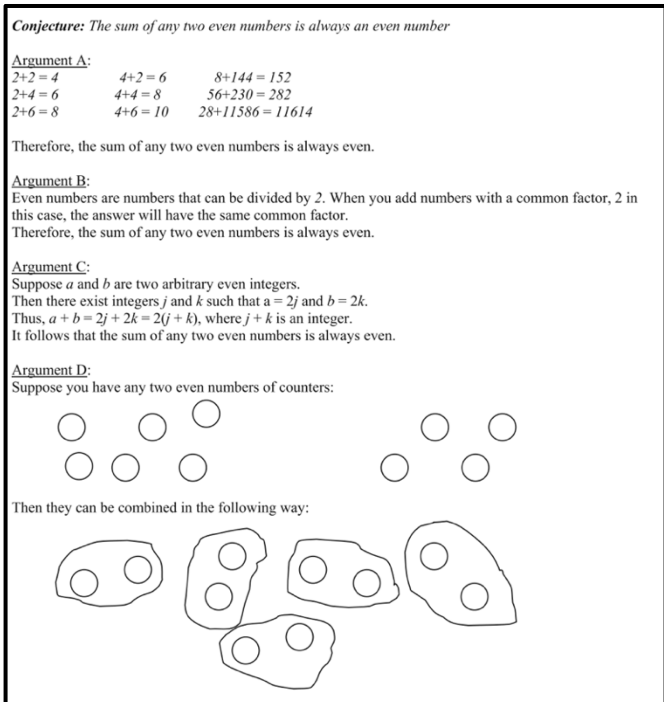

Figure 1. Mathematical arguments for each of four given mathematical conjectures, adapted from Stylianou et al. (2015).

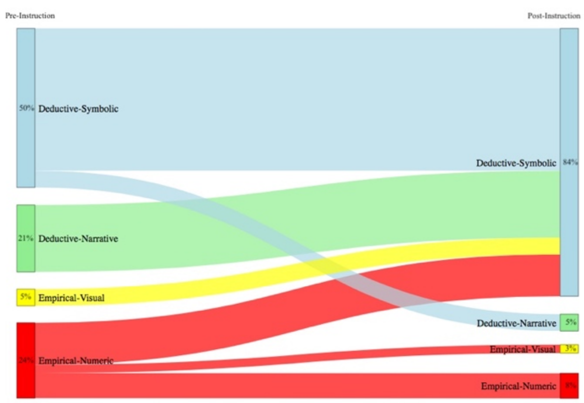

Figure 2. Sankey diagram for the closest to their own approach – Conjecture 1.

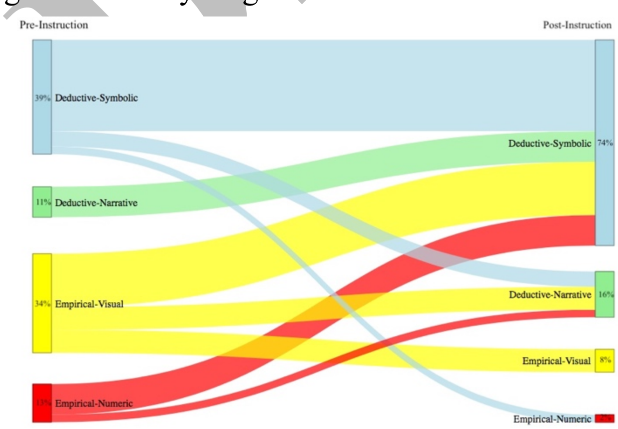

Figure 3. Sankey diagram for closest to their own approach – Conjecture 4.

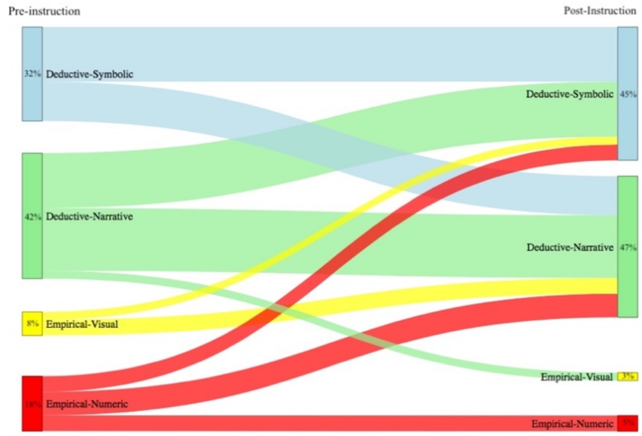

Figure 4. Sankey diagram for closest to their own approach - Conjecture 3.

