Journal of Mathematical Behavior 53 (2019) 284–303

Contents lists available at ScienceDirect

The Journal of Mathematical Behavior

journal homepage: www.elsevier.com/locate/jmathb

From examples to proof: Purposes, strategies, and affordances of example use

Zekiye Ozgur[a,*], Amy B. Ellis[b], Rebecca Vinsonhaler[a], Muhammed Fatih Dogan[c], Eric Knuth[a]

[a] Department of Curriculum and Instruction, University of Wisconsin-Madison, 225 N. Mills Street, Madison, WI, 53706, USA
[b] Department of Mathematics and Science Education, University of Georgia, College of Education, 110 Carlton Street, 105/111L Aderhold Hall, The University of Georgia, Athens, GA 30602, USA
[c] Adiyaman University, Altinsehir Mah. 3005. Sok. No:13, 02040 Merkez, Adiyaman, Turkey

ARTICLE INFO

Article history:
Received 31 January 2017
Accepted 29 March 2017
Available online 23 July 2017

Keywords:
Example use
Conjecturing and proving

ABSTRACT

Examples can be a powerful tool for students to learn to prove, particularly if used purposefully and strategically, but there is a pressing need to better understand the nature of productive example use. Therefore, we examined the characteristics of the successful and unsuccessful cases of proving in the context of a number theory task across the three student populations (middle school, high school, undergraduate), where by *successful case* we mean the ability to develop a viable justification that accounts for why the conjecture must be true. We present the characteristics of the successful and unsuccessful provers regarding the purposes, strategies, and affordances of example use, offer detailed accounts of a few illustrative cases of students' proving processes, and highlight a couple factors that seemed to have hindered the unsuccessful provers' ability to gain greater affordances from examples. We conclude with a discussion of the instructional implications of the results.

1. Introduction

Although learning to prove has been recently cherished as an important educational goal for learners of mathematics by both mathematics educators (e.g., Ball, Hoyles, Jahnke, & Movshovitz-Hadar, 2002; Knuth, 2002; Sowder & Harel, 1998; Yackel & Hanna, 2003) and policy documents (e.g., National Governors Association Center/Council of Chief State School Officers, 2010; RAND Mathematics Study Panel, 2002), it is also widely documented that students across all grade levels find proof difficult and often hold narrow conceptions of proof (Balacheff, 1988; Bell, 1976; Chazan, 1993; Coe & Ruthven, 1994; Harel & Sowder, 1998; Healy & Hoyles, 2000; Knuth, Choppin & Bieda, 2009). A line of research attempting to address this issue suggests that students ought to understand the limitation of examples as a means of justification so that they can develop an appreciation for the need for a proof (e.g., Stylianides & Stylianides, 2009; Zaslavksy, Nickerson, Stylianides, Kidron, & Winicki-Landman, 2012). However, Weber (2010a) found that the undergraduate students in his study did not find examples as sufficient for proof, yet they still had difficulty in comprehending mathematical arguments presented as proofs. Hence, Weber argues that improving students' comprehension of mathematical proofs does not rely on their awareness of the limitations of examples, and points to the need to help students improve the ways in which they process

* Corresponding author at: Department of Curriculum and Instruction, University of Wisconsin-Madison, 225 N. Mills Street, Madison, WI, 53706, USA.
E-mail addresses: zozgur@wisc.edu (Z. Ozgur), amyellis@uga.edu (A.B. Ellis), vinsonhaler@wisc.edu (R. Vinsonhaler), mfatihdogan@adiyaman.edu.tr (M.F. Dogan), knuth@education.wisc.edu (E. Knuth).

https://doi.org/10.1016/j.jmathb.2017.03.004
0732-3123/

arguments. Accordingly, we contend that understanding the limitations of examples is important, but it is not enough in itself for supporting students' proof production. Instead, we maintain that examples can be a powerful tool for students to learn to prove, particularly if used purposefully and strategically. However, students seem to be largely unaware about how to leverage examples in proving; thus, learning to use examples effectively appear to be an untapped resource for students, which can support students' proving abilities in powerful ways.

As mathematicians often make use of examples in exploring mathematical phenomena and developing and proving conjectures (Lockwood, Ellis, & Lynch, 2016; Lynch & Lockwood, 2019; Weber, 2008), students too may greatly benefit from thinking with examples in proof activities. Following this premise, Iannone, Inglis, Mejia-Ramos, Simpson, and Weber (2011) investigated whether example generation helps undergraduate students produce proofs, but they found that "simply asking students to generate examples about a concept may not substantially improve their abilities to write proofs about that concept, at least not more so than providing students with examples to read" (p. 11). The researchers concluded that further empirical research is needed to characterize the nature of example use that supports proof production. Similarly, Pedemonte and Buchbinder (2011) found that not all examples were helpful in developing proofs; some examples were only helpful for constructing a conjecture, but failed to inform how to prove the conjecture, while in some cases examples did also help students prove their conjecture. They argue that examples are helpful in developing proof only if there is a cognitive unity and structural continuity between the argumentation leading to a conjecture and its subsequent proof.

Hence, there is a pressing need for further research to unpack the nature of productive example use if we are to help students use examples purposefully and strategically, and thus, learn to prove. In our attempt to address this issue, we sought to examine the characteristics of successful and unsuccessful cases of proving in the context of a number theory task across the three student groups, ranging from middle school to undergraduate level, where by *successful case* we mean the ability to develop a viable justification that accounts for why the conjecture must be true. In dividing the data into successful and unsuccessful cases of proving, independent of the grade level of the students, we aimed to identify distinguishing characteristics between the two groups. To accomplish this goal, we drew on the Criteria, Affordances, Purposes, Strategies (CAPS) framework described by Ellis et al. (2019), and examined the patterns regarding the purposes, strategies, and affordances of students' example use. In discussing the nature of example use among the successful and unsuccessful cases of proving, we also drew on the literature on the relationship between example use and proof production. In what follows we present a brief theoretical background of the study by outlining some of the key studies that inform this study, together with a brief description of the CAPS framework.

2. Theoretical background and relevant literature

2.1. Relationship between example use and proof production

Following Goldenberg and Mason (2008), we view examples as cultural mediating tools between learners and mathematics concepts, theorems, and techniques, which are situated within the learner's understanding. Hence, examples play a vital role in learning and doing mathematics as they enable mathematical communication, either with oneself or with others. The role of examples in learning mathematics, therefore, has become an important research interest for many scholars, resulting in studies that examine the types of examples students generate (e.g., Buchbinder & Zaslavsky, 2011; Ellis, Lockwood, Williams, Dogan, & Knuth, 2012), the relationship between example use and learning of new mathematics concepts (e.g., Watson & Shipman, 2008), the nature of mathematicians' and advanced mathematics students' use of example (e.g., Alcock & Inglis, 2008; Lockwood et al., 2016), and the relationship between example use and proof (e.g., Antonini, 2003; Iannone et al., 2011; Pedemonte & Buchbinder, 2011; Rowland, 2001). Within the literature on the relationship between individuals' example use and proof development, we highlight three sets of constructs that are particularly relevant to our study: (a) attending to mathematical structure in example use, (b) relationships between types of generalizations and example use, and (c) generic examples as a bridge between empirical and deductive reasoning.

2.1.1. Attending to mathematical structure in example use

Attending to mathematical structure is often alluded to when describing successful provers' example use, and similarly, structural thinking is highlighted as the core of deductive proof and mathematical understanding (Kuchemann & Hoyles, 2009; Mason, Stephens, & Watson, 2009). Watson and Shipman (2008) define '*structure*' as "how elements and properties of mathematical expressions are related to each other" (p. 98). But, Mason et al. (2009) caution that simply detecting a relationship between two or more objects does not necessarily mean structural thinking. They describe structural thinking as "a disposition to use, explicate, and connect these properties in one's mathematical thinking" (p. 11). Watson and Shipman (2008) note that "comparison *is* the way to perceive the structures, dependencies and relationships which characterize mathematical abstraction (Davydov, 1972, p. 93)" (p. 99, emphasis original). Furthermore, they highlight the importance of variation and reflection on the generated examples for productive example use that lead to learning gains. Watson and Shipman argue that individuals need to compare similar examples with an aim to discern critical features, reflect on their example use and consider the effects of the variations they made in generating their examples. Accordingly, it seems that goal-oriented example use, which is intended to find a structure across the generated examples, is likely to generate affordances for proof development.

2.1.2. Relationships between types of generalizations and example use

The activities of conjecturing and proving through example exploration naturally involves the act of generalizing, too, and thus the generalizations students make merit special attention to better understand students' proving abilities. Researchers have identified different types of generalizations. For instance, Bills and Rowland (1999) described generalizations made from a small number of cases or observations – based on the form of the results or observed relationships – as *empirical generalizations*, whereas the generalizations made based on an example being treated as a generic representative of a class, focusing on the underlying meanings, structures or properties, *as structural generalizations*. Thus, Bills and Rowland posit empirical generalizations as inferior to structural generalizations, as the former does not provide insight into why a conjecture must be true. Harel (2001) identified a similar distinction in students' generalizations in the context of mathematical induction. In one type of generalization, students focus on the regularity in the results, which he called *result pattern generalization* (RPG), whereas in the other type students focus on regularity in the process, which he called *process pattern generalization* (PPG). Moreover, Harel underscored the link between the types of generalizations students made and their proof scheme. He observed that the students whose understanding of mathematical induction was based on RPG had an empirical proof scheme, in which students relied on "evidence from perception or examples of direct measurement of quantities, substitution of specific numbers in algebraic expressions" (Harel & Sowder, 1998). In contrast, the students whose understanding was based on PPG exhibited the transformational proof scheme, which consists of a consideration of the generality of the conjecture, goal oriented (mental) operations, and deduction (Harel & Sowder, 1998). Harel stressed that PPG requires students to attend to the underlying structure of a pattern or relationship, and thus posited it as more likely to help students develop mathematical understanding. Similarly, Pedemonte and Buchbinder (2011) underscored the importance of attending to mathematical structure when generalizing, noting that students' example exploration may yield generalizations that are based on the results of examples, devoid of an understanding of the underlying structure. This may be sufficient for conjecture development, but is often insufficient for proof development.

2.1.3. Generic examples as a bridge between empirical and deductive reasoning

Mason and Pimm (1984) described a *generic example* as a particular example that is "*presented such a way* as to bring out its intended role as the carrier of the general" (p. 287), stressing the caveat that an example is generic only if the individual sees the general structure in the particular example. Rowland (2001) points to the significance of generic examples to support students' proof processes, highlighting that generic examples provide insight about why a conjecture must be true through that particular example. Numerous researchers have discussed generic examples as a powerful tool to potentially help students transition from empirical reasoning to deductive reasoning. For instance, Pedemonte and Buchbinder (2011) argue that generic examples, which were formed based on *process pattern generalization*, enabled cognitive unity and structural continuity between argumentation and proof. Similarly, Alcock and Inglis (2008) suggest that generic examples play a key role in translating the insight gained from semantic reasoning to syntactic product, and thus advocate generic examples as a valuable pedagogical tool.

In sum, attending to mathematical structure, different types of generalizations, and generic examples emerge as key constructs in characterizing the nature of students' example use, which are also instantiated in the CAPS framework.

