



## **International Journal of Contemporary Educational Research (IJCER)**

[www.ijcer.net](http://www.ijcer.net)

### **Pre-Service Teachers' Criteria for Evaluating Mathematical Arguments That Include Generic Examples**

**Muhammed Fatih Doğan<sup>1</sup>**  
<sup>1</sup>Adiyaman University

#### **To cite this article:**

Doğan, M.F. (2020). Pre-service teachers' criteria for evaluating mathematical arguments that include generic examples. *International Journal of Contemporary Educational Research*, 7(1), 267-279. DOI: <https://doi.org/10.33200/ijcer.721136>

This article may be used for research, teaching, and private study purposes.

Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

Authors alone are responsible for the contents of their articles. The journal owns the copyright of the articles.

The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of the research material.

## Pre-Service Teachers' Criteria for Evaluating Mathematical Arguments That Include Generic Examples

Muhammed Fatih Doğan<sup>1\*</sup>

<sup>1</sup>Adiyaman University

### Abstract

This study investigated how pre-service teachers evaluate mathematical arguments including generic examples. By using written responses of 71 PSTs, the results revealed six criteria used by PSTs, which were being explanatory, being general, correctness, mode of representation, mode of argumentation, and structure of the argument. The criteria suggest what PSTs considered and might value while evaluating arguments. Also, PSTs found deductive arguments more convincing than generic examples arguments. While evaluating arguments with generic examples nature, PSTs considered generic example with visual representation more valid and convincing than with numeric representation. PSTs seemed to be relatively adept at evaluating arguments; however, many had difficulty with identifying the structure of the generic examples. Overall, this study suggests a more coherent approach for integrating generic examples in teacher education programs and directions for further research.

**Key words:** Reasoning and Proof, Proof evaluation, Generic examples

### Introduction

Proof has received significant attention in mathematics at all grade levels and is an important part of every student's education. Proof is a central concept of mathematics and an important tool for teaching and learning mathematics (Knuth, 2002). Thus, both mathematics educators (e.g., Ball, Hoyles, Jahnke, & Movshovitz-Hadar, 2002; Hanna, 2018; Healy & Hoyles, 2000; Reid & Knipping, 2010) and reform documents (e.g., National Council of Teachers of Mathematics [NCTM], 2000; National Governors Association Center for Best Practices & Council of Chief State School Officers [NGA & CCSSO], 2010; Department for Education, 2014) advocate for placing reasoning and proof as a central part of mathematics education. For example, NCTM (2000) states that proof is an essential part of mathematical reasoning, stating that instruction focused on reasoning and proof from prekindergarten through grade 12 should enable all students to "recognize reasoning and proof as fundamental aspects of mathematics, make and investigate mathematical conjectures, develop and evaluate mathematical arguments and proofs, and select and use various types of reasoning and methods of proof" (NCTM, 2000, p. 56). However, the corpus of existing literature on learning and teaching proof indicates that students at all grade levels struggle to understand and construct proofs (e.g. Chazan, 1993; Harel & Sowder, 1998; Weber, 2010), and that teachers often have difficulty effectively fostering students' learning to justify and prove (Knuth, 2002).

To make reasoning and proof more accessible and meaningful to both students and teachers, the notion of generic example has received special attention in the literature because of its explanatory power (e.g., Balacheff, 1988; Harel & Sowder, 1998; A. Stylianides, 2007). Generic examples are particularly powerful in that they provide both explanation and conviction by assigning and illustrating a particular number and providing the foundation for a more general argument. Mason and Pimm (1984) define a generic example as a particular example that does not rely on any specific properties of that example. "A generic example is an actual example, but one presented in such a way as to bring out its intended role as the carrier of *general*" (Mason and Pimm, 1984, p. 284, italics added). Specifically, a generic example is a particular example that reveals the general structure of reasoning without relying upon the individual properties of that particular example. "The generic proof, although given in terms of a particular number, nowhere relies on any specific properties of that number" (Mason and Pimm, 1984, p. 284). Thus, the term *generic example* is used as making explicit the reasons for the truth of an assertion (Balacheff, 1988), not for another purpose.

\* Corresponding Author: Muhammed Fatih Dogan, [mfatihdogan@adiyaman.edu.tr](mailto:mfatihdogan@adiyaman.edu.tr)

To support their students, teachers need to be able to read and analyze students' arguments, which include generic examples, as their notion of what constitutes proof is crucial for teaching and learning reasoning and proof. In this paper, pre-service teachers (PST) were asked to analyze hypothetical student arguments that include generic examples. This paper examines pre-service teachers' notion of generic example in evaluating arguments and the criteria used to justify their responses. The research question is: *What types of criteria do pre-service teachers use to evaluate students' arguments?*

## Background and Theoretical Framework

In the following section, I first present the definition of proof and introduce generic examples. Then, I review the literature on proof evaluation.

**Definition of Proof.** Proof is “a means of convincing oneself whilst trying to convince others” (Alibert & Thomas, 1991, p. 215), and “an essential public activity” (Bell, 1976, p. 24) in which a person convinces themselves or others about the truth or falseness of mathematical propositions. Stylianides (2007) defines proof as follows:

Proof is a mathematical argument, a connected sequence of assertions for or against a mathematical claim, with the following characteristics:

1. It uses statements accepted by the classroom community (set of accepted statements) that are true and available without further justification;
  2. It employs forms of reasoning (modes of argumentation) that are valid and known to, or within the conceptual reach of, the classroom community; and
  3. It is communicated with forms of expression (modes of argument representation) that are appropriate and known to, or within the conceptual reach of, the classroom community.
- (p. 291)

This study embraces proof as a convincing argument and focuses on modes of argumentation and modes of representation.