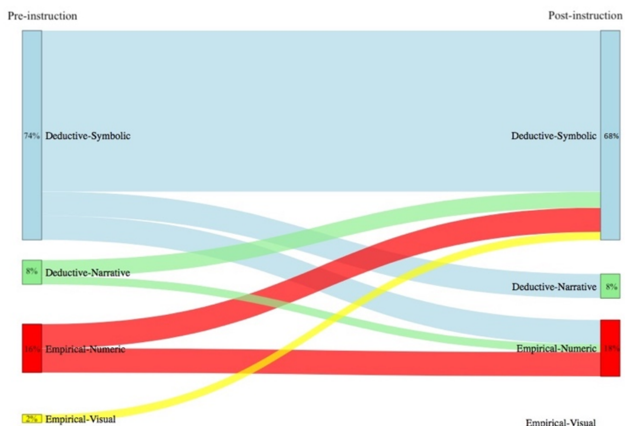

Figure 5. Sankey Diagram for most rigorous - Conjecture 1.

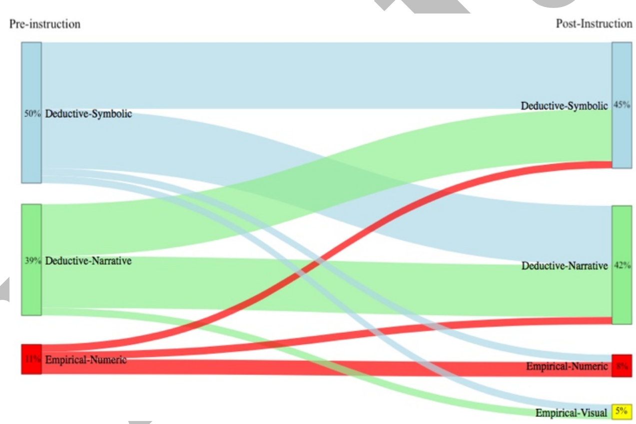

Figure 6. Sankey diagram for most rigorous - Conjecture 3.

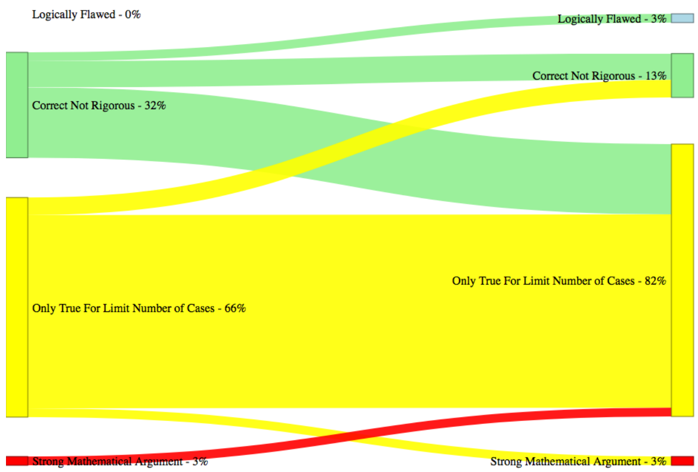

Figure 7. Sankey diagram for empirical-numeric argument - Conjecture 1.

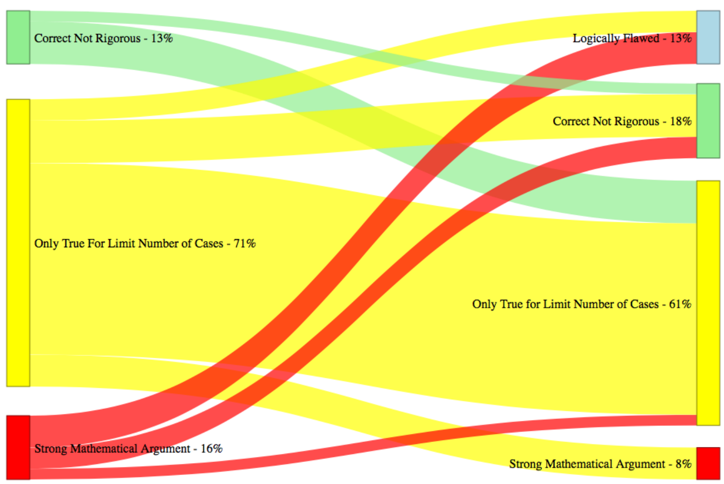

Figure 8. Sankey diagram for empirical-numeric argument – Conjecture 3.