2.2. The CAPS framework

To better understand the role examples played in students' exploration, development, and justification of a mathematical conjecture, we draw on the Criteria, Affordances, Purposes, and Strategies (CAPS) framework, described by Ellis et al. (2019). To recap briefly, the CAPS framework offers an analytical tool to examine students' conjecturing and proving activity with regards to their criteria for choosing examples, their purposes for examples, their strategies for choosing and using examples, as well as the benefits they gain from examples when evaluating, developing, and proving conjectures. For the purposes of this paper, we will focus on three categories of the framework (affordances, purposes, and strategies), and provide a brief description of each category below.

The *affordances* category organizes various benefits students gain from their example use as they investigate a conjecture. Those affordances include *gaining insight* into a conjecture (either understanding why a conjecture must be true or seeing a structural element in the examples), *generalizing* those insights to broader cases, *supporting conjecture development* (via revising an existing conjecture or developing a new conjecture), *supporting justification development* (either in the form of a viable but incomplete proof or a complete proof), as well as *understanding the limitations of examples*.

The *purposes* category identifies the ways in which students intend to use examples. Purposes encompass a wide range of intentions, including *understanding what* the conjecture says, *testing the truth* of a conjecture, *confirming a belief* about the conjecture, *exploring the truth domain* of the conjecture, or *refuting* a conjecture. Students' example use may also be aimed at *understanding why* the conjecture is true or false, *conveying a general argument, understanding a representation* through a specific case or *illustrating a representation* to the interviewer. Or, students' intent may be simply to *placate the interviewer*.

Lastly, the *strategies* category distinguishes two types of strategies students employ: (a) strategies for *choosing* examples, and (b) strategies for *using* examples. The strategies for choosing examples include choosing a diverse set of examples (*diversity*), systematically varying one or more elements of a set of examples (*systematic variation*), or considering mathematical properties of examples (*properties*). The strategies for using examples, on the other hand, address what students do with examples once they have chosen them. For instance, students may use examples to *attempt to disprove* the conjecture, or try

2. This question involves consecutive numbers. For example, 2, 3 and 4 are consecutive numbers, but 2, 3, and 8 are not consecutive numbers.

   Tyson came up with a conjecture about consecutive whole numbers that states: If you add any number of consecutive whole numbers together, the sum will be a multiple of however many numbers you added up.

Fig. 1. The sum of consecutive numbers task, Part 1.

to see multiple cases through a structural lens, searching for a common meaningful feature across a set of examples (*structure*). Or, they may attempt to *build formality* from examples, bridging what is seen in examples to a general representation. For more detailed description of the framework we suggest that readers see Ellis et al., 2019.

3. Methods

We conducted hour-long, semi-structured, task-based interviews with 12 middle school (MS) students, 16 high school (HS) students, and 10 undergraduate (U) students majoring in mathematics-related fields. The interview items included tasks that were designed to elicit students' example use in exploring, developing, and justifying conjectures and were also intended to be accessible to students with a wide range of mathematical abilities and resources. Thus, to reduce the potential difficulty students may have with domain-specific mathematical content knowledge a task may require, which could, in turn, obstruct our understanding of the nature of the students' example use, the interview protocols consisted mainly of number theory tasks, with the exception of one geometry task in each population. More specifically, the interview protocol for the secondary students consisted of eight tasks and was identical for middle school and high school students, except for the geometry task. The interview protocol for the undergraduate students had seven tasks, including four identical tasks with the secondary protocol to allow for comparison across the three populations. For more information about the interview protocols and participants, see Knuth, Ellis, & Zaslavsky, 2019.

We acknowledge that students' proving activities are nuanced and may be context-dependent; their strategies and purposes for examples may vary depending on the nature of task and the mathematical content of the task. Accordingly, in order to minimize the potential variations that might have occurred due to other factors related to the nature of tasks, we focus on one common proving task (instead of the entire interview) to investigate the characteristics of the successful and unsuccessful provers' example use in proving. Moreover, given that the overall differences and similarities across the populations in students' example use were discussed by Ellis et al. (2019), we aim to instead provide a more focused, in-depth analysis of students' example use. Therefore, for the purposes of this paper we analyzed the sum of consecutive numbers task, as it turned out to be a fruitful task for many students and therefore provided a rich context for us to examine the characteristics of students' proving processes. Also, being the second task in the interview protocol, it was less likely that students' example use would have been influenced by their activity in previous tasks. The first task in the protocol was an incorrect conjecture, so the only potential influence could have been a sensitivity to counterexamples and thus an impetus to search for counterexamples.

3.1. The sum of consecutive numbers task

In the secondary student interviews, the interviewer presented the task (Fig. 1) to the students and reminded them what consecutive numbers meant, if needed. The interviewer then asked, "Can you give an example of how the conjecture might work if you use five consecutive numbers?" Requesting an example for five consecutive numbers was a deliberate choice to scaffold students' understanding of the general conjecture by encouraging them to think about the conjecture through a specific case first, as well as to make sure that the students understood the meaning of the conjecture before progressing further on the task. Thus, the first part of the task was an initial exploration of the conjecture, and the students' example explorations and justifications were often limited to the case of five consecutive numbers.

Once students had finished exploring the case of five, the interviewer provided them with the second part of the task, as shown in Fig. 2. This prompted students to explore the truth domain of the conjecture, often leading to breaking the conjecture into multiple sub-conjectures (e.g., the case of three consecutive numbers, etc.).

The undergraduate students were presented the same task with a different phrasing, appropriate to their level, as follows: "Tyson examined the following rule. 'If you add k consecutive numbers together, then the sum will be divisible by k.Ḧe came up with a conjecture that this rule holds for any k greater than 2. What do you think of Tyson's conjecture?'

3.2. Data analysis

All interviews were transcribed verbatim and enhanced by adding images of student work and gestures to contextualize the student's and interviewer's utterances. To be consistent with the unit of analysis across coders, we first parsed the enhanced transcripts into episodes prior to coding. When parsing into episodes, we looked for shifts in students' thinking

Tyson thinks that the conjecture will always be true no matter how many consecutive numbers you use or which consecutive numbers you choose. So he thinks that if you add any 3 consecutive numbers, the answer will be a multiple of 3, or if you add any 6 consecutive numbers, the answer will be a multiple of 6, and so on.

Do you think the conjecture is true for any set of consecutive numbers, not just when you pick five consecutive numbers?

Fig. 2. The sum of consecutive numbers task, Part 2.

or activity, either self-directed or interviewer guided. For instance, suppose a student began working on the task by testing an example, then tried another example, and then shifted to an algebraic representation. In such a case, we would have considered that activity as occurring in three distinct episodes. For each transcript, one member of the research team parsed the transcript into episodes, and then another member checked it to ensure consistency in parsing. Each parsed interview was then independently coded by two researchers via NVivo per the CAPS framework (for more detail see Ellis et al., 2019).

To answer our research question (*What are the characteristics of the successful and unsuccessful provers' example use with respect to the purposes, strategies, and affordances of their example use?),* we then divided the data into two groups as (a) *successful cases of proving* and (b) *unsuccessful cases of proving* for the sum of consecutive numbers task. Our criteria for successful cases of proving was whether a student could produce a justification that treats the general case, which includes justifications considered to be proofs as well as justifications that are viable but incomplete (i.e., justifications that treated the general case and could become a formal proof, but did not reach the level of completeness or formality to merit being coded as a proof).[1] We populated students who could produce a justification together (who we refer as *successful provers*), and similarly populated students who could not produce a justification together (referring to as *unsuccessful provers*), and examined the trends in their example use in terms of affordances, purposes, and strategies of examples in two ways.

First, we identified the percentage of students within each group who exhibited a given category (e.g., *purposes of examples*) and the codes within each category (e.g., *explore the truth domain*). Second, we determined the average frequency of each code within the groups. For example, 75% of the successful provers (12 out of 16 students) and 86% of the unsuccessful provers (19 out of 22 students) used at least one example with the purpose of *exploring the truth domain*. On the other hand, the total number of instances in which an example was used to *explore the truth domain* among the successful provers was 26, with an average of 1.63 instances per student, and it was 32 among the unsuccessful students, with an average of 1.45 instances per student. Because the group sizes are different, we report the percentage of students rather than the number of students to enable comparisons across groups. In addition, we report the average frequency of codes within each group to see if there was any pattern between groups in terms of how frequently they tended to exhibit a code. In what follows, we present the results by comparing the characteristics of the successful and unsuccessful provers.

4. Results

We report and discuss the findings in two parts. First, we consider the successful and unsuccessful cases of proving (independent of grade level), and present an overall comparison of these two groups in terms of the affordances gained from example use as well as the purposes of and strategies for their use of examples. Next, we provide detailed accounts of the students' proving processes to complement these broader level comparisons. We present two illustrative cases to exemplify the characteristics of example use observed among the successful provers. Lastly, we highlight some factors that seemed to have hindered students' abilities to gain greater affordances from their example use and, thus, develop a proof.

4.1. Comparison of successful and unsuccessful provers' example use

For the sum of the consecutive integers task, 16 students (6 MS, 4 HS, and 6 U students) were considered to be successful in developing a proof, while 22 students (6 MS, 12 HS, and 4 U students) were considered to be unsuccessful. As explained in the methods section, in order to be categorized as a successful prover a student must have produced a justification accounting for why the conjecture was true in general (which was coded as *produce a proof*), or produced a viable justification that was similar in nature to *produce a proof*, but was not complete yet (*viable but incomplete justification).* For example, Esther, a high school student, could *produce a proof* for the case of five consecutive numbers by constructing her proof based on a generic example (2+ 3 + 4 +5 + 6). She explained why the conjecture was true for *all* five consecutive numbers as follows:

[1] We use justification and proof interchangeably to refer to arguments comprised of a series of logically connected assertions that one makes to justify why a mathematical claim must be true in general.

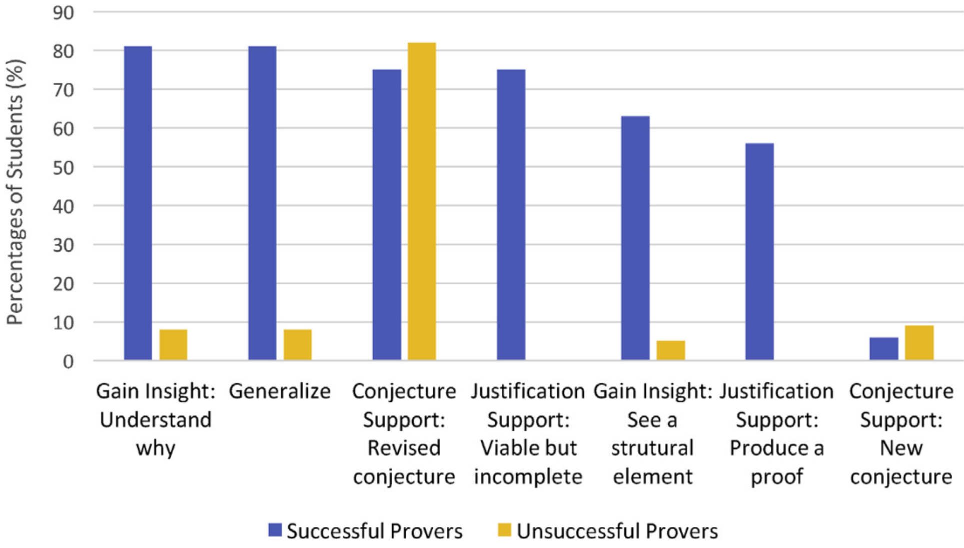

Fig. 3. Affordances of example use.

