INVESTIGATING A MATHEMATICS TEACHER AND HIS EIGHTH-GRADE STUDENTS' PROOF EVALUATIONS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

HALİL İBRAHİM ŞİMŞEK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
MATHEMATICS EDUCATION IN MATHEMATICS AND SCIENCE
EDUCATION

SEPTEMBER 2023

Approval of the thesis:

INVESTIGATING A MATHEMATICS TEACHER AND HIS EIGHTH-GRADE STUDENTS' PROOF EVALUATIONS

submitted by **HALİL İBRAHİM ŞİMŞEK** in partial fulfillment of the requirements for the degree of **Master of Science** i**n Mathematics Education in Mathematics and Science Education, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** ________________

Prof. Dr. Mine Işıksal Bostan
Head of the Department, **Mathematics and Science Education** ________________

Assist. Prof. Dr. Işıl İşler Baykal
Supervisor, **Mathematics and Science Education, METU** ________________

Examining Committee Members:

Assist. Prof. Dr. Zekiye Özgür
Mathematics and Science Education, DEU ________________

Assist. Prof. Dr. Işıl İşler Baykal
Mathematics and Science Education, METU ________________

Prof. Dr. Safure Bulut
Mathematics and Science Education, METU ________________

Date: 07.09.2023

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name Last name : Halil İbrahim Şimşek

Signature :

ABSTRACT

INVESTIGATING A MATHEMATICS TEACHER AND HIS EIGHTH-GRADE STUDENTS' PROOF EVALUATIONS

Şimşek, Halil İbrahim
Master of Science, Mathematics Education in Mathematics and Science Education
Supervisor : Assist. Prof. Dr. Işıl İşler Baykal

September 2023, 126 pages

Proof and reasoning can be considered as an essential aspect of mathematics education that should be developed by students from an early age. However, it should be noted that an argument without appropriate reasoning cannot be considered valid proof. Past studies show that students and teachers have difficulty distinguishing between valid and invalid proofs (e.g., Stylianides et al., 2017). In order to understand this problem more deeply, a case study was conducted with a mathematics teacher and his 20 eighth-grade students, aiming to reveal the similarities and differences in the proof evaluation of the teachers and the students. The data were gathered through written questionnaires and interviews. The results showed that while the axiomatic and inductive (example-based) arguments were convincing for both the teachers and the students, the authoritarian arguments were not. In addition, while the transformational and perceptual arguments were convincing for the teacher, the students did not find them convincing. In terms of showing that the mathematical statement always works, while the students considered all inductive arguments valid, more than half of the students did not consider axiomatic arguments valid. On the other hand, the teacher considered all axiomatic arguments valid, but only one inductive argument was valid for him. Also, it was found that the teacher mostly focused on examples, validity for all cases, and using algebra to evaluate the convincingness and validity of the arguments and to score arguments like they were his students’ responses. The students mostly focused on examples and understandability of the argument to evaluate convincingness, validity, and teacher expectancy.

Keywords: Proof Evaluation, Teacher Proof Evaluation, Student Proof Evaluation

ÖZ

BİR MATEMATİK ÖĞRETMENİ VE ONUN SEKİZİNCİ SINIF ÖĞRENCİLERİNİN İSPAT DEĞERLENDİRMELERİNİN İNCELENMESİ

Şimşek, Halil İbrahim
Yüksek Lisans, Matematik Eğitimi, Matematik ve Fen Bilimleri Eğitimi
Tez Yöneticisi: Dr. Öğr. Üyesi Işıl İşler Baykal