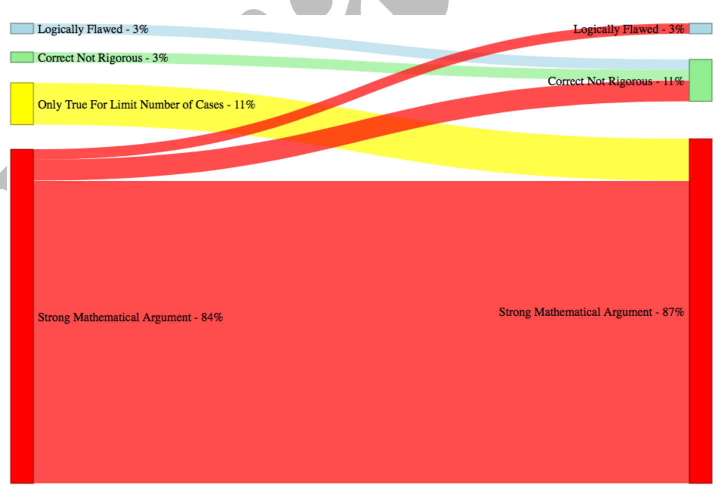

Figure 9: Sankey diagram for deductive-symbolic argument - Conjecture 1.

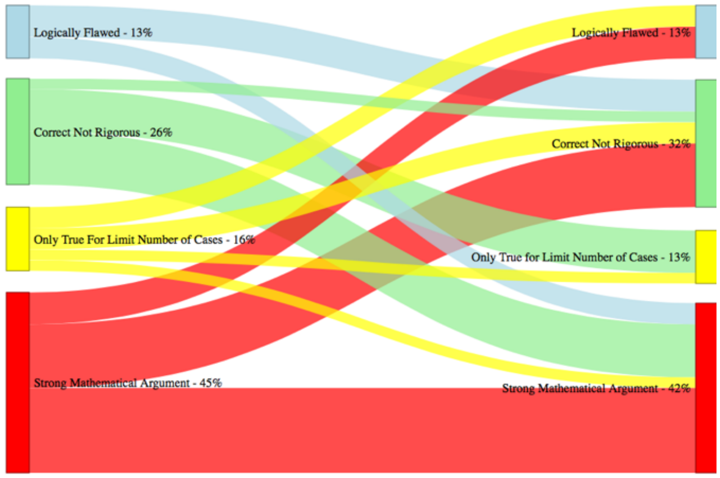

Figure 10. Sankey diagram for deductive-symbolic argument - Conjecture 3.

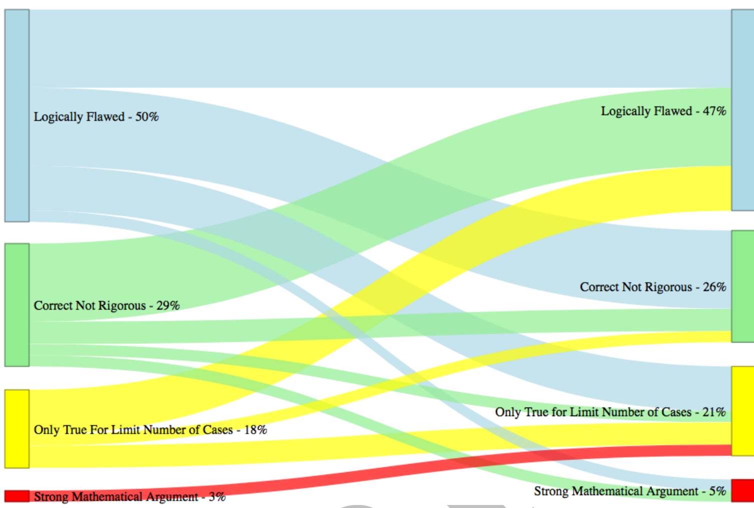

Figure 11. Sankey diagram for empirical-visual argument – Conjecture 2.