> (I) subtracted numbers from each of these numbers to get a 2. And, since there's five numbers and in each one of them (there) is 2, it's already a multiple of 5, and basically, as a number goes on, which is like, you know, (you do) plus one more, plus one more, plus one more, all of them added together is 10, and that would happen every single time, because you have to do consecutive numbers, so the difference between the first number and all the other differences together would be 10, which is also a multiple of 5. So, you would always get a multiple of 5.

In another successful case of proving, Karl, a middle school student, could develop a *viable but incomplete justification* for his revised version of the conjecture, that is the initial conjecture would be true for all odd consecutive numbers. Karl used a generic example (3 + 4 + 5 + 6 + 7 + 8 + 9) to explain his justification, too, but his justification was based on the fact that odd number strings of consecutive numbers have a middle number. He elaborated:

> You would find the center, 6. So, you would take away 3 from this (9), and add it to this (3), so then this (3) would be 6 and this (9) would be 6. Take away 2 from this (8) and add 2 to this (4); it would be 6 and 6. Take away 1 and 1, and 6, 6, 6. And it would be – if you added all of these up, that would be multiplied by 7.

Although Karl also treated the general case, we considered this as an incomplete proof because there was no additional evidence to distinguish whether Karl viewed this as a generic example of the case of seven consecutive integers or more generally as a generic example of any odd number string of consecutive numbers.

Seventy-five percent of the students deemed successful (n = 12) could develop a viable but incomplete proof, while about 42% of them (n = 5) were also able to reach a complete proof later. Twenty-five percent of the successful students (n = 4) could directly produce a complete proof. Having distinguished the successful and unsuccessful provers in this way, we then examined each groups' example use by calculating the percentage of students within each group who exhibited a given code and the average frequency of the code occurrences. We begin with presenting the trends and patterns between the two groups in terms of the affordances they gained from examples. Next, we present the trends in students' purposes of and strategies for example use, respectively, and then discuss the interplay of the trends across the three categories (i.e., affordances, purposes, and strategies).

4.1.1. *Affordances of examples*

Whether successful or not in developing a proof, most of the students (at all grade levels) engaged in a variety of example use during their proving-related activities (e.g., exploring, developing, and justifying conjectures). A distinguishing feature between the two groups, however, was the extent to which they gained valuable affordances from their use of examples in those activities. In fact, we found that the successful and unsuccessful provers looked most different in terms of the affordances they gained from their example use. Fig. 3 presents the percentages of the students (y-axis) within each group and the different affordances gained as a result of their example use (x-axis). The types of affordances were ordered from the most common to the least common among the successful provers, showing the distribution of the types of affordances more visibly.

Fig. 3 also reveals an interesting finding that while the successful provers' example use frequently enabled them to *gain insight* into why the conjecture must be true, to *generalize* their understanding, to *revise* the given conjecture, and ultimately, to *prove* the conjecture, the unsuccessful provers demonstrated very few of these affordances (with the exception of their example use activity resulting in a *revision of the conjecture*). Thus, in addition to their ability to develop viable

Table 1
Distribution and frequency of affordances gained from examples.

Affordances	Percentage of Students		Average Frequency of Code Occurrences	
	Successful Provers (n = 16)	Unsuccessful Provers (n = 22)	Successful Provers (n = 16)	Unsuccessful Provers (n = 22)
Understand why	81%	8%	1.69	0.18
Generalize	81%	8%	1.5	0.23
Viable but incomplete proof	75%	0%	1.31	0
Revise the conjecture	75%	82%	1.13	1.41
See a structural element	63%	5%	1.25	0.05
Produce a proof	56%	0%	0.94	0
Develop a new conjecture	6%	9%	0.06	0.14
Understand limitations	0%	0%	0	0
TOTAL	100%	82%	11.31	2.86

Table 2
Distribution and frequency of the purposes of examples.

Purposes	Percentage of Students		Average Frequency of Code Occurrences	
	Successful Provers (n = 16)	Unsuccessful Provers (n = 22)	Successful Provers (n = 16)	Unsuccessful Provers (n = 22)
Test the truth	94%	95%	1.94	2.23
Convey a general argument	81%	14%	1.63	0.23
Explore the truth domain	75%	86%	1.63	1.45
Placate the interviewer	63%	73%	1.06	1.27
Belief confirmation	50%	55%	0.63	1
Convey the claim	38%	23%	0.69	0.36
Understand what	31%	68%	0.31	0.73
Understand a representation	25%	9%	0.63	0.27
Illustrate a representation	19%	5%	0.25	0.05
Understand why	13%	5%	0.13	0.09
Refute	13%	14%	0.13	0.14
Conjecture development	6%	14%	0.06	0.18
TOTAL	100%	100%	9.06	8

justifications, the successful and unsuccessful provers also differed significantly in terms of the affordances of *gaining insight* and *generalizing*. This is fairly expected, as the affordances of gaining insight and generalizing are ones that can often foster the ideas that lead to generic arguments, deductive arguments, and proofs.

In particular, while all successful provers gained multiple affordances from examples, 18% of the unsuccessful provers did not gain any affordance at all. As shown in Table 1, many of the students who were able to produce a justification had gained insight from examples through *seeing a structural element* (63%) and *understanding why* the conjecture was true (81%), and also *made generalizations* (81%), while very few students among the unsuccessful provers showed gains from those affordances (5%, 8%, and 8%, respectively). For the unsuccessful provers, it seems that their example use did not offer critical insight in most cases, and as a result, they were typically unable to make progress toward developing generalizations and justifications. For both groups, the affordance of *gaining insight* (both understanding why and seeing a structural element) and the affordance of *generalizing* were closely linked; the occurrences of these codes were concurrently high among successful provers and concurrently low among unsuccessful provers. Hence, the findings corroborate that gaining insight and generalizing, which were often precursors to the students' justifications, are critical for the successful development of a proof. In particular, *gaining insight* is reminiscent to Raman's (2003) notion of "key idea" that shows why a claim is true, and thus gives a sense of both conviction and understanding.

Table 1 presents further details about the trends in the affordances gained from example use among the successful and unsuccessful provers, showing the average frequency of code occurrences as well as the percentages of students who gained affordances within each group. Overall, successful provers in this task gained 11.31 affordances on average, whereas unsuccessful provers only gained 2.86 affordances on average. Furthermore, about half of the affordances for the unsuccessful provers came in the form of making revisions to the conjecture. In contrast, the average frequency of the affordances successful provers gained was quite evenly distributed across *understanding why* (1.69), *generalizing* (1.5), *producing a viable but incomplete proof* (1.31), *seeing a structural element* (1.25), *revising the conjecture* (1.13), *and producing a proof* (0.94). These striking differences between two groups raise the question of what accounts for them. We turn now to the students' purposes of and strategies for example use, respectively, to address this question.

4.1.2. Purposes of examples

Table 2 shows the percentage of students within each group who used an example for a given purpose and the average frequencies of code occurrences for that purpose. The two groups looked similar in terms of the average number of examples they used, as the successful provers used 9.06 examples on average while the unsuccessful provers used 8 examples. However,

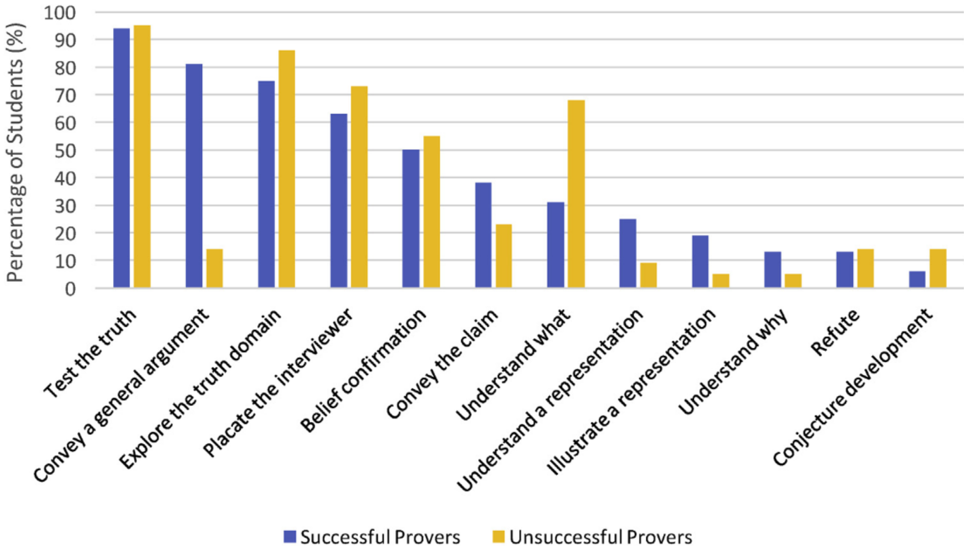

Fig. 4. Purposes of example use.

the distribution of the various purposes that their example use served within each group reveals some interesting differences as well as similarities (see Fig. 4 for a visual representation of the distribution). First, the most common purpose of example use in both groups was to *test the truth* of the conjecture (94% and 95%, respectively), with similar average frequencies across the groups (1.94 vs 2.23). Similarly, the students in both groups commonly used examples *to explore the truth domain* of the conjecture (75% of the successful provers, and 86% of unsuccessful provers), again with similar average frequencies (1.63 vs 1.45).

In contrast, the two groups differed remarkably in the use of examples for *conveying a general argument* as 81% of the successful provers used an example to convey an argument supporting his or her assertion, that is a *generic example*, whereas only 14% of the unsuccessful provers used an example for that purpose. Furthermore, the average frequency of this code was also noticeably higher among the successful students: While the successful students used 1.63 examples on average to convey a general argument, the unsuccessful students used 0.23 examples on average for that purpose. Yet, this stark difference in the use of examples for conveying a general argument is not very illuminating for explaining the differences between the groups in terms of the affordances they gained from examples, as conveying a general argument typically requires students to have already gained some insight into why the conjecture must be true.

Another considerable difference was observed in the use of an example to understand the meaning of a conjecture (*understand what*). Although not many successful provers (31% of them) needed to use an example to make sense of the conjecture, 68% of the unsuccessful provers used an example for that purpose, underscoring a potential difficulty that unsuccessful provers had in understanding the meaning of the conjecture, including the mathematics involved in the given conjecture. On the other hand, there was a decreasing trend in students' use of examples for *understanding what* the conjecture was saying when shifting from the middle school to the undergraduate population; while 67% of the MS students used an example to *understand what*, only 20% of the undergraduate students did so, which is likely due to students' increased experiences with mathematics and more familiarity with understanding how mathematical ideas are expressed – a consistent trend with the overall comparisons across the grades bands (). This decreasing trend may also be explained by differences in students' "personal example spaces and technical tools" that Sandefur, Mason, Stylianides and Watson (2013) discussed as one of the qualities of students that contribute to productive example use in proving. That is, the undergraduate students' more expanded set of "technical tools" might have supported their understanding of the conjecture, and thus, unlike the many middle school students, they did not need to use an example to understand what the conjecture was saying.

Shifting our attention to some of the less frequently used purposes of examples, three other purposes emerge as somewhat different between the groups: to *understand a representation*, to *illustrate a representation*, and to *understand why*. Even though the use of these three purposes were not very common in either group, we consider the distinctions between the groups helpful, to some extent, to account for the differences in the affordances gained from their use of examples. Both the purposes of *understanding a representation* and *illustrating a representation* indicate an attempt to shift from examples to the general by representing an argument algebraically or visually. Students may often use an example to understand a representation as they figure out how a representation relates to the given conjecture; once they determine that relationship, they may use an example to convey the relationship to others (the interviewer in this context). Thus, these purposes were often in place when students' example exploration included a back and forth activity between examples and (an attempt of) algebraic representation – this is reminiscent to the back-and-forth activity in which mathematicians engage, with examples serving to illuminate their ongoing proof development (Lockwood et al., 2016). We found that the middle school students' nature

Table 3
Distribution and frequency of the strategies for example use.