**Generic Examples.** The notion of a generic example is seen as an important part of teaching and learning of proof. However, there is an ongoing debate on whether a generic example proves an assertion or a theorem. Balacheff (1988) categorizes *generic example* at a different and higher level than *naive empiricism*, and *the crucial experiment*, but sees as a lower level of *the thought experiment* which is counted as valid proof. Thus, he does not accept generic example as a full proof as it only depends on a particular representation and lacks for the (formal) linguistic expression that makes proof explicit. Harel and Sowder (1998) also see a generic example as an incomplete proof as it “reflects students' inability to express their justification in general terms” (p. 43). Similarly, Leron & Zaslavsky (2013) claim that “the main weakness of a generic proof is that it does not prove the theorem” (p. 27) as the facts for the proof are observed from an example. On the contrary, Rowland (2001, 2002) sees generic example as valid proofs and states it has the quality of a structural generalization to lead formal proof. Rowland (2002) notes:

The generic example serves not only to present a confirming instance of a proposition - which it certainly is - but to provide insight as to *why* the proposition holds for that single instance. The transparent presentation of the example is such that analogy with other instances is readily achieved, and their truth is thereby made manifest. Ultimately the audience can conceive of no possible instance in which the analogy could not be achieved. (p. 161, italics original)

Similarly, Yopp and Ely (2016) state that “a formal proof often uses general representations, such as quantified variables or symbolic placeholders. But, in a generic example argument, the generality lies not in the representation but in the way the example is appealed to” (p. 41). Therefore, generic examples can be seen as general proofs. I agree with Bills (1996), who states that “generic example might be a half-way house between empirical generalization and generalized formal proof” (p. 84). One of the strengths of generic examples is that they provide a bridge between empirical proofs and deductive proofs so that learners can leverage the increased generality of generic examples to progress closer to complete deductive proofs (G. Stylianides, 2008). Since a generic example can reduce the level of abstraction, it can also serve to make an argument or generic example

more accessible to students. Therefore, generic examples can play a crucial role in learning and teaching proof, especially at the K-12 level.

Generic examples have been used primarily to categorize different types of mathematical reasoning and arguments (Balacheff, 1988), and while researchers have recommended generic examples as a powerful tool for supporting proof (e.g., Rowland, 2002; Stylianides, 2007; Yopp & Ely, 2016), there is little agreement on what constitutes a generic example. I follow Yopp & Ely (2016) in their definition of a generic example, which they describe as reasoning around a particular example that nonetheless can be applied to all cases in the example's domain. Balacheff (1988) highlights the importance of the actions performed upon the generic example as determining whether or not the example is generic - for example, an action taken upon an example that relies upon a particular characteristic of the example that is not common to the entire domain identifies an example as non-generic. Reid & Vallejo Vargas (2018) highlight two important criteria to decide what counts as generic examples: Evidence of awareness of generality and Mathematical evidence of reasoning. Awareness of generality is required because students need to be aware that they are not only using examples as empirical evidence, but they are identifying a general structure through the use of their examples. Thus, they know that their argument is general enough to be counted as valid proof as it works for all cases in the domain. The mathematical reasoning reveals the reasoning behind their argument, in particular, why the identified structure may work for other cases from that specific case. This is grounded on the knowledge that specific community shares. Their categorization is based on psychological and social factors. They state that "psychologically, for a generic argument to be proof it must result in a general deductive reasoning process occurring in the mind of the reader that convinces the reader that there exists a fully deductive inference structure behind the argument. Socially, for a generic argument to be proof it must conform to the social conventions of the context" (p. 250).

I view GEs as less rigorous than a formal proof, but I agree with A. Stylianides (2007), Reid & Vallejo Vargas (2018) and Rowland (2002) that generic examples are valid mathematical arguments (Yopp & Ely, 2016). An important consequence of this view is that mathematics learners can still authentically engage in mathematical proving practices even without facility with formal proof representation. In short, here generic examples are considered a legitimate proving component.

**Proof Evaluation.** The relevant literature demonstrates that students and teachers have difficulties not only in constructing proofs (Martin & Harel 1989, Harel & Sowder 1998, Knuth, 2002) but also in determining what constitutes proof and whether a proof is valid (Dogan, 2015; Dogan and Williams-Pierce, 2019; Healy and Hoyles, 2000; Knuth, 2002; Selden & Selden 2003). This literature reveals that proof is not only difficult for students to learn but also for teachers to teach. Indeed, G. Stylianides, A. Stylianides, & Weber (2017) note that we still do not know enough about how to help all students develop a meaningful understanding of proof, or successfully produce and analyze mathematical proofs. Reid and Knipping (2010) stated, "how you teach proof depends on what you mean by 'proof' and what you think proofs are for" (p. 211). Thus, similar to "the mathematical knowledge needed to carry out the work of teaching mathematics" (Ball, Thames & Phelps, 2008, p. 395), teaching proof requires a certain kind of knowledge (Stylianides & Stylianides, 2008) to be able to teach what counts as acceptable mathematical proof and to evaluate students' arguments. In the same breath, Powers, Craviotto, and Grassl (2010, p. 501) state that the "ability to validate proofs is a much-needed skill for future teachers." Therefore, proof evaluation is a crucial part of proving activity because it allows one not only to decide the validity of an argument but also to clarify important mathematical principles (Lannin, Ellis, Elloitt, 2011). As stated by Lannin et al. (2011), the evaluation is to determine whether an argument includes "correct or mistaken assumptions, valid conclusions with erroneous logic, or valid arguments that nevertheless explain the only portion of the statement" (p. 45).