Eylül 2023, 126 sayfa

İspat ve akıl yürütme, matematik eğitiminin öğrenciler tarafından erken yaşlardan itibaren geliştirilmesi gereken önemli bir yönü olarak düşünülebilir. Ancak şunu da belirtmek gerekir ki, uygun akıl yürütmeye dayandırılmayan bir argüman geçerli bir ispat olarak değerlendirilemez. Geçmiş çalışmalar, öğrenci ve öğretmenlerin geçerli ve geçersiz ispatları ayırt etmekte zorlandıklarını göstermektedir (Örn., Stylianides vd., 2017). Bu sorunun daha derinlemesine anlaşılabilmesi için bir matematik öğretmeni ve onun 20 sekizinci sınıf öğrencisi ile, öğretmen ve öğrencilerin ispat değerlendirmelerindeki benzerlik ve farklılıkları ortaya çıkarmayı amaçlayan bir durum çalışması yapılmıştır. Veriler yazılı anketler ve görüşmeler yoluyla toplanmıştır. Sonuçlar, aksiyomatik ve tümevarımsal (örnek temelli) argümanların hem öğretmenler hem de öğrenciler için ikna edici olduğunu ancak otoriter argümanların ikna edici olmadığını göstermiştir. Ayrıca dönüşümsel ve algısal argümanlar öğretmen açısından ikna edici bulunurken öğrenciler tarafından ikna edici bulunmamıştır. Matematiksel ifadenin her zaman işe yaradığını gösterme açısından öğrenciler tümevarımsal argümanların tamamını geçerli görürken yarıdan fazlası aksiyomatik argümanları geçerli bulmamıştır. Öte yandan öğretmen tüm aksiyomatik argümanları geçerli bulduğunu ancak kendisi için yalnızca bir tümevarımsal argümanın geçerli olduğunu belirtmiştir. Ayrıca öğretmenin, argümanların ikna ediciliğini ve geçerliliğini değerlendirmek ve argümanları kendi öğrencilerinin cevaplarıymış gibi puanlamak için çoğunlukla örneklere, tüm durumlar için geçerliliğe ve cebir kullanımına odaklandığı gözlemlenmiştir. Öğrenciler ise ikna edicilik, geçerlilik ve öğretmen beklentisine göre değerlendirme yaparken çoğunlukla örneklere ve argümanın anlaşılırlığına odaklanmışlardır.

Anahtar Kelimeler: İspat Değerlendirmesi, Öğretmen İspat Değerlendirmesi,
Öğrenci İspat Değerlendirmesi

To my love and my family

ACKNOWLEDGMENTS

Firstly, I want to thank my respectful advisor, Assist. Prof. Dr. Işıl İŞLER BAYKAL. It is lucky to have such an advisor who is a devoted, altruistic, supportive, and wise expert mentor. Her guidance taught me to be a researcher, and her feedback and advice made the study more quality throughout the journey. She encouraged me to get up when I stumbled. When I lost, she lightened my way with her wisdom and positive energy. I think there is no way to express my deepest gratitude fully, and I can never thank her enough.

I also want to present sincere thanks to dear committee members Assist. Prof. Dr. Zekiye ÖZGÜR and Prof. Dr. Safure BULUT for their constructive and valuable comments, suggestions, and contributions, which helped to enhance the study.

A special thanks goes to Assoc. Prof. Dr. Muhammed Fatih DOĞAN and Asst. Prof. Dr. Emine Gaye ÇONTAY for their time, feedback, and valuable advice on the tools. Similarly, I would like to thank the teachers and all the students who participated in the study for their voluntariness, time, and contributions.

Throughout the journey, my love Saime Nur GEDİK has stood by me and believed in me to achieve great work. I want to give lovely thanks to her because she was always there during my mental breakdowns to support and motivate me even when I lost faith in myself. She has been turning the storm clouds into April showers in my life. I am so fortunate to have her.