Strategies	Percentage of Students		Average Frequency of Code Occurrences	
	Successful Provers (n = 16)	Unsuccessful Provers (n = 22)	Successful Provers (n = 16)	Unsuccessful Provers (n = 22)
Strategies for example choice	**75%**	**86%**	**2.20**	**2.22**
Systematic variation: Initial	56%	64%	0.88	0.86
Systematic variation: Continuation	50%	36%	0.69	0.41
Properties	38%	36%	0.50	0.45
Diversity	13%	41%	0.13	0.50
Strategies for example use	**88%**	**32%**	**2.20**	**0.56**
Structure	38%	18%	0.44	0.23
Informal induction	38%	5%	1.06	0.05
Building formality	25%	9%	0.25	0.14
Jumping to formality	13%	5%	0.19	0.05
Attempt to disprove	13%	5%	0.13	0.09
Pattern search	13%	0%	0.13	0
TOTAL	100%	91%	4.38	2.77

of example use did not typically include exploiting examples to understand a representation or to illustrate one, while most of the successful undergraduate students did employ examples to understand a representation. Thus, while acknowledging that the more abstract task formulation used in the undergraduate protocol may have encouraged such example use among the undergraduate students, it also seems that as students' mathematical experiences increased, their purposes of examples were enriched with more sophisticated use of examples. This trend was also compatible with the increasing progression observed in students' strategies that relied on general representations, as the students progressed from middle school to undergraduate level.

Lastly, using examples to *understand why* a conjecture is true is also important as this curiosity to understand the underlying reasons that make the conjecture true often led to more strategic example use. Thus, the higher frequencies of these codes potentially point to more productive example use. However, to our surprise, the use of examples for the purpose of *understanding why* was quite low even among the successful provers. This is particularly interesting for two reasons. First, from a theoretical perspective, one would think that a deliberate search for understanding the underlying reasons would trigger more thoughtful and strategic example use, which would then yield to more affordances. Second, from an empirical perspective, in the overall comparisons across the grade bands we have found a clear increasing progression (from MS to U students) in the use of examples for *understanding why*, parallel to the progression seen in the development of proof (), hinting to a potential link between the two. Thus, at first glance, the low occurrence of the *understand why* code among the successful provers in this conjecture may be seen as a contradiction to our hypothesis that example use with the purpose of *understand why* is likely to contribute to proof development. But, we rather see this as an indication of the intricate interrelationships between the purposes, strategies, and affordances of examples. In other words, these findings together suggest that there may not be a straightforward relationship between any specific purpose, strategy, and successful proof development; rather, it is the dynamic interactions among purposes, strategies, and affordances of examples that guides and supports productive example use. We suspect that the successful provers for this particular conjecture were already able to gain useful insights from their initial example exploration, perhaps because the conjecture was easy enough for them to make sense of from their initial examples alone. Hence, we contend that if students had already gained insight into why the conjecture might be true from their initial examples that were used for other purposes (or from algebra), then they may not have needed to try additional examples with the deliberate purpose of understanding why.

4.1.3. Strategies for choosing and using examples

As outlined in Section 2.2, we identified two types of strategies in students' example exploration and proving activities: (a) strategies for *choosing examples* and (b) strategies for *using examples*. This distinction between the strategies for choice and use proved to be helpful in characterizing the proving processes of the successful and unsuccessful provers. The distribution of the percentage of the students within each group who used a given strategy, as well as the average frequencies, are presented in Table 3 (see Fig. 5 for a more visual representation of the distribution).

Comparison of the two groups in terms of the overall percentage of the students who used at least one strategy does not show much difference (100% versus 91%). However, the average number of strategies the successful provers used (4.38) was almost double the unsuccessful provers' average frequency (2.77), suggesting that perhaps the successful provers' more frequent use of different strategies might have helped them benefit from their example use more, and thus led to gaining more affordances, including the development of a proof. Taking a closer look at the distribution of the *strategies for example choice* and *strategies for example use* within each group reveals more telling differences. As seen in Table 3, 88% of the successful provers employed a strategy for *using* examples, and 75% of them employed a strategy for *choosing* examples. In contrast, only 32% of the unsuccessful provers employed a strategy for *using* examples although 86% of them employed a

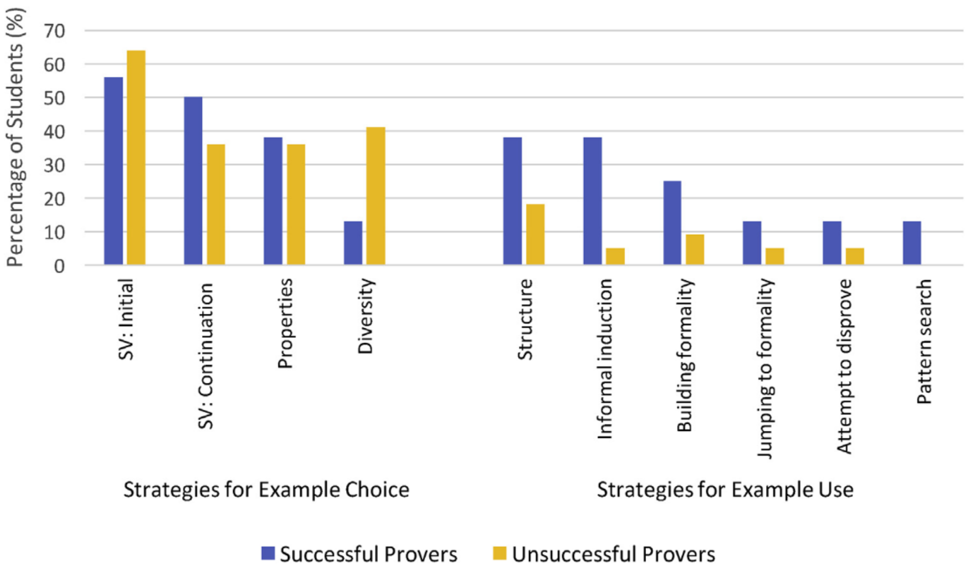

Fig. 5. Strategies for example use.

strategy for *choosing* examples. In this case, it seems that the unsuccessful provers, in general, were quite strategic about how to choose their examples, but they lacked the strategies for how to effectively use their examples.

Moreover, the successful provers' total average frequencies of the strategies were evenly split between the strategies for *choosing* examples (2.2) and the strategies for *using* examples (2.2), indicating that the successful students were strategic both in choosing their examples and subsequently using them to develop a proof. In contrast, about 80% of the unsuccessful provers' strategies were for *choosing* examples, and thus only about 20% of them were for *using* examples (total average frequency of 0.56). This discrepancy further corroborates that the unsuccessful provers' strategies were mainly limited to *choosing* examples and that they did not seem to know how to effectively leverage their chosen examples to shift from examples to a general argument.

In terms of the strategies for example choice, both the successful and unsuccessful provers similarly chose their examples by systematically varying one or more element of each successive example (*systematic variation*) and considering the mathematical properties of examples (*properties*). While 41% of the unsuccessful provers also strategically tested a diverse set of examples to justify the conjecture (*diversity*), it was not a common strategy among the successful provers, as only 13% used that strategy. Given that the unsuccessful provers did not typically have other strategies for *using* examples at their disposal, they seem to have relied more heavily on this strategy. Although showing that a conjecture holds true for a diverse set of examples does not prove that the conjecture is *always* true, it nevertheless seemed to serve to convince students that it is likely true. In fact, this tendency to test a diverse set of examples is quite common among students (Balacheff, 1988; Chazan, 1993; Harel & Sowder, 1998; Knuth, Choppin, Slaughter, & Sutherland, 2002; Knuth et al., 2009). Balacheff (1988) described one aspect of this behavior, which he called "crucial experiment", as a tendency to test extreme or "random" cases as an attempt to prove a given mathematical statement. This approach heightens students' confidence in the truth of a conjecture, because students feel assured that they have at least considered a variety of cases, and thus reduced the risk of the outcome holding true only for a limited set of cases. The strategy of *diversity*, however, is typically less fruitful for providing insight into why a conjecture is true, compared to the strategies of *systematic variation* and *properties*, especially if the sole purpose in diversification of examples is to gain conviction of the truth of a conjecture on the account that it passes the veracity test for various cases.

What seems to be a key distinction between the successful and unsuccessful provers is the nature of the strategies employed in *using* examples. As shown in Fig. 5, the unsuccessful provers very minimally engaged in strategies for using examples, whereas the successful provers more frequently used one or more strategies. The successful provers searched for a pattern or a mathematical structure through a set of examples and attempted to use the set of examples to identify general features (*structure*), attempted to show that the conjecture was true in general by arguing that all cases of a string of n consecutive numbers could be built on a base case (*informal induction*), and used examples to develop a formal representation as an expression of what was the same across all examples (*building formality*). Although less common, some of the successful provers also assumed the conjecture untrue and tried to disprove it with examples prior to trying to prove it true (*attempt to disprove*). The successful provers also showed higher evidence of the less productive strategies. For instance, some successful provers attempted to find similar features or patterns from prior examples without consideration of any structural or explanatory elements (*pattern search*), or they started with a formal representation prior to using any examples or abruptly introduced a formal representation unrelated to the examples considered (*jumping to formality*). Although pattern search and jumping to formality are not necessarily productive strategies, nonetheless, taken together those strategies show that the successful provers' example use in proving was characterized by an attempt to find a common structure across examples that could be used for a general argument. In contrast, the unsuccessful provers did not engage in those strategies much, which may point to either a lack of curiosity for how to demonstrate the conjecture's truth or a lack of experience in such strategies.

Comparing findings across grade bands revealed a decreasing trend in the strategies for *choosing* examples from secondary to undergraduate students (MS: 83%, HS: 88%, U: 70%), while there was an increasing trend in the strategies for *using* examples (MS: 42%, HS: 50%, U: 90). While students at all grade levels employed all four types of strategies for *choosing* examples, the number of types of strategies for *using* examples the students used in each population increased from middle school to undergraduate level. Specifically, the MS students employed only two types of strategies for using examples (*structure* and *informal induction*), whereas the HS students employed 4 (all strategies but, *jumping to formality* and *pattern search*) and the U students employed 5 strategies for using examples (all but, *informal induction*). Interestingly, the strategy of *informal induction* was unique to the secondary students. We suspect that because the secondary students overall had limited facility with algebra compared with the undergraduate students, they therefore devised other creative ways of thinking with examples, one being the strategy of informal induction. These patterns should be considered promising, as the students appeared to have become more strategic in using examples as they progressed through the grade levels, which also may account for the increasing progression observed in developing a proof.

4.1.4. Summary

In the context of the consecutive integers task, we examined the characteristics of the successful and unsuccessful provers' example use, revealing some interesting distinctions between the two groups. First, in addition to being able to develop a justification, the two groups were markedly different in terms of the other affordances they gained from examples. Particularly, we found that the successful provers' example use generally enabled them to gain insight into the conjecture, either helping them see a structural element or understand why the conjecture was true, and to generalize and revise conjectures. On the other hand, the unsuccessful provers' gains from examples were often limited to revising the given conjecture. The interplay between the purposes and strategies of examples appeared to influence the provers' subsequent example use, which accounts for the striking differences observed in the affordances gained.