However, pre-service teachers and in-service teachers cannot often determine if a given argument is valid proof or not. Knuth (2002) found that in-service teachers had a hard time identifying what constituted valid proof. Teachers were more convinced by empirical arguments than deductive ones, and many teachers were not able to distinguish between proofs and non-proof arguments, as they frequently accepted invalid arguments as proof. More specifically, Knuth found that teachers mostly focused on surface features (such as correctness of algebraic manipulations), rather than deep features (such as nature of proof, overall logic, etc.) and found arguments convincing based on concrete features, specific examples, and visual representations. Martin & Harel (1989) investigated pre-service elementary school teachers' abilities to assess the validity of mathematical arguments by asking them to evaluate various arguments and to rate each argument. Like Knuth (2002), they found that many teachers (40%) accepted empirically-based arguments as proof. The authors also found that the form of argument affected whether or not teachers accepted a given proof. For example, teachers were likely to accept algebraic-symbolic proofs as valid without focusing on the validity of the actual argument. Morris

(2007) also found that pre-service teachers accept empirical arguments as valid proofs. These studies highlight teachers' limited notion of proof, but there are a few studies that found contrary results. For example, Bleiler, Thompson, and Krajčevski (2014) found that pre-service teachers were aware of the limitation of empirical arguments and did not accept them as valid proofs. Similar results were found by Ko & Knuth (2013), but the participants in their study accepted incorrect deductive arguments as valid proofs. Overall, the results of these studies show that both in-service and pre-service teachers' notion of proof may not be sufficient for teaching proof effectively.

Two similar studies, Ko and Hagen (2013) and Lovin, Cavey, and Whitenack (2004) investigated how in-service and pre-service teachers evaluated arguments including example-based, algebraic and generic example (visual) arguments. Ko and Hagen (2013) asked 55 in-service middle school teachers to evaluate three arguments. They found that 16 teachers were convinced by example-based reasoning, while 39 did not accept it as valid proof. Also, 37 teachers determined the algebraic argument to be a valid proof while only 27 teachers accepted the visual argument as valid proof. While evaluating arguments, Ko and Hagen found that the main criterion used by teachers was the use of generality. Lovin et al. (2004) worked with 280 pre-service teachers and asked them to evaluate three arguments. The algebraic argument was accepted by 83% of the pre-service teachers, the generic example argument was accepted by 63% of the students and the empirical arguments was accepted by 17% of the students as valid proof. Both studies emphasized that the participants recognized the power of algebraic and visual arguments, and the limitations of example-based arguments. Isler (2015) investigated elementary teachers' expectations of reasoning and proof in school mathematics and, as a part of her dissertation, she showed teachers mathematical arguments that includes generic examples. She found that the majority of the in-service elementary teachers identified generic examples as a general argument and valid proofs. However, the participants distinguished between generic examples that include visuals and just numeric examples, as they recognized visual generic examples more than the numeric ones; this means that the teachers did not recognize numeric generic examples as much as the visual generic examples. This suggests that teachers may prefer the generic argument not because of its generic nature but rather because of the inclusion of visuals. Dogan (2015) found similar results in his study with in-service secondary teachers. Teachers in his study put an extra emphasis on visual representations and accepted all visual representations as a generic example without focusing on the mathematical component of the argument.

## Method

### Participants and Study Context

This research was conducted in the mathematics education department at a public university in Turkey. The participants were pre-service teachers enrolled in an elective course called *Mathematical Reasoning and Proof* in 2017, 2018, and 2019, which means that the participants were different students each semester in the same class. The course was 2 hours each week for 14 weeks and designed to support PSTs' learning of how to teach reasoning, justification, and proof. The participants of this study were 71 students in their 3<sup>rd</sup> or last year of the course work in the program. Six of the students had a degree in mathematics, one had in engineering, and one had in nursing.

### Data Collection

Data were collected through written assessment items including 3 parallel tasks administered during one class period early in each semester. These items were adapted from literature (Lovin, Cavey, and Whitenack, 2004; Isler, 2015; and Dogan, 2015). Thus, the main data for this study comes from PSTs' written responses to the tasks for which they had to evaluate hypothetical student arguments. All tasks were related to number theory and designed to highlight generic examples arguments including visual representation and verbal representation along with a formal argument and an example-based argument. An example of tasks is presented below and Table 1 shows a summary of the characteristics of each hypothetical student argument.

The students responded to these three tasks at the beginning of the course. Here, the results of these tasks were presented and discussed. It is important to note that all three tasks had similar results, so I focus specifically on the first task as illustrative of the whole.

PSTs' were also asked to rank the students' arguments based on which one was the most and less convincing argument for them and to explain their ranking.

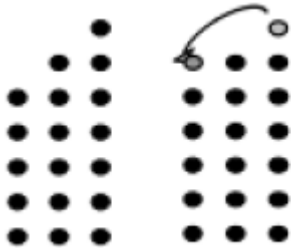
**Task 1- three consecutive number sum task**

Four students are discussing whether the following conjecture is always true.

The sum of any three consecutive numbers is equal to three times the middle number. For example, 4, 5 and 6 are consecutive numbers and  $4 + 5 + 6$  equals 15, which equals three times the middle number, 5. Show that the sum of any three consecutive numbers is always equal to three times the middle number. (Isler, 2015, p.79)

Their explanations are shown below. Do you think their explanations counts as valid proof? Explain your reasoning.