Finally, I especially thank my parents, Ramazan and Seher ŞİMŞEK, and my brothers Oğuzhan and Ahmet ŞİMŞEK for their unconditional love and continuous support during my whole life. It is glad to have such a family who raised me to have the courage to go further and give their emotional support to make this path easier.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

NCTM	National Council of Teachers of Mathematics
MoNE	Ministry of National Education

CHAPTER 1

INTRODUCTION

Mathematics is not just terms and operations but also the logic and reasoning behind them. Moreover, reasoning is called mathematical argumentation and justification, and the proof is often the end of the argumentation (Yackel & Hanna, 2003). Reasoning and proof are accepted as one of the process standards by the National Council of Teachers of Mathematics (NCTM, 2000), and many researchers consider the proof as a central point for mathematics and an important aspect of mathematics education from early grades (e.g., Ball et al., 2002; Knuth, 2002a). Although the Turkish mathematics curriculum does not include specific terms of "proof" at the middle school level, it aims, "*In the problem-solving process, they (the students) will be able to easily express their thoughts and reasoning and will be able to see the deficiencies or gaps in the mathematical reasoning of others*." (Ministry of National Education [MoNE], 2018, p. 9). So, although the Turkish mathematics curriculum does not mention the term "proof," it does "mathematical reasoning." Such explanation and communication are the most highlighted functions of proof for mathematics education (Yackel & Hanna, 2003). Also, Ellis et al. (2012) suggested proof be embedded and a regular part of mathematics courses instead of being a separate subject.

Although proof is considered an essential aspect of early grades, proof might not be in the same form in each grade. Kline (1970) mentioned the importance of deductive structure in a proof, but he emphasized intuition before the deductive structure and suggested that rigor proof should enter gradually based on the student's mathematical development levels. Such gradual progress can be ensured by examples. Balacheff (1988, p. 219) mentioned that generic examples involve explicit reasoning and carry “a characteristic representative of its class” with explicit reasoning, and these examples may ensure a transition from showings to formulations. , Ellis et al. (2019) found that noticing similarities in the structures of different examples was beneficial and ensured ways for deductive reasoning to students from middle school through the undergraduate level. Being aware of such examples could help the teachers direct the students to deductive reasoning. It is also essential for teachers to notice deductive structures in students’ arguments because, as Kline (1970, p. 272) said, the students could recreate mathematics “with the aid of a teacher.”

Not all arguments have to be qualified as proof. While an argumentation is “*a reasoned discourse that is not necessarily deductive*,” mathematical proof is “*a chain of well-organized deductive inferences that uses arguments of necessity*,” and helping students to have such deductive reasoning is one essential role for educators (Hanna & de Villiers, 2008, p. 331). Thus, both teachers and students need to recognize the value of proof and proving and be able to evaluate arguments to distinguish between deductive and non-deductive ones. Such distinction is critical in terms of improving mathematical reasoning (Morris, 2007). However, based on the key findings from research about proof, teachers and students have difficulties distinguishing valid and invalid arguments; moreover, they can be convinced by empirical arguments (Stylianides et al., 2017). As a result, proof conceptions of both students and teachers need to be improved and investigated in terms of how they approach and evaluate the arguments.

Some studies (e.g., Bieda & Lepak, 2014; Heinze & Reiss, 2003) indicated that the empirical arguments convinced students. Even if they were aware of the limitations of examples, they used empirical arguments in their proof construction (Healy & Hoyles, 2000). The students might not find the algebraic deduction reliable or notice the generalizability power of algebra (Liua et al., 2016). Thus, the students’ proof conceptions should be investigated, and the ways to enhance the proof conception of the students should be found.

Although the importance of proof in early grades was emphasized (e.g., Ball et al., 2002; Knuth, 2002a), the teachers thought the proof might not be appropriate for all grades (Knuth, 2002b). Morris (2007) also found that preservice teachers tended to accept examples as proof if they did not see deductively valid arguments in a discussion. Teachers' beliefs about proof in schools and proof conception might affect the students' proof conceptions. As Lin et al. (2012) also emphasized, teachers could implement activities and lead the class about which kind of argument can be qualified proof. Ozgur (2017) stressed that the proof conception of the teacher shapes these instructional activities, and she found similarities between the proof conceptions between the teacher and the students. Therefore, like the proof conception of the students, the proof conception of the teachers should be investigated; in that way, they can help the students better understand proofs.