The most distinct purposes of examples between the groups were the purposes of *conveying a general argument* and *understand what*, the former being a characteristic purpose of the successful provers, while the latter was a common purpose among the unsuccessful provers. The successful provers also slightly more commonly made use of examples to *understand* and *illustrate representations*, as well as to *understand why*, compared to the unsuccessful provers. However, the distribution of the types of strategies employed by the two groups revealed the most interesting and stark differences. We found that the unsuccessful provers were quite strategic for *choosing* examples, but lacked the strategies for *using* examples, as only 20% of their strategies were of the latter kind. Thus, the unsuccessful provers relied more heavily on the strategy of *diversity* to convince themselves of the truth of the conjecture. In contrast, the successful provers were strategic both in *choosing* and *using* examples, as their strategies were evenly split between the two kinds of strategies. Unlike the unsuccessful provers, the successful provers could leverage their examples more effectively through the strategies of looking for *structure*, *building formality*, and *informal induction*.

4.2. Exemplifying students' proving processes: from examples to proof

Now that we have presented the broad level comparisons of the characteristics of the successful and unsuccessful cases of proving, we offer detailed accounts of a few illustrative cases of students' proving processes. First, we share the work of one high school student and one middle school student and describe how these students' purposeful and strategic example use led to various affordances, as well as success in developing a viable justification. We also compare the ways in which the different strategies the students used afforded and impeded each student in developing a proof. We follow with highlighting a few factors that seemed to have hindered the student's ability to gain greater affordances from examples, and thus to develop a proof.

4.2.1. Illustrating successful cases of proving process: the case of Rehnuma and Caden

In this section, we first present the work of a HS student, Rehnuma, to exemplify a wide variety of example use, highlighting the strategies of *structure* and *building formality* that she employed in her example use. Then we examine the work of a MS student, Caden, as another case of productive example use, highlighting the strategy of *informal induction*. In doing so, we aim to illustrate the use of the three strategies that distinguished the successful provers' example use overall, and also to show how different strategies led to different affordances, emphasizing the importance of attending to students' mathematical resources available to them at each grade level for their proving activities.

4.2.1.1. Rehnuma's proving process: the strategies of building formality and structure. At the time of the interview, Rehnuma had recently finished ninth grade and had taken algebra and geometry courses. She was quite adept at finding commonalities and generalizing patterns. Rehnuma explored the sum of the consecutive integers task case by case, first exploring the case of five consecutive numbers and then the case of three, and proving each case true. At the request of the interviewer, she then explored the case of four, followed up with the case of six without prompting, and gained a key insight into the conjecture

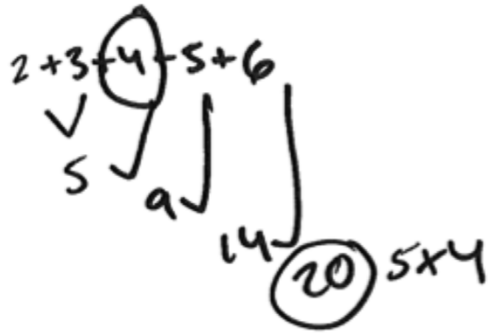

Fig. 6. Rehnuma's first example.

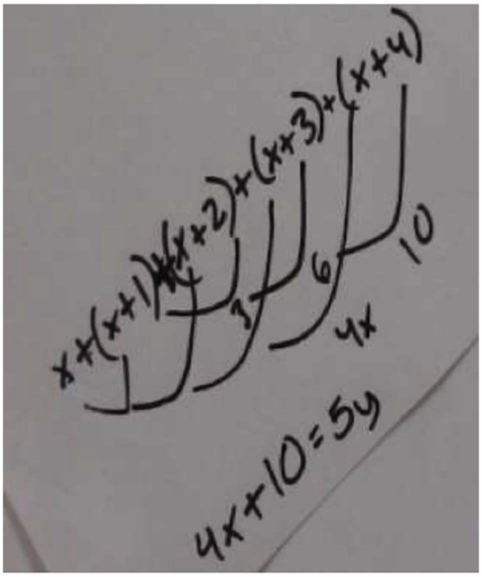

Fig. 7. Rehnuma building formality from the examples.

that led her to correctly revise the original conjecture, and to develop a new conjecture for the sum of n consecutive integers when n is even. Below we discuss some of the highlights of her proving activity.

The interviewer presented Rehnuma the first part of the task: "Tyson came up with a conjecture about consecutive numbers that states, if you add any number of consecutive numbers together, the sum will be a multiple of however many numbers you added up. Can you give an example of how the conjecture might work if you use five consecutive numbers?" Rehnuma picked 2 + 3 + 4 + 5 + 6 as her first example, to *placate the interviewer* and to *understand what* the conjecture was saying. As seen in Fig. 6, Rehnuma used the example to connect the premises and the claim of the conjecture; she added up the consecutive numbers and then wrote the sum (20) as a product of 5 (the number of consecutive numbers added) and 4. Note that the number 4, the middle term, was not circled at this point.

Rehnuma was careful not to make a general claim about the truth of the conjecture based on one example, and thus continued exploring the conjecture with another example (11 + 12 + 13 + 14 + 15), this time choosing a two-digit number to *diversify* the set of examples she tested. After two confirming examples, the interviewer asked whether she believed the conjecture was true for any five consecutive numbers, which led her to shift to algebra in responding to the question. Rehnuma replied, "Actually you can make a formula for this, x plus x plus 1 plus...", pausing to put parentheses around x + 1. She continued: "Or, no. x + (x + 1) + (x + 2) + (x + 3) + . . . one, two, three, four. . .", counting the terms that she had written (Fig. 7). She *built formality* in the sense that she generalized her activity from those two examples and represented how she thought about the examples more formally with algebraic notation.

Rehnuma attempted to express the conjecture generally (for the case of five consecutive numbers) by algebraically representing the premise of the conjecture as the sum of any five consecutive numbers, and representing the claim as 5 times a variable: "And then it's one, two, three, four *x*'s and three, six, ten. So 4*x* plus10 um, is equal to 5 times I guess another variable." However, she made a mistake in counting the number of *x* terms and incorrectly represented the sum of the string of five consecutive numbers as $4x + 10$ rather than as $5x + 10$. The interviewer asked Rehnuma to explain her

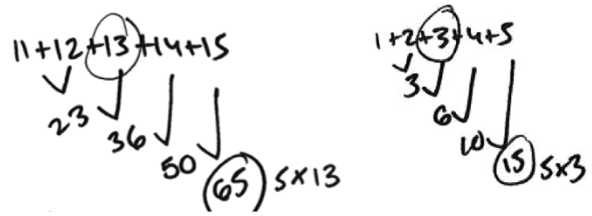

Fig. 8. Rehnuma referring to previous examples to illustrate representation and to test her theory.

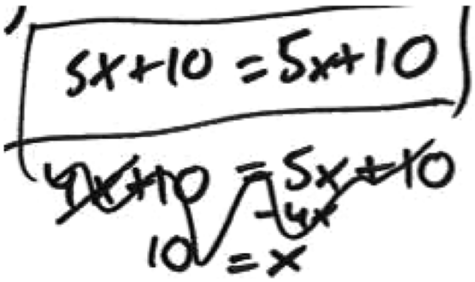

Fig. 9. Rehnuma's proof for the case of five consecutive numbers.

work, hoping that explaining would enable her to notice the error. In response, she referred to the first example she used, 2 + 3 + 4 + 5 + 6, to *illustrate her algebraic representation* to the interviewer, which helped her recognize that what she called the "other variable" (and represented with y) was the middle number. Thus, Rehnuma began to *gain insight* into the role of the middle number. She explained,

> So, if x is the first number you start with, each number has to be consecutive, so x plus 1 which like – 2, 2 plus 1 is 3 (*pointing back to 2 + 3 + 4 + 5 + 6*), and then x plus 2, 2 plus 2 is 4, um, until you have all five numbers and then I– you don't need the parentheses, but I combined the x's and then the numbers, and that's $4x$ plus 10. And then you know it has to, according to the conjecture, it has to equal 5 times a certain number, but, oh– the number will be the middle term.

The activity of relating her algebraic expression to a specific example enabled Rehnuma to *see structure* in the example that she did not see before; she recognized the string of 5 numbers as having a middle number, and that the sum equaled 5 times the middle number. With this insight, Rehnuma adjusted her algebraic representation accordingly and wrote $4x + 10 = 5 \cdot (x + 2)$, still unaware of her error of miscounting the number of x terms. She then referred to her previous example (11 + 12 + 13 + 14 + 15) to *confirm her belief* that the sum was a multiple of the middle number. Having confirmed her insight, Rehnuma then used another example, 1 + 2 + 3 + 4 + 5, to *convey her argument* to the interviewer (Fig. 8). Thus, in revisiting her previous examples, she *tried to see examples through a structural lens*, looking for a common structure in both examples in relation to the claim of the conjecture, and gained an important insight into the conjecture.

In order to explain further what x meant in her representation, Rehnuma used another example to *illustrate her algebraic representation*, which helped her realize that she made a calculation error. She was then able to show that the sum of five consecutive numbers was equal to five times the middle number.

> If x is equal to 10 and you plug it back in, that's 4 times 10 which is 40, plus 10 is 50. And then, um, oh no, it doesn't work. Oh, no! Hmm. Because there are $5x$'s, not 4. There, that was my mistake. Um... So, that's– that's 5, which is 50, and then plus 10, so that's 60. And then... That would have to change (*crossing out work*), so it would be $5x$ plus 10 is equal to $5x$ plus 10. There we go, that makes more sense. So, that proves that it works, because $5x$ plus 10 is equal to $5x$ plus 10 (Fig. 9).

Pointing to the left side of the equation, Rehnuma elaborated, "This side is setting up the consecutive numbers as they were with variables. And, this side –", pointing to the right side, " – is going by the idea that whatever the um, the sum is, is going to be 5 times the middle value." Thus, she showed that the conjecture was true for any 5 consecutive numbers, albeit not knowing yet why the conjecture was true. Hence, Rehnuma could develop a viable but incomplete justification for the case of five consecutive numbers. Her strategy to use examples for *building formality*, together with *trying to see examples through a structural lens*, enabled her to *gain insight* into the conjecture. Rehnuma made use of her initial examples, repurposing them as she continually shifted back and forth between her algebraic representation and the examples; and thus, the insight she gained from the examples supported developing her algebraic expression, and her increasing understanding of the algebraic representation boosted the affordances she gained from the examples vice versa. But, because her insight was a result of

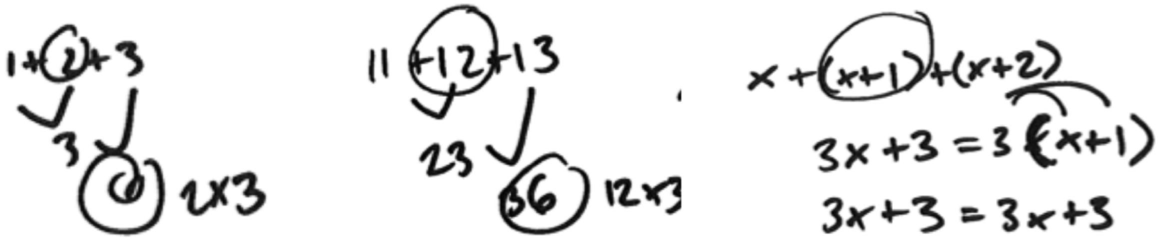

Fig. 10. Rehnuma's proof for the case of three consecutive numbers.

only noticing a commonality between the examples – *result pattern generalization* in Harel's (2001) terms – Rehnuma did not know why the conjecture was true, or why the middle number times five equaled to the sum.