Table 1. Summary of four hypothetical student arguments of three consecutive number sum task that are given to PSTs

Argument	Argument Summary	
<p>Emir: I found a way using marbles. I can make three columns of marbles representing any three consecutive numbers. The first column represents the first number; the second column represents the middle number, and the third column represents the last number. I can take the top marble from the last column and move it to the first column. This makes the number of marbles in each column the same as the number of marbles in the middle column. Since the total number of marbles is always three times the number in the middle column, I know the conjecture is always true. (Isler, 2015, p.80)</p>	<p>Generic Example</p>	<p>Valid Proof</p>
		
<p>Damla: 5, 6, and 7 are three consecutive numbers and <math>5 + 6 + 7 = 18</math>, and <math>3 \times 6 = 18</math>. 7, 8, and 9 are three consecutive numbers and <math>7 + 8 + 9 = 24</math>, and <math>3 \times 8 = 24</math>. 569, 570, and 571 are three consecutive numbers and <math>569 + 570 + 571 = 1710</math>, and <math>3 \times 570 = 1710</math>. Since it works in these three examples, I know the conjecture is always true. (Isler, 2015, p.80; Lovin et al., 2004, p.1203)</p>	<p>Example-based reasoning</p>	<p>Invalid proof</p>
<p>Kenan: I'll show you using 4, 5 and 6. I can write 4 as <math>(5-1)</math> and 6 as <math>(5+1)</math>. So, it will be <math>(5-1) + 5 + (5+1)</math>. Since adding 1 and taking away 1 cancels each other, there will be three 5's. So, you see that it equals adding three times the middle number that is 5. (Isler, 2015, p.80)</p>		
<p>Doga: I can show three consecutive numbers as: <math>x-1</math>, <math>x</math>, <math>x+1</math>. So I can write the following equation to show that when you add 3 consecutive numbers, you always get 3 times the middle number:</p>	<p>Deductive Reasoning: Algebraic manipulation of the conjecture</p>	<p>Valid proof</p>
<p><math>(x-1) + x + (x+1) = 3x</math></p> <p>Since the 3 consecutive numbers add up to 3 times the middle number, I show that the conjecture is always true.</p>		

PSTs were just given the arguments not the categorization and summary of the proof types. These four hypothetical arguments used different reasoning and PSTs were expected to identify that reasoning. The first argument, Emir's argument, is a valid argument that used generic examples by using visual representations. In terms of Stylianides' proof definition, Emir uses a valid mode of argumentation to prove the conjecture by providing a general case. He also uses a valid visual representation that can be accepted at the middle school level. When one considers Reid & Vallejo Vargas's (2018) criteria for generic example, they can see awareness of generality as his argument states "*the total number of marbles is always three times the number in the middle column*" and mathematical evidence of reasoning as it represents a structure of the conjecture that shows that it works for any case. It is also important to note that Emir's argument does not rely on a specific case, but a general argument. The second argument, Damla's argument, is an obvious instance of example-based reasoning as she tries three examples and then generalizes her argument. This argument fails to have a valid mode of argumentation and mode of representation. The third one, Kenan's argument, is an incomplete generic example. He clearly shows the structure of the conjecture, but it is not clear if he provides a general argument. In Reid & Vallejo Vargas's (2018) terms, his argument has valid mathematical evidence of reasoning, but we cannot be sure about evidence of generality. We need more information to identify if Kenan sees *the general* through his example, or he just employs it for the case of 4, 5, and 6. Isler (2015, p.40) classifies his arguments as a complete generic example, but for this study, this counted as an incomplete generic example and PSTs were expected to identify the difference between Emir's argument and Kenan's argument. The final argument, Doga's argument, is a classic argument that counts as a deductive argument that uses algebraic representation. Thus, this argument had a valid mode of argumentation and mode of representation at the middle school level.

PSTs, then, were told, "*Rank each explanation through most convincing to least convincing, give 4 points for most convincing and 1 for least convincing*".

### Data Analysis

PSTs' evaluation of the hypothetical student arguments was first to analyze if they see the given argument as valid proof or not. After that, PSTs' written response for each argument (284 arguments of task 1) is analyzed by using open coding. Thus, I adapted Glaser and Strauss's (1967) constant comparison method to analyze students' responses on written tasks. All responses were systematically compared and contrasted within and across PSTs' responses. The goal was to identify regularities or patterns in PSTs responses. After having an initial coding framework on PSTs criteria for given arguments, emergent themes were identified and all data were coded based on the final coding framework. Six main criteria used by PSTs identified are presented in the results section.

To check the reliability of the coding scheme, two mathematics education graduate students read and coded all written responses. Only for six out of 284 arguments, there were disagreements. The disagreements were discussed and resolved by comparing the codes.

### Results

The purpose of this study was to investigate how PSTs evaluated arguments that include generic examples. Below, the frequency of which arguments PSTs considered as valid proof was presented. Secondly, the criteria used by PSTs while evaluating arguments were presented. Finally, how they ranked each argument was presented.

The results show that PSTs did not agree with which argument is a valid proof except for Doga's and Damla's arguments. Table 2 shows their decision of each argument.

Table 2. PSTs' responses of arguments validity

	Valid Proof	Invalid Proof
Emir's Argument	42 (59%)	29 (41%)
Damla's Argument	10 (14%)	61 (86%)
Kenan's Argument	31 (44%)	40 (56%)
Doga's Argument	62 (87%)	9 (13%)

As it can be seen from Table 2, PSTs considered Doga's argument as a valid proof, while they did not accept Damla's argument as a valid proof. Doga's argument was a deductive argument including algebraic representation and most of the PSTs considered it as proof. On the other hand, Damla's argument was an empirical argument and most of the PSTs did not consider it as proof. Thus, the results showed that PSTs were consistent with these two arguments. However, for Emir's and Kenan's arguments, there was a big difference between accepting as a valid argument or not. While around 60% of the PSTs accepted Emir's argument as a valid argument, 41% did not accept as proof. On the contrary, PSTs did not see Kenan's argument as equally valid as Emir's argument. 56% of the participants did not see Kenan's argument as valid, but 44% of them accepted it as proof. Considering similarities between Emir's and Kenan's arguments, these results reveal that PSTs did see the generic nature of those two arguments. As will be presented below, PSTs focused on the visual representation of Emir's argument but considered Kenan's argument as an empirical argument similar to Damla's argument. One important finding of this section was that the PSTs from the mathematics department (six participants) only considered Doga's argument as proof and others as invalid proofs mainly because of algebraic representation of his argument. One of those students did not accept any of these arguments as valid proof because none of them did use the axiomatic nature of mathematics.