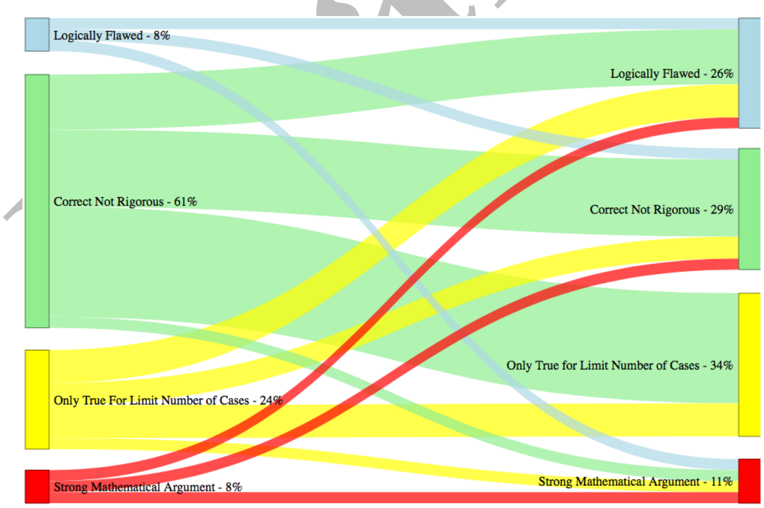

Figure 12: Sankey diagram for empirical-visual argument - Conjecture 3.

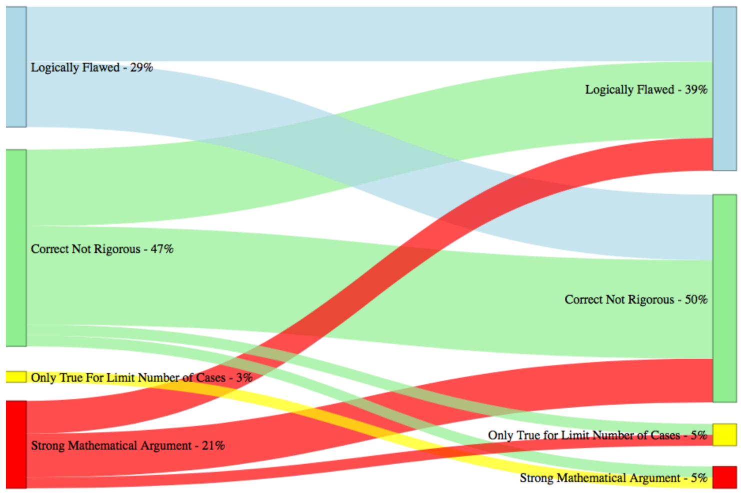

Figure 13: Sankey diagram for deductive-narrative argument – Conjecture 2

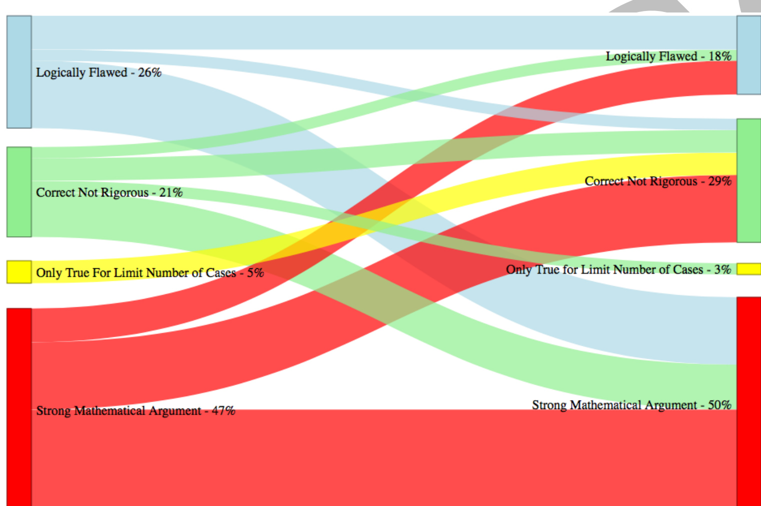

Figure 14: Sankey diagram for deductive-narrative argument - Conjecture 3