During the second part of the task, Rehnuma *explored the truth domain* of the conjecture, beginning with the case of 3 consecutive numbers. She *generalized* her way of thinking with the string of 5 consecutive numbers to the string of 3 numbers by extending her reasoning to a new case. Rehnuma drew on the insight she gained from her previous examples and employed the same strategy of *diversity* for choosing examples and of *building formality* for using examples, as shown in Fig. 10. Thus, her meaning for the examples (the strings of 3 numbers) were influenced by how she began to think of the strings of 5 numbers, as having a middle term. Without prompting, Rehnuma shifted to algebra again, and repeating the same structure, she showed that the sum was equal to 3 times the middle term, and thus the conjecture was true for any 3 consecutive numbers. Hence, Rehnuma was able to develop a similar justification for the case of three consecutive numbers. Alcock and Inglis (2008) assert that students rely on the same strategy if it happens to be useful to them, which is what we see with Rehnuma maintaining the same set of strategies in her example use following the case of 5 consecutive numbers.

As she continued exploring the truth domain of the conjecture, Rehnuma found that the conjecture was not true for 4 consecutive numbers. Reflecting on why the conjecture did not hold true for that case, she *gained insight*, realizing that there was no middle value for even number strings of consecutive numbers, which enabled her to correctly *revise the conjecture*: "So, it needs to be an odd amount of consecutive numbers." In order to *confirm her hypothesis*, Rehnuma tested another example for the case of 4 consecutive numbers to see whether this new example would conform to the same set of relationships she had identified with the prior example. Thus, she attempted to find a structural reason to account for why the conjecture was not true for the case of four, in particular, and for even strings of cases, in general. Therefore, Rehnuma's example use shifted from a particular example to a generic example. In the end, Rehnuma could revise the original conjecture to account for when it would be true, but she was unable to prove why the conjecture must be true for odd number of cases in general.

Rehnuma's proving process exemplifies the various purposeful (*convey a general argument, illustrate a representation*) and strategic (*building formality, structure*) example use that we found as distinguishing characteristics of the successful provers in this study. Most distinctly, Rehnuma attempted to build formality from the examples she used, and she was successful at shifting to a general representation. This was possible because Rehnuma chose a diverse set of examples early on by systematically varying the digit numbers of her examples. Reflecting on her examples, considering the effects of variation in her examples, and trying to see a structure, Rehnuma gained insight into the conjecture and could develop a viable justification, though not a complete proof.

4.2.1.2. Caden's proving process: the strategy of informal induction. We now turn to Caden, a rising eighth grade student who was intending to take a geometry course the following year, to illustrate another successful case of proving. Caden employed a different strategy than Rehnuma, but one that was quite common among the secondary students (especially among the MS students): the strategy of *informal induction*. Specifically, Caden saw a structural element in the very first example he used and subsequently constructed all of his examples to conform to that structure. Beginning with the case of 5 consecutive numbers, Caden first showed that the conjecture was true for the base case $(1+2+3+4+5)$. He then argued that all strings of 5 consecutive numbers could be formed by adding 1 to each quantity in the series, thus adding 5 to the sum of the base case (15) for each iteration. Showing that each addend was a multiple of 5, Caden proved that the sum of them was a multiple of 5 too, and therefore showed that any string of 5 consecutive numbers must be divisible by 5. Like Rehnuma, Caden also applied the same strategy to explore the other cases, gained insight from his example use, and appropriately revised the original conjecture to account for when the conjecture was true and when it was not.

In response to the interviewer's request to show how the conjecture worked for the case of 5 consecutive numbers, Caden picked $1+2+3+4+5$, finding that the sum was 15, thus divisible by 5. He stated, "That's true for this one." The interviewer asked if he thought the conjecture would be true for any 5 consecutive numbers, to which Caden replied, "Yes", and explained, "because if you add 1 to each of these numbers it will be 5 more than 15." The fact that he confidently responded that the conjecture would be true for any 5 consecutive numbers and supported his claim with an explanation suggests that Caden could view this relationship holding true in general (*seeing a structural element*). Prompted by the interviewer, Caden provided another example, $(2+3+4+5+6)$, to *convey his argument*. He *systematically varied* his initial example, adding 1 to each number in the set to get the next example. Note that this was the seed of the strategy of *informal* *induction*, which relies on constructing a general argument based on showing the truth of a base case and then extending the reasoning to build all possible sets.

The interviewer followed up, pressing Caden to explain why the conjecture would always work, and asked him, "What if I gave you, I don't know... 16, 17, 18, 19, 20. How do you know that it's going to work?". In response, Caden tested 16 + 17 + 18 + 19 + 20. Instead of applying his reasoning to explain why this example would work, he added up the numbers and divided the sum by 5. In other words, Caden used this example to *test the truth* of the conjecture. We argue that Caden could not extend his "add 1 to each digit" idea to this example, because the leap was too large from 2 to 16. Nevertheless, he believed that the conjecture was true for any 5 consecutive numbers. We suspect that Caden's first example had enabled him to see a general idea, and the other examples helped him to confirm his idea, but he was not yet able to articulate his insight. We consider this significant as it shows that multiple examples may not be always necessary to gain insight; what is important is the generality of the example, which comes from reflectively abstracting on one's activity (Piaget, 2001), that is powerful, convincing, and can support deductive argumentation.

The interviewer then shifted the focus to the general case: "Now thinking back to the conjecture in general, what if you looked at any number of consecutive numbers... Do you think the conjecture will be true for any set of consecutive numbers?" Caden responded, "No", stating that it would not work for two consecutive numbers. We suspect that Caden mentally tested examples and found a counterexample right away, and based on this counterexample he concluded that the conjecture was not true in general. In response to the interviewer's query to why he thought that the conjecture was not true, Caden provided a general argument in relation to the parity of numbers: "For 2 [consecutive numbers], it would be an even and an odd number. So, it wouldn't be a multiple of 2 since it would be an odd number." Hence, for the string of 2 numbers, Caden could form a generic argument without ever having articulated an example to the interviewer.

The interviewer encouraged Caden to modify the conjecture so that it would be true, leading him to *explore the truth domain* of the conjecture. After thinking for a while, Caden responded that "3 would work" and explained with an example: "Well, 1, 2, 3 is a multiple of 3. So, if you added 1 or the same amount of number to each of them then it would." Note that he extended his reasoning to a new case, the case of 3 consecutive numbers, and employed the same strategy, *informal induction*. Caden argued, "So, 1, 2, 3, so that equals 6. And if you added 1 to each of these, it would be adding 3 to the whole equation, so it would be still a multiple of 3. So, if you did that for any amount of times, it would still be a multiple of 3." Caden's thinking with this example, 1 + 2 + 3, enabled him to bootstrap to an even more general argument. He was no longer restricted to thinking about just adding one number to each term in the string; he could now think about adding the "same amount of number" to each. Caden was then able to justify his reasoning for the case of 3; he knew that the sum would increase by $3n$ if each term was increased by n units, and therefore the sum would still have to remain a multiple of 3. Note that Caden began to justify for a string of 5 with a limited explanation, and as he had continued to reflect on this idea of adding 1 or the same number of units to each term in a string, his thinking had become more general and resulted in a more explicit justification. This did not arise from using many examples; it was instead due to his ability to reflect on his activity from the very first example and his focus on the process, rather than the results. In other words, Caden's justification was based on *process pattern generalization* (Harel, 2001).

To further probe Caden's thinking, the interviewer asked, "Using your reasoning, what is it that you are doing to this example?", and suggested that he explain by using 10 + 11 + 12 as an example. Though the example was provided by the interviewer, Caden used it *to convey a general argument*. He argued, "So, you're adding 9 to each of these, so 9 times 3 is 27, and that's a multiple of 3. And this is a multiple of 3, so if you're just adding those two together, then it's a multiple of 3." Thus, this time when the interviewer jumped to a large example, Caden could now make that leap because he had generalized from adding 1 unit to each term to adding n units to each term. He had reflected and generalized his thinking, focusing on the process (PPG), to the point where he could make the leap he was unable to make before.

Caden continued to *explore the truth domain* by testing the case of four consecutive numbers by using 1 + 2 + 3 + 4, and concluded that, "Four wouldn't work, because 1, 2, 3, 4 is 10 and that's not a multiple of 4." He offered another example, 3 + 4 + 5 + 6, *to convey his argument,* and explained, "Then using the same thing for 3 it wouldn't work... So, like adding a certain amount of number to each one." He added, "Because then it will always be two, two too many to be a multiple of 4." Apparently, Caden applied the same insight he got from the string of three numbers to the string of four. Systematically employing the strategy of *informal induction*, Caden could see why the conjecture was not true for the case of 4 consecutive numbers, which in turn seeded an insight into why the conjecture was not true for even number of cases.

The interviewer prompted Caden again to consider when the conjecture would be true. Caden hypothesized that odd cases would work, and he developed similar justifications for the case of 7 and 9 consecutive numbers by extending his reasoning and employing the same strategy of *informal induction.* Therefore, Caden could develop deductive arguments that accounted for all possible sets of examples within each case, explaining why the conjecture worked or did not work for a given case. Thus, he could prove any given sub-conjecture by means of his strategy of informal induction, though he could not construct a single general proof for why the conjecture must work for all odd number strings of consecutive numbers.

In sum, like Ruhnema, Caden's proving process exemplifies the purposeful (*convey a general argument*) and strategic (*informal induction*) example use that we found as distinguishing characteristics of the successful provers. Caden could recognize a structural element in the very first example he used, and subsequently built all his examples based on his initial example by using the strategy of informal induction. It was critical that Caden systematically varied his initial example by adding 1 to each number to construct his next example, which enabled him to recognize that every set of string of 5 numbers could be created. In a way, his initial example has become a generic example for Caden, through which he could develop a general argument. Through systematic variation of examples, attending to structure, and reflection, Caden was able to develop general deductive arguments that explained why the conjecture was true for a given sub-conjecture. Thus, although the strategy of informal induction that Caden used is not the same as the formal method of proving by mathematical induction, it relies on similar ideas that the secondary students could devise in accordance with their mathematical capabilities. While Lange (2009) argues that proof by mathematical induction is never explanatory, our findings (considering the resemblance of *informal induction* to proving by mathematical induction) lend credence to Stylianides, Sandefur and Watson (2016) in their argument that proving by mathematical induction can be explanatory for students when certain conditions are satisfied.

4.2.1.3. Comparison of Rehnuma's and Caden's proving processes. Rehnuma's proving was characterized by her activity of *building formality* from examples. She was able to make generalizations in a couple of different ways. First, she could formalize her activity with two examples, and represented how she thought about those examples with algebraic notation. Second, Rehnuma could notice, after representing her activity algebraically and then looking back at one of her examples, that both examples had a structure to them – a middle number. She noticed a pattern, a regularity – that in both cases, the sum was equal to 5 times the middle. Because this abstraction and resulting generalization was a result of noticing a regularity, Rehnuma did not understand why the conjecture must be true. So, Rehnuma's justification was based on *result pattern generalization*. In contrast, Caden's proving was characterized by reliance on the strategy of *informal induction*. He consistently reflected on his activity with just one example from the very beginning, in the way that enabled him to form an image that he could manipulate in an anticipatory manner, constructing a general way of thinking that included an understanding of the logical necessity for why for any given string of numbers, if one example is a multiple of n, then any string of the same number of numbers must also be a multiple of n. Thus, Caden's justification was based on *process pattern generalization*.