The analysis revealed six main criteria that PSTs used while evaluating hypothetical student arguments. These criteria, as presented in table 3, are *Being Explanatory*, *Being General*, *Correctness*, *Mode of Representation*, *Mode of Argumentation*, and *Structure of the argument*. These criteria capture how PST's evaluate student's arguments. Below I discuss each criterion and illustrate with excerpts from PSTs' written responses.

Table 3. Criteria used while evaluating arguments

Criteria		Emir's Argument	Damla's Argument	Kenan's Argument	Doga's Argument	Total
<b>No Reason</b>		3	1	4	2	10
<b>Being explanatory</b>	Explanatory	12	0	4	2	18
	Not Explanatory	1	5	0	5	11
<b>Being general</b>	General	23	3	12	47	85
	Not general	22	38	24	0	84
<b>Structure of the argument</b>		22	0	25	2	49
<b>Correctness</b>		2	15	8	2	27
<b>Mode of Representations</b>	Visual model	41	0	0	0	41
	algebraic	0	0	0	57	57
	examples	1	2	7	0	10
<b>Mode of Argumentation</b>	Deductive arguments	0	2	1	13	16
	Example based reasoning	2	59	29	0	90
	Generic example	2	0	1	0	3

As can be seen from Table 3, 10 participants only stated if they think the argument is valid or not, and did not provide any reason for their decision. The most used criteria while evaluating arguments were *being general* or not (used 169 times) for all four arguments. All participants that used *generality* criteria considered Doga's



argument as general, most probably because of the mode of representation (algebraic). They did not see Damla's argument as general except three PST's who believed she provided a general argument. These two results were expected, but PSTs did not have a similar agreement for Emir's and Kenan's arguments. Most of the PTSs who used *generality* as a criterion for their evaluation did not see Kenan's argument as general. 24 of them stated that Kenan did not provide a general argument, while 12 of them stated Kenan's argument was general. On the other hand, PSTs used *being general* 23 times and not being general 22 times for Emir's argument. Half of the participants who used *generality* as a criterion saw Emir's argument as general, but the other half did not see as general. This was an interesting result about how PSTs perceived generic examples. The difference between the use of criteria of *generality* in Kenan's and Emir's arguments might be because of the mode of representation that each argument used. PSTs saw Kenan's argument as similar to Damla's argument rather than Emir's which includes visual representation. Similar criteria used by PSTs were the *structure of the argument*. This criterion was mostly used for the arguments that have a generic nature. They mentioned the *structure of the argument* 22 times for Emir's argument and 25 times for Kenan's argument. It is important to note that not seeing Kenan's argument general but identifying the structure of Kenan's argument highlights that PSTs did see a difference between Kenan's argument and Damla's argument which was example based. Some representative samples of PST's evaluating four hypothetical arguments are:

Damla's argument:

*This only generalizes based on 3 examples, so it cannot say anything for other numbers. It is only proof of those three special cases, not the conjecture.*

*I think it is a proof; it tests cases for the conjecture and then generalizes by comparing those cases. Also, there is not any counterexample, so it is proof.*

Kenan's argument

*The argument is proof. Kenan made logical reasoning. He did not make any computational errors. He did not use any formal proof methods, but his argument shows why the conjecture is true.*

*What he did is partially correct, but he just did it for the case of five, so it is not a proof. It should have something like this  $(n-1) + n + (n+1)$ .*

*This is not a proof. Because the student rewrites the conjecture in terms of the middle number. This is true for all three consecutive numbers, but it is only an example, it is not general. He should have used algebraic notations.*

Emir's argument:

*It is not a proof. Because you cannot find that many marbles to work with big numbers, so it is not generalizable.*

*It is not a proof. He only shows that the conjecture is true for a case and it does not have anything that shows why it is true. He also claimed that it is always true, he made a generalization based on his example, this shows his argument is incorrect.*

*I think this is a proof. Because when you give the extra one on the third column to the first column, all three columns would be equal, it does not matter which number you are using. In other words, the number of the first column and third column would be equal to the second column. The reason why I am considering this as a proof is that the number of marbles is not important. We know that the number of marbles on the third column is one more than the number of marbles on the second column and two more than the first column. When we add the extra marble on the third column to the first column they all will be equal and the sum of marbles would be three times the number of the marbles on the second column.*

*I think it is not a clear proof. When you look at his conclusion, you see that he generalized based on visual representation. But, I could not decide if it has a mathematical value or not. I think we can say he proved it since he provided a general argument.*

*We cannot say that Emir proved the conjecture. You have to make generalizations for proof, you need to write a formula. Here, Emir uses marbles as visuals, if he used "n" instead of marbles, he could prove it.*

Doga's argument:

*Doga starts with a general statement by using algebraic notations. He proves it, even we can say he has a mathematical proof. Because as a result, he has a general statement. To prove something you need to have this kind of results.*

PST's also mentioned the *correctness of the argument* as a criterion. Interestingly, they used this criterion mostly for Damla's (15 times) and Kenan's (8 times) arguments; because they wanted to emphasize what they are doing is mathematically correct, but not general to be counted as a valid proof. For example, they said:

*Damla gives three instances of the conjecture by using examples. She shows the correctness of the conjecture for these three cases. She does not answer if the conjecture holds for all numbers. Thus, this argument is not a valid proof. Proof needs to be general.*

*Kenan's argument is not a proof, it is an example that shows it's mathematically correct. He needs to show that all possible cases work. Since there are infinite numbers, this cannot be done.*