1.1 The Purpose of the Study

The study aimed to investigate the proof evaluation of a middle school mathematics teacher and his students and compare their proof evaluations to determine which way they are similar or different. In that way, gaining more insight into the teacher's and students' proof conceptions was aimed. On the other hand, the proof evaluation can be approached in different aspects. The purpose of this study was to specifically investigate the proof evaluations of the students and the teacher in terms of the convincingness and validity of the arguments and their expectations of each other.

1.2 Research Questions

The following questions with the sub-questions were the research questions that guided this study:

What are the similarities and differences between the proof evaluation of a mathematics teacher and his 8th-grade students?

A. What criteria do the students use to evaluate an argument in terms of convincingness?

B. What criteria do the students use to evaluate an argument in terms of validity?

C. What criteria do the students expect their teacher to use for scoring an argument?

D. What criteria does the teacher use to evaluate an argument in terms of convincingness?

E. What criteria does the teacher use to evaluate an argument in terms of validity?

F. What criteria does the teacher use to score the arguments like they were responses of his students?

1.3 The Significance of the Study

Although proof is significant for mathematical understanding, many students consider proof as a meaningless ritual (Ball et al., 2002). However, students at all levels need to see the importance of deductive reasoning and proof, and educators have a critical role in helping them (Hanna & de Villiers, 2008). As emphasized, both the teachers and students were convinced by empirical arguments (Stylianides et al., 2017). Therefore, investigating the proof evaluation of the teachers and the students is significant. In addition, Stylianides et al. (2016) mentioned the limited number of studies with middle and elementary school students. This study can contribute to this gap in the literature. Moreover, there were a limited number of studies in the Turkish context. Thus, the proof evaluation in different contexts is significant to be investigated. Furthermore, many studies focused on the proof evaluation of the teachers and the students in separate ways. The comparison of teachers' and their students' evaluations for proof arguments is significant to be researched in terms of getting a deeper understanding of their proof concepts. Finally, the findings can be beneficial for the improvement of classroom instruction, curriculum, and teacher education programs.

1.4 Definitions of the Important Terms

Proof

In the current study, the definition of Stylianides (2007), which was prepared for school mathematics, was used:

"*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 definition was used in the study because it is flexible to focus on proof in different classroom contexts and levels.

Proof evaluation

Pfeiffer (2011) defined proof evaluation as:

"*…determining whether a proof is correct and establishes the truth of a statement (validation) and also how good it is regarding a wider range of features such as clarity, context, sufficiency without excess, insight, convincingness, or enhancement of understanding*." (p. 5)

CHAPTER 2

LITERATURE REVIEW

The study aims to investigate the proof evaluations of a mathematics teacher and his students. Firstly, the framework is presented to understand what kind of arguments can convince the participants. Secondly, the studies about proof evaluation of teachers were reviewed, and then the third section focused on the studies about the proof evaluation of students.

2.1 The Proof Scheme Framework

A framework belongs to Balacheff (1988), which occurred based on pupils' experimental approach. He mentioned four main types of proof: naïve empiricism, which refers to checking truthness with several examples; the crucial experiment, which is selecting crucial examples such as complicated or extreme where the outcome of the experiment should be different based on the hypothesis; the generic example which includes "*the characteristic properties and structures of a class,*" and the thought experiment which "*invokes action by internalizing it and detaching itself from a particular representation*" (p. 219). And he mentioned there is a hierarchy of these proofs.

The framework of Harel and Sowder (2007 & 1998) includes three main titles: External Conviction Proof Schemes, which refer to removing doubts by external elements such as ritual forms and sources of knowledge; Empirical Proof Schemes which refer to validating by examples and perceptions, and Deductive Proof Schemes which refer to validating by logical deductions. Each title has subtitles based on what kind of argument can convince the individual or community. Figure

1 shows the map of these categories based on the adaptation of the definitions in Harel and Sowder (2007).