Rehnuma produced a "proof that proves" rather than a "proof that explains" (Hanna, 2000). Her thinking was limited because she did not understand why the sum would always be n times the middle number, and it was difficult for her to generalize properly. Caden, on the other hand, produced a justification that explained why the conjecture must be true, and he understood the logical necessity for why one string of 3 consecutive numbers being a multiple of 3 implies that any string will be a multiple of 3. However, his thinking was also limited because he could only show that every string of n consecutive digits will work once one example for that number works. It did not enable him to see why *any* string of odd numbers would work, because he had not developed the insight about the middle term and relating the numbers to one another vis-a-vis that term.

Despite these limitations both students demonstrated strategic and purposeful use of examples. The students' purposes of examples included a wide variety, ranging from *placating the interviewer*, *understanding what the conjecture said*, and *testing the conjecture* in the beginning to *exploring the truth domain*, *illustrating a representation*, and *conveying a general argument* in the later phases of their proving activities. Both Rehnuma and Caden examined the examples by searching for a structural element across them. While Rehnuma attempted to build algebraic representations of the structure that she identified, Caden attempted to construct a general argument built on one base example. Each student's approach enabled him or her to gain many affordances, despite the limitations we outlined. But, what was also remarkable about these students, and other successful provers, was their curiosity to understand when the conjecture would be true in general and to understand why it would be true, as much as they could.

Collectively, these two successful cases of proving illustrate students' example use with the purposes of *conveying a general argument* and *illustrating a representation*; with the strategies of *structure*, *building formality*, and *informal induction*; and with the affordances of *gaining insight*, *generalizing*, and *justification support* – all of which were distinctive characteristics of the successful provers as discussed earlier. Also noteworthy is that all are closely related to attending to mathematical structure, generalizing, and generic examples – the three constructs that we highlighted in the literature on example use and proof.

4.2.2. Some impeding factors for gaining affordances from examples

A common characteristic we recognized among the secondary students who were unable to prove the consecutive integers task was a lack of curiosity about how to explore why the conjecture must be true, which was often coupled with a poor notion of proof. In some cases, the students' impoverished notion of proof manifested as a reliance on "authority" (Harel & Sowder, 1998). Such students' proving approach was often characterized by using examples mainly to test the veracity of the conjecture, with no apparent strategy either for *choosing* or *using* examples. For instance, Matt, a 9th-grade trigonometry student, used examples mainly to test the truth of the conjecture, with no apparent strategy either for choosing or using examples. Matt picked multiple strings of 5 consecutive numbers to *test the truth* of the conjecture and to *understand what* it said, but his example use was limited to verification. Matt was at a loss for how to show the conjecture might be true for any string of 5 consecutive numbers, stating that he would need "a mathematical law" to be fully convinced in the conjecture's truth. Matt explained, "I would probably do it multiple times and then maybe look on the internet and find if there is a rule that stated that . . . to see like – like a law, a mathematical law that stated that any five numbers added together would be divisible by five."

Matt's lack of interest in exploring when the conjecture might be true (and why it was true for some cases but not for others) was coupled with a disbelief in his ability to prove, which prevented him from gaining affordances from his examples and, thus, developing a proof. Hence, it seems important for mathematics teachers to cultivate a positive disposition in students to proving, supporting students' self-confidence by modeling how strategic and purposeful example use can lead to proofs that students too can develop.

Another factor that was a hindrance for some students, especially for the undergraduate students, was jumping to formality as the first recourse to proving the conjecture. These students were often eager to prove, but the strategy of *jumping to formality* hindered their proof attempts. As we have observed during the student interviews, recourse to algebra was simply not fruitful in some tasks, and skipping example exploration made the task more challenging. The task of Tyson's conjecture was an example of this phenomenon. For instance, Ivy, a second-year engineering student who had taken several undergraduate mathematics courses such as calculus, linear algebra, and multivariate differential equations, approached the task by attempting to algebraically prove the general conjecture, rather than by first exploring the conjecture for a specific case (such as the case of 5 consecutive numbers). Ivy first represented the conjecture as "n + n + 1 + n + 2 + n + 3 + . . . + n + (k-1), and then as '1 + 2 + 3 + . . .(k-1)' in an attempt to show that the sum was a multiple of k. This line of exploration did not prove fruitful, and so Ivy began to consider cases when k was 4, 5, and 6. These examples served to help Ivy build formality by adjusting her representation to (1 + 2 + 3 + . . . + k-3 + k-2 + k-1)/k. This could have been a key insight leading to a proof, but it is not trivial to simultaneously consider this representation for the general conjecture and for specific cases. Ivy remained confused about the conjecture's truth or how to prove it, and it was only after a prolonged attempt to prove with algebra that Ivy shifted to examples. Once she chose examples to *understand the conjecture*, specifically (10 + 11 + 12) and (10 + 11 + 12 + 13 + 14 + 15), Ivy could conclude that it was not true for the case of 6 consecutive numbers. However, at this point, Ivy did not further pursue the conjecture, commenting that she did not know how to prove it.

Although Ivy engaged in a very sophisticated reasoning and could express the conjecture in general terms, she struggled to prove the conjecture. Considering the sophisticated work that she produced and her mathematical experience, we are confident that Ivy could prove why the conjecture must be true if she had made use of examples to explore the conjecture for particular cases. Because she aimed to prove the conjecture for *any* number of string of consecutive numbers, Ivy set an extra challenge for herself, which prevented her from gaining greater affordances from examples. We consider two plausible explanations for Ivy's reluctance to use examples in her attempt to prove the conjecture. First, although the task was the same for all students, it was more formally stated in the undergraduate interview protocol. Thus, besides her advanced mathematical background, the more abstract formulation of the task might have encouraged Ivy to attempt to prove the conjecture algebraically, without first trying to explore the conjecture with examples. However, given that *jumping to formality* was a common strategy observed among the undergraduate students even when engaging with tasks that were identically stated for all students, this may also be due to her lack of awareness of how examples can be leveraged in successful proof development. In fact, Sandefur et al. (2013) identified that "experience in utility of examples in proving" was a common quality of the students in their study, who were able to effectively use examples in developing a proof. Thus, Ivy's case points to the importance of creating opportunities in order for students to appreciate the potential utility of examples in proving.

5. Discussion and implications

Producing a proof is not the only goal, or affordance, of example use, but we consider it a key instance of the many potential affordances that examples can provide. Although we distinguished the cases in terms of being able to produce a justification or not, we contend that all types of affordances are valuable as they are important building blocks leading to proof. Thus, we argue that in aiming to help students to learn to prove, these various affordances may function as intermediate goals for teachers to help students experience and appreciate them in their attempt at proving.

Accordingly, we conclude that examples were helpful for all students, but with varying degrees. For some students, examples were mainly helpful in revising conjectures, while for others they were also helpful in the development of generalizations and justifications. The students who could develop viable justifications often used an example to convey a general argument, whereas many students who could not develop a justification needed to use an example to understand what the conjecture was saying. This trend suggests that the successful provers usually could see a general feature in examples. In other words, they could "look through" examples, and then subsequently use examples to communicate the generality they had recognized (Watson & Mason, 2005), as we saw in Rehnuma's and Caden's proving processes. On the other hand, many of the unsuccessful students had difficulty in grasping what the conjecture meant to begin with. Dahlberg and Housman (1997) claim that understanding the concepts and definitions involved in a conjecture is critical in proof development, and that generating examples is a means of helping students to develop that understanding. Thus, many students turned to examples to make sense of the conjecture.

Furthermore, a key distinction between the successful and unsuccessful provers was the trends in the strategies they employed for choosing and using examples. While the successful provers' strategies were characterized by an almost even distribution between the strategies for *choosing* and *using* examples, the unsuccessful students' strategies were largely restricted to *choosing* examples, pointing to an area that needs to be supported. The successful provers also employed strategies more frequently, at almost double the frequency of the unsuccessful provers' strategies. Taken together, these findings suggest that the successful provers' frequent and varied use of strategies enabled them to gain greater affordances from examples, and subsequently supported their justification development. Underlying these striking differences, however, we contend that there appears to also be dispositional differences between the two groups. The fact that the successful provers approached the task with a broader set of strategies may also point to those students' curiosity about the conjectures and their eagerness to understand why the conjecture was true.

In contrast, the unsuccessful provers often spent less time on the task compared to the successful provers, as they were not as interested in investigating the conjecture beyond verifying it for specific cases. The unsuccessful students often had to be prompted by the interviewer to explore the truth domain for when the conjecture might be true. Moreover, the students' perception of themselves as incapable of producing proof also contributed to their lack of perseverance. It may be the case that viewing oneself as less capable, which in turn creates a need to rely on an "authority" for proof, and the lack of curiosity are reflexive, each breeding the other. In fact, it was the continued exploration and sustained thinking on when and why the conjecture was true that enabled the successful provers to develop a proof. Hence, this underscores an important instructional implication: How can we educators cultivate in students a view of themselves as capable of proving and in instilling in them a curiosity to explore mathematical relationships purposefully, strategically, and persistently? Students need to learn to use examples not only to convince themselves that a conjecture is true, but also to understand why it is true.

Relatedly, students' perceptions of confirming examples as sufficient proof was another hurdle to overcome. Many unsuccessful provers were not driven by purposeful and strategic example use; they merely tested examples to see if the conjecture held true. There was no indication of a desire to move beyond verifying the conjecture to understanding why it might be true. Hence, those students had limited gains from their example exploration. We therefore argue that mathematics teachers should continually emphasize the need to understand why a conjecture must be always true. Harel (2007) argues that not seeing an intellectual need to prove may be an underlying reason for students' difficulties with proving self-evident or obvious mathematical statements. Similarly, Buchbinder and Zaslavsky (2011) propose that (an appropriate level of) uncertainty may evoke a need for proof in individuals. These authors claim that students may be more likely to develop a proof if they see the uncertainty of the truth of a conjecture as resolvable, and they may give up on proving if the uncertainty is too strong.

Furthermore, Stylianides et al. (2016) maintain that "experience with utility of examples in proving" may facilitate students' productive use of examples when developing a proof. Hence, to support students' learning to prove, we first need to have students appreciate the value of examples in proving. Thus, students should be familiarized with "how constructing examples can be used not just to verify but also to expose structural relationships and generate conjectures" (Stylianides et al., 2016, p. 24). Moreover, Weber (2010b) discussed that "proofs that develop insight" not only explain why a mathematical statement is true but also extend one's understanding of the mathematical concepts and provide one with new methods that can be applied to new problems. Accordingly, we contend that classroom proving opportunities should be organized to realize these goals.

In addition, given that attending to mathematical structure in examples and thus generalizing the regularities observed in the process and relationships to new cases (PPG) – rather than generalizing the regularities in results (RPG) – were common activities among the successful provers, we follow Mason et al. (2009) in recommending that mathematics teachers continually encourage their students to try to discern structural elements in their examples, to systematically vary their examples and anticipate the effects of variation, and to reflect on the outcomes of the examples they used. Mason, Stephens and Watson also assert that teachers who appreciate mathematical structure, perceiving situations as instances of properties, are better equipped to support similar appreciation in their students. Thus, they encourage teachers to request their students to "justify their anticipations and actions on the basis of properties which have been discussed and articulated, rather than on the basis of inductive-empirical experience" (p. 29).