*Mode of Argumentation* and *Mode of Representation* were also used by PSTs while evaluating arguments. It is important to note that these two criteria are related. PSTs identified different representations for each argument. All participants who used this criterion stated that Emir's argument is a visual model that shows how the conjecture works, but one of them saw that as only an example of the conjecture. Similarly, they stated Doga's argument used an algebraic representation and was formal. Two PSTs saw *mode of representation* for Damla's argument as an example and accepted her argument as a valid proof. Also, seven of them stated that Kenan's argument was a valid argument that uses examples as a *mode of representation*. PSTs used example-based reasoning as a *mode of argumentation* when they did not see the arguments as valid proof. They stated that the *mode of argumentation* for Damla's (59 times) and Kenan's (29) arguments were examples and could not be seen as proof. Two of them also considered Emir's argument as example-based reasoning and did not consider as a valid proof. PSTs stated deductive *mode of argumentation* for Doga's (13 times), Damla's (2 times), and Kenan's (once) arguments. Very limited participants (3 of them) also mentioned using a generic example as *the mode of argumentation* for Emir's and Kenan's arguments. They said, for example, Emir used a visual model to show the conjecture works for all numbers.

Besides these criteria, PSTs also mentioned *being convincing* (3 times) and *counterexample* (12 times). *Being convincing* was mentioned by PSTs two times for Emir's argument and once for Kenan's argument. The *criterion of counterexample* was an interesting one as PSTs used it to explain both why some arguments cannot be considered as valid proof or can be considered as proof. They used the word counterexample for Damla's argument nine times to explain example based arguments are not valid proofs. For example:

*"You cannot prove this conjecture as Damla did. Because there is an infinite number of three consecutive number. If you could find a counterexample that would prove the conjecture, but just using numbers to show it works does not prove it."*

They also used the word counterexample for Doga's argument (3 times) as the generality of his argument. For example, one PST said: *"This argument is general, it works for all numbers. You cannot find any counterexample for it."* Thus, it can be said that they used the notion of counterexample to show the generality of the argument and limitation of example-based reasoning.

PST's were also asked to rank the arguments considering how convincing they think they are and to grade them between 4 to 1 as 4 being the most convincing and 1 being the least convincing. Table 4 shows the frequency and average point for each argument.

Table 4. PSTs rankings of the arguments

	Most Convincing	Least Convincing
Emir (2,62)	10	15
Damla (1,29)	0	48
Kenan (2,29)	9	6
Doga (3,67)	51	3

As it can be seen from Table 4, most of the participants found Doga's argument most convincing with the average of 3,67 out of 4. 51 participants ranked Doga as most convincing while none of them ranked Damla as most convincing. Damla got an average point of 1,29 out of 4, and 48 participants found her argument least convincing. Emir's and Kenan's arguments had close ranks while Emir's argument was 2,62 and Kenan's was 2,29. An interesting result was that 15 participants found Emir's argument least convincing while only 6 found Kenan's least convincing. This result was contradicted to participants' perception of each argument as 59% of them accepted Emir's argument as a valid proof but only 44% of them accepted Kenan's argument as a valid proof. Some examples of their justification are:

*Since Doga used a formal language to represent all consecutive numbers, it is most convincing for me. Emir's argument uses concrete visual representations, so I found it convincing. Both Kenan and Damla produce their arguments based on examples, so I did not find them convincing at all.*

*I think Emir's argument is the most convincing one because it has not only a visual representation but also a logical structure of the conjecture. The model sticks in your mind. Doga uses formal language but it does not explain why the conjecture works, so I did not find it convincing. Both Kenan and Damla use examples which is the simplest method to prove. Their arguments are example-based, so I don't see them as valid proof. They are the least convincing.*

*Doga's argument is the most convincing. Because it uses algebraic general representation, it shows which number you put it will always work. But, Emir's argument is well-developed too and it is a valid one. Kenan did a similar argument as Emir, but it would be better if Kenan used a visual model, which would make his more convincing. I do not see Damla's argument correct because it does not prove.*

*Doga's argument is the most convincing. Emir and Kenan used different representations, but they both made the same argument. Damla's argument is incorrect. Doga's argument is mathematically true and easy to understand, so that argument is the most convincing.*

## Discussion

Proof evaluation is an important ability for teachers to have effective teaching of reasoning and proof. Specifically, teachers need to understand and identify types of arguments generated by their students in order to help them develop an understanding of proof that is advocated by reform initiatives (e.g., NGA & CCSSO, 2010; Hanna, 2018; Powers et al., 2010; Suominen, Conner, & Park, 2018). In this study, PSTs' notion of generic example was examined by assessing the proving criteria they used while evaluating hypothetical student arguments. The criteria used by PST's were: Being Explanatory, Being General, Correctness, Mode of Representation, Mode of Argumentation, and Structure of the Argument. The PST's mostly looked for generality when evaluating arguments. They also identified the structure of the arguments mostly for Emir's and Kenan's argument which had a generic nature.

The results reveal some promising findings but also some problematic ones. PSTs successfully evaluated arguments that were either deductive or example-based reasoning. Most of the participants considered Doga's argument as a valid proof while they did not consider Damla's argument as a valid proof. Thus, this result showed that the PST's were aware of the limitations of examples-based reasoning. This finding supports Ko and Knuth (2013) and Bleiler et al., (2014) results which also showed that PST's were aware of the limitation of empirical arguments. Also, this result aligns with Isler (2015) who found most of the in-service teachers recognized invalid empirical arguments, Ko and Hagen (2013) who found most of the in-service teachers accepted algebraic arguments as valid and example-based ones as invalid, and Lovin et al., (2004) who found most of the PST's considered example-based arguments as invalid, but algebraic arguments as proof. Some of the literature showed the pre-service teachers and in-service teachers rely on example-based reasoning and accepted empirical arguments as proof (Knuth, 2002; Martin & Harel, 1989) especially while generating arguments. One of the results of this study suggests that the participants do not use example-based arguments in the same way when evaluating arguments. Overall, this result was promising as it highlights that PSTs were aware of invalid arguments.