6. Conclusion

If one is able to see the general relationship that makes a conjecture true right away, then examples may not be necessary. But, this is rarely the case. In most cases the relationship is not obvious, thus mathematicians often draw on examples to investigate and gain insight into a conjecture or a novel problem. Given that students do not possess the mathematical tools, expertise, and metacognition skills that mathematicians do, identifying general mathematical relationships can be even more difficult. Thus, learning to use examples can be rich, largely untapped resource for students. Students can benefit vastly by learning to use examples in the service of understanding problems and justifying their solutions. We have shown that examples can be helpful to students across the grade levels in exploring, developing and revising conjectures, as well as in generalizing and justifying. But, students need to be supported in learning to use examples more purposefully and strategically.

In particular, we need to help students go beyond using examples only for conviction. As outlined in the CAPS framework (see) and shown in this paper, there are numerous purposes of example use and varied strategies that govern example choice and use, which, when applied attentively, could lead to many affordances, including successful development of proof. The identification of these rich possibilities of example use is an important contribution of the CAPS framework, while distinguishing the patterns of example use regarding the purposes, strategies, and affordances of example use among the successful provers is the contribution of this paper. We argue that students may greatly benefit from thinking with examples if they are introduced to the extensive array of purposeful and strategic example use. Teachers could model these productive purposes and strategies as they use examples in investigating, developing, and proving conjectures. Teachers could also encourage students to choose their examples with a goal in mind, asking them to anticipate the result of their example, and to continually reflect on their example use to inform their subsequent decisions.

We identified several distinguishing characteristics of the successful provers' example use in the context of a number theory task. These distinctions provide a helpful ground on which to build further examinations of the interplay among the categories of example use (criteria, affordances, purposes, and strategies) and the characteristics of example use associated with successful development of proof. However, once again, we underscore that students' example use is likely to depend on the context and domain of the task at hand and remind the reader that these identified characteristics emerged in the context of one number theory task. While our decision to focus on one common task enabled us to characterize the students' nature of example use free of potential variations that might have arisen due to the nature of different tasks, it also meant that the results are somewhat limited. Yet, this analytical choice also allowed us to conduct an in-depth examination of the proving processes of a fairly large number of students across a wide range of student populations, and thus the study provides a rich groundwork on which to build future research. Hence, we encourage further research to examine characteristics of students' example use leading to successful development of proof in the context of other domains and tasks as well.

Acknowledgements

This work was supported by the National Science Foundation (REC-1220623). The views expressed do not reflect official positions of the NSF. The authors wish to thank Oscar Bernal, Tracy Carolan, Gillian Galle, Elise Lockwood, Alison Gordon Lynch, Inbar Aricha Metzer, Pooneh Sabouri, Michael Thoms, and Orit Zaslavsky for their contributions to the project.

References

Alcock, L. J., & Inglis, M. (2008). Doctoral students' use of examples in evaluating and proving conjectures. *Educational Studies in Mathematics*, *69*, 111–129.
Antonini, S. (2003). Non-examples and proof by contradiction. In N. A. Pateman, B.J. Dougherty, & J. Zilliox (Eds.). *Proceedings of the 27th conference of the international group for the psychology of mathematics education*, *Vol. 2*, 49–56.
Balacheff, N. (1988). Aspects of proof in pupils' practice of school mathematics. In D. Pimm (Ed.), *Mathematics, teachers, and children* (pp. 216–235). London: Hodder & Stoughton.
Ball, D. L., Hoyles, C., Jhanke, H. N., & Movshovitz-Hadar, N. (2002). The teaching of proof. In L. I. Tatsien (Ed.), *Proceedings of the international congress of mathematicians* (Vol. III) (pp. 907–920). Beijing: Higher Education Press.
Bell, A. W. (1976). A study of pupil's proof explanations in mathematical situations. *Educational Studies in Mathematics*, *7*, 23–40.
Bills, L., & Rowland, T. (1999). Examples, generalization and proof. In L. Brown (Ed.), *Making meaning in mathematics. Advances in mathematics education* (pp. 103–116). York, UK: QED.
Buchbinder, O., & Zaslavsky, O. (2011). Is this a coincidence? The role of examples in fostering a need for proof. *ZDM –The International Journal on Mathematics Education*, *43*(2), 269–281.
Chazan, D. (1993). High school geometry students' justifications for their views of empirical evidence and mathematical proof. *Educational Studies in Mathematics*, *24*, 359–387.
Coe, R., & Ruthven, K. (1994). Proof practices and constructs of advanced mathematics students. *British Educational Research Journal*, *20*, 41–53.
Dahlberg, R. P., & Housman, D. L. (1997). Facilitating learning events through example generation. *Educational Studies in Mathematics*, *33*(3), 283–299.
Ellis, A. B., Lockwood, E., Williams, C., Dogan, M. F., & Knuth, E. (2012). Middle school students' example use in conjecture exploration and justification. In L. R. Van Zoest, J. J. Lo, & J. L. Kratky (Eds.), *Proceedings of the 34th annual meeting of the north American chapter of the international group for the psychology of mathematics education* (pp. 135–142). Kalamazoo, MI: Western Michigan University.
Ellis, A. B., Ozgur, Z., Vinsonhaler, R., Carolan, T., Dogan, M. F., Lockwood, E., Lynch, A. G., Sabouri, P., Knuth, E., & Zaslavksy, O. (2019). Student thinking with examples: The criteria-affordances-purposes-strategies framework. *Journal of Mathematical Behavior*, *53*, 263–283.
Goldenberg, P., & Mason, J. (2008). Spreading light on and with example spaces. *Educational Studies in Mathematics*, *69*(2), 183–194.
Hanna, G. (2000). Proof: Explanation and exploration: An overview. *Educational Studies in Mathematics*, *44*, 5–23.
Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. In A. Schoenfeld, J. Kaput, & E. Dubinsky (Eds.), *Research in collegiate mathematics education III* (pp. 234–283). Washington, DC: Mathematical Association of America.
Harel, G. (2001). The development of mathematical induction as a proof scheme: A model for DNR-based instruction. In S. Campbell, & R. Zazkis (Eds.), *Learning and teaching number theory: Research in cognition and instruction* (pp. 185–212). Norwood, NJ: Ablex Publishing Corporation.
Harel, G. (2007). The DNR system as a conceptual framework for curriculum development and instruction. In R. Lesh, J. Kaput, & E. Hamilton (Eds.), *Foundations for the future in mathematics education* (pp. 1–21). Erlbaum.
Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education*, *31*(4), 396–428.
Iannone, P., Inglis, M., Mejia-Ramos, J. P., Simpson, A., & Weber, K. (2011). Does generating examples aid proof production? *Educational Studies in Mathematics*, *77*, 1–14.
Knuth, E. J., Choppin, J., Slaughter, M., & Sutherland, J. (2002). Mapping the conceptual terrain of middle school students' competencies in justifying and proving. In D. Mewborn, P. Sztajin, D. White, H. Wiegel, R. Bryant, & K. Noony (Eds.), *Proceedings of the twenty-fourth annual meeting of the North American chapter of the International Group for the Psychology of mathematics education* (Vol. 4) (pp. 1693–1700). Columbus, OH: Educational Resources Information Center.
Knuth, E. J., Choppin, J., & Bieda, K. (2009). Middle school students' production of mathematical justifications. In D. A. Stylianou, M. L. Blanton, & E. J. Knuth (Eds.), *Teaching and learning proof across the grades: A K-16 perspective* (pp. 153–170). New York, NY: Routledge.
Knuth, E., Zaslavsky, O., & Ellis, A. B. (2019). The role and use of examples in learning to prove. *Journal of Mathematical Behavior*, *53*, 256–262.
Knuth, E. (2002). Secondary school mathematics teachers' conceptions of proof. *Journal for Research in Mathematics Education*, *33*(5), 379–405.
Kuchemann, D., & Hoyles, C. (2009). From empirical to structural reasoning in mathematics: Tracking changes over time. In D. Stylianou, M. Blanton, & E. Knuth (Eds.), *Teaching and learning proof across the grades: A K-16 perspective* (pp. 171–190). Mahwah, NJ: Taylor & Francis Group.
Lange, M. (2009). Why proofs by mathematical induction are generally not explanatory. *Analysis*, *69*, 203–211.
Lockwood, E., Ellis, A. B., & Lynch, A. G. (2016). Mathematicians' example-related activity when exploring and proving conjectures. *International Journal of Research in Undergraduate Mathematics Education*, *2*(2), 165–199.
Lynch, A. G., & Lockwood, E. (2019). A comparison between mathematicians' and students' use of examples for conjecturing and proving. *Journal of Mathematical Behavior*, *53*, 323–338.
Mason, J., & Pimm, D. (1984). Generic examples: Seeing the general in the particular. *Educational Studies in Mathematics*, *15*, 277–289.
Mason, J., Stephan, M., & Watson, A. (2009). Appreciating mathematical structure for all. *Mathematics Education Research Journal*, *21*(2), 10–21.
National Governors Association Center/Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Washington, DC: Council of Chief State School Officers.
Pedemonte, B., & Buchbinder, O. (2011). Examining the role of examples in proving processes through a cognitive lens: The case of triangular numbers. *ZDM: The International Journal on Mathematics Education*, *43*(2), 257–267. http://dx.doi.org/10.1007/s11858-011-0311-z

Piaget, J. (2001). Studies in reflecting abstraction. In R. Campbell (Ed.),. Sussex: Psychology Press.
RAND Mathematics Study Panel. (2002). *Mathematical proficiency for all students: toward a strategic research and development program in mathematics education (DRU-2773-OERI)*. Arlington, VA: RAND Education and Science and Technology Policy Institute.
Raman, M. (2003). Key ideas: What are they and how can they help us understand how people view proof? *Educational Studies in Mathematics*, *52*, 319–325.
Rowland, T. (2001). Generic proofs in number theory. In S. Campbell, & R. Zazkis (Eds.), *Learning and teaching number theory: Research in cognition and instruction* (pp. 157–184). Westport, CT: Ablex Publishing.
Sandefur, J., Mason, J., Stylianides, G. J., & Watson, A. (2013). Generating and using examples in the proving process. *Educational Studies in Mathematics*, *83*(3), 323–340.
Stylianides, G. J., & Stylianides, A. J. (2009). Facilitating the transition from empirical arguments to proof. *Journal for Research in Mathematics Education*, *40*(3), 314–352.
Stylianides, G. J., Sandefur, J., & Watson, A. (2016). Conditions for proving by mathematical induction to be explanatory. *Journal of Mathematical Behavior*, *43*, 20–34.
Watson, A., & Mason, J. (2005). *Mathematics as a constructivist activity: Learners generating examples*. Mahwah: Erlbaum.
Watson, A., & Shipman, S. (2008). Using learner generated examples to introduce new concepts. *Educational Studies in Mathematics*, *69*(2), 97–109.
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. (2010a). Mathematics majors' perceptions of conviction, validity, and proof. *Mathematical Thinking and Learning*, *12*, 306–336.
Weber, K. (2010b). Proofs that develop insight? *For the Learning of Mathematics*, *30*(1), 32–36.
Yackel, E., & Hanna, G. (2003). Reasoning and proof. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to the Principles and Standards for School Mathematics* (pp. 227–236). Reston, VA: NCTM.
Zaslavksy, O., Nickerson, S. D., Stylianides, A. J., Kidron, I., & Winicki-Landman, G. (2012). The need for proof and proving: Mathematical and pedagogical perspectives. In G. Hanna, & M. de Villiers (Eds.), *Proof and proving in mathematics education: the 19th ICMI study (New ICMI study series)* (Vol. 15) (pp. 215–229). New York, NY: Springer.