One important note to have here is that the PSTs who have a degree in mathematics only accepted deductive arguments as valid proof, while they did not consider other arguments as valid. This might be because of their perception of proof in mathematics courses where the axiomatic nature of proof is emphasized (Weber, 2010).

Another important result of this study was the PSTs' notion of generic example arguments. Even though PSTs did not have any former instruction about generic example, 59% of them recognize generic example argument as a valid proof. Both Ko and Hagen (2013) and Lovin et al., (2004) found similar results with a close percentage of the participants who accepted generic examples as valid proof. One important difference of this study from Ko and Hagen (2013) and Lovin et al., (2004) studies was the number of arguments that include generic examples. In these studies, only one generic example with a visual argument was presented to the participants for evaluating. In this study, I used two different arguments; a complete generic example with visual representation and an incomplete generic example that uses an example as representation. Isler (2015) used two generic examples that include visual and numeric examples, and this study used those arguments. However, as discussed in the method section, for the purpose of this study Emir's argument was accepted as a valid proof, while Kenan's argument was considered as an incomplete generic example as it does not show evidence for generality (Reid & Vallejo Vargas, 2018). The findings related to Emir's argument seems to be consistent with the results of Isler (2015). As the participants of her study, more than half of the PSTs identified generic examples as proof.

The findings of this study show the participants did not see Kenan's argument as they accepted Emir's argument. One of the reason for this finding might be because of having a visual representation for Emir's argument while having a numeric representation for Kenan's argument. Indeed, Isler (2015), Ko and Hagen (2013) and Dogan (2015; 2019) found that teachers tend to accept visual examples as general, while hesitant to consider generic examples with numeric representation, not valid arguments. As discussed by Isler (2015), this finding might be problematic as teachers may accept visual arguments as general but do not accept generic examples with numeric representations. In-service teachers who participated in Dogan's (2015; 2019) study indeed accepted all visual representations as generic examples without any further consideration, while they had a hard time to distinguish between empirical arguments and generic examples with numbers when evaluating arguments. Another reason might be the opaque nature of generic examples (Rø and Arnesen, 2020). The participants of Rø and Arnesen's study were asked to use generic examples to construct proofs. They found that none of the PSTs provided complete generic arguments as the participant might not be aware of the structure of a generic argument. They suggested that teacher educators should provide a more coherent approach when teaching proof with a generic example by emphasizing the structure of the arguments. My findings show that almost one-third of the PSTs were aware of the structure of generic examples for Emir's and Kenan's arguments. They considered Emir's argument more convincing than Kenan's argument but still did not find them as convincing or valid arguments as Doga's argument. Not seeing generic examples as deductive arguments might be problematic especially for teaching proof at the middle school level as explanatory power of the generic example (Hanna, 2018). Indeed, Harel and Sowder (1998) consider generic examples under the deductive proof scheme which suggests teachers need to embrace generic examples as valid proof.

## Conclusion

Teachers' view of what constitutes as proof is a crucial role in teaching and learning proof. Teachers should be able to construct, read, and evaluate proofs in order to help their students. This study investigated how PSTs evaluate hypothetical student arguments. The results mainly suggest that they were adapted to evaluate arguments, but had a lot of room to increase their conceptions of proof. One important conclusion of this study is that PSTs need to have more opportunities in engaging in proving activities that include generic examples. They found deductive arguments including algebraic representations more convincing and valid than generic examples that include visual and numeric representations. Thus the emphasis of generic examples needs to be put in teacher education programs to have a more adequate notion of proof for teaching and learning. In addition, the results of this study were limited to written responses, which allowed to identify criteria for evaluating arguments but cannot say more about why they use those criteria for evaluating. Studies that include both interviews and classroom observations are needed to better understand PSTs conceptions of what constitutes for valid arguments and generic example.

## References

- Alibert, D. & Thomas, M. (1991). Research on mathematical proof. In D. Tall (Ed.) *Advanced Mathematical Thinking* (pp. 215-230). Kluwer: The Netherlands.
- 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., Hoyles, C., Jahnke, H., & Movshovitz-Hadar, N. (2002). The teaching of proof. Paper presented at the International Congress of Mathematicians, Beijing, China. In L. Tatsien (Ed.), *Proceedings of the international congress of mathematicians* (Vol. 3, pp. 907–920). Beijing: Higher Education Press.
- Ball, D.L., Thames, M.H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389-407.
- Bell, A. (1976). A study of pupils' proof – explanations in mathematical situations. *Educational Studies in Mathematics*, 7, 23-40.
- Bieda, K. (2010). Enacting proof-related tasks in middle school mathematics: challenges and opportunities. *Journal for Research in Mathematics Education*, 41(4), 351-382.
- Bills, L. (1996). The Use of examples in the teaching and learning of mathematics. In Puig L. & Gutierrez A. (Eds.), *Proceedings of the 20<sup>th</sup> Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 81–88). Valencia, Spain: PME.
- Bleiler, S. K., Thompson, D. R., & Krajčevski, M. (2014). Providing written feedback on students' mathematical arguments: Proof validations of prospective secondary mathematics teachers. *Journal of Mathematics Teacher Education*, 17(2), 105-127.
- Department for Education. (2014). *Mathematics programmes of study: key stages 1 and 2: National curriculum in England*. Retrieved from <https://www.gov.uk/government/publications/national-curriculum-in-englandmathematics-programmes-of-study>
- Dogan, M. F. (2015). The nature of middle school in-service teachers' engagements in proving-related activities. [Unpublished doctoral dissertation]. University of Wisconsin-Madison, USA.
- Dogan, M. F. (2019). The nature of middle school in-service teachers' engagements in proving-related activities. *Cukurova University Faculty of Education Journal*, 48(1), 100-130.
- Dogan, M. F., & Williams-Pierce, C. (2019). Supporting teacher proving practices with three phases of proof. *Teacher Education Advancement Network Journal*, 11(3), 48-59.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: strategies for qualitative theory*. New Brunswick: Aldine Transaction.
- Hanna, G. (2018). Reflections on proof as explanation. In *Advances in Mathematics Education Research on Proof and Proving* (pp. 3-18). Cham, Springer.
- Harel, G. & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. In A. H. Schoenfeld, J. Kaput, & E. Dubinsky (Eds.), *Issues in mathematics education: Vol. 7. Research in collegiate mathematics education III* (pp. 234-283). Providence, RI: American Mathematical Society.
- Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education*, 31(4), 396–428.
- Isler, I. (2015). *An Investigation of Elementary Teachers' Proving Eyes and Ears*. Unpublished Doctoral dissertation. The University of Wisconsin-Madison, USA.
- Knuth, E. (2002). Teachers conceptions of proof in the context of secondary school mathematics. *Journal for Research in Mathematics Education*, 5(1), 61-88.
- Ko, Y. Y., & Hagen, C. J. (2013). Conviction and validity: Middle school mathematics teachers' proof evaluations. In M. Martinez, & A. Castro Superfine (Eds.). *Proceedings of the Thirty-Fifth Annual Conference of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 801-804). Chicago, IL.
- Ko, Y. Y., & Knuth, E. J. (2013). Validating proofs and counterexamples across content domains: Practices of importance for mathematics majors. *The Journal of Mathematical Behavior*, 32(1), 20-35.
- Lannin, J. K., Ellis, A. B., & Elliott, R. (2011). *Developing essential understanding of mathematical reasoning for teaching mathematics in prekindergarten-grade 8*. Reston, VA: National Council of Teachers of Mathematics.
- Leron, U., & Zaslavsky, O. (2013). Generic proving: Reflections on scope and method. *For the Learning of Mathematics*, 33(3), 24–30.
- Lovin, L. A., Cavey, L. O., Whitenack, J. W. (2004) Evidence and justification: Prospective K–8 teachers' proof-making and proof-evaluating. In D. E. McDougall and J. A. Ross (Eds.). *the Proceedings of the Twenty-Sixth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 1201–1208). Toronto, Canada.
- Martin, W.G., & Harel, G. (1989). Proof frames of preservice elementary teachers. *Journal of Research in Mathematics Education*, 20(1), 41-51.
- Mason, J., & Pimm, D. (1984). Generic examples: Seeing the general in the particular. *Educational Studies in Mathematics*, 15, 227-289.
- Morris, A. K. (2007). Factors affecting pre-service teachers' evaluations of the validity of students' mathematical arguments in classroom contexts. *Cognition and Instruction*, 25(4), 479-522.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

- National Governors Association (2010). *Common Core State Standards for Mathematics*. Washington, DC: Council of Chief State School Officers.
- Powers, R. A., Craviotto, C., & Grassl, R. M. (2010). Impact of proof validation on proof writing in abstract algebra. *International Journal of Mathematical Education in Science and Technology*, 41(4), 501-514.
- Reid, D., & Knipping, C. (2010). *Proof in mathematics education: Research, learning, and teaching*. Rotterdam: Sense Publishers.
- Reid, D., & Vallejo Vargas, E. (2018). When is a generic argument a proof? In A. J. Stylianides, & G. Harel (Eds.). *Advances in mathematics education research on proof and proving* (pp. 239–251). Cham: Springer International Publishing.
- Rø, K., & Arnesen, K. K. (2020). The opaque nature of generic examples: The structure of student teachers' arguments in multiplicative reasoning. *The Journal of Mathematical Behavior*, 100755.
- Rowland, T. (2001). Generic proofs: Setting a good example. *Mathematics Teaching*, 177, 40-43.
- Rowland, T. (2002). Generic proofs in number theory. In S. R. Campbell & R. Zazkis (Eds.), *Learning and teaching number theory: Research in cognition and instruction* (pp. 157–183). Westport, CT: Ablex Publishing.
- Selden, A., & Selden, J. (2003). Validations of proofs considered as texts: Can undergraduates tell whether an argument proves a theorem? *Journal for Research in Mathematics Education*, 34(1), 4-36.
- Stylianides, A.J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education*, 38, 289-321.
- Stylianides, G. J. (2008). An analytic framework of reasoning-and-proving. *For the Learning of Mathematics*, 28(1), 9-16.
- Stylianides, G. J., & Stylianides, A. J. (2008). Proof in school mathematics: Insights from psychological research into students' ability for deductive reasoning. *Mathematical Thinking and Learning*, 10(2), 103-133.
- Stylianides, G. J., Stylianides, A. J., & Weber, K. (2017). Research on the teaching and learning of proof: Taking stock and moving forward. In J. Cai (Ed.), *Compendium for research in mathematics education*. (pp. 237–266). Reston, VA: National Council of Teachers of Mathematics.
- Suominen, A. L., Conner, A., & Park, H. (2018). Prospective mathematics teachers' expectations for middle grades students' arguments. *School Science and Mathematics*, 118(6), 218-231.
- Yopp, D. A., & Ely, R. (2016). When does an argument use a generic example? *Educational Studies in Mathematics*, 91(1), 37-53.
- Weber, K. (2010). Mathematics majors' perceptions of conviction, validity, and proof. *Mathematical Thinking and Learning*, 12(4), 306-336.