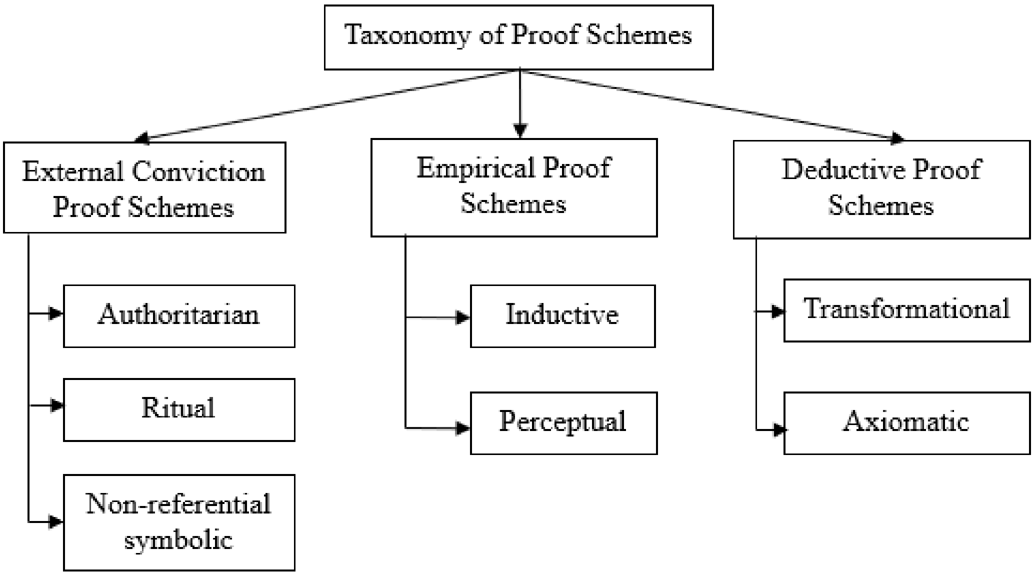


Figure 2.1 The Taxonomy of Proof Schemes (adapted from Harel & Sowder, 2007, p. 7)

The first proof scheme of Harel and Sowder (2007 & 1998) is the *External Conviction Proof Scheme*. In this scheme, one can be convinced by a teacher or book as the authority, which is called an *authoritarian proof scheme*. What an authoritarian figure says can remove doubts. Or, rigorously used argument appearance (e.g., proof needs to be seen as the format they learned in class) can convince an individual or a community, which is called a *ritual proof scheme*. This scheme is similar to the syntactic level in the model of Martin and Harel (1998). As the third subtitle is called as a *non-referential symbolic proof scheme,* the arguments include symbols but not referencing any meaning. In short, in the External Conviction Proof Schemes, the external elements play a role in convincing an individual or community.

The second one is *Empirical Proof Schemes*. Sometimes, one can be convinced by testing the conjecture with some examples, called an *inductive proof scheme*. These examples can be direct measurements or numbers. The naïve empiricism and the crucial experiment in the framework of Balacheff (1988) can be considered as inductive proof schemes. Or, an individual or community can be convinced by the perception, called a *perception proof scheme*.

The third and last one is *Deductive Proof Schemes*. As the name implies, the mode of argumentation is deductive reasoning in this scheme. Convincing arguments are general for all related cases and include logical inferences. There could be an example case, but these cases are purposefully selected, and the explanation is logical, namely a *transformational proof scheme*. The generic example in Balacheff (1988) can be considered as an example of this scheme. Or, in the deductive proof scheme, there could be a chain of arguments in the logic of p => q => r, which may be called an *axiomatic proof scheme*. It is an extension of the transformational proof scheme with an understanding of the proving process, where it starts with facts and accepted statements.

To sum up, the External Conviction Proof Scheme depends on an external authority, which can be a teacher, classroom practices, or mathematical symbols. The Empirical Proof Schemes mostly focus on empirical and perceptional arguments, and the Deductive Proof Schemes, which refers to valid proof, use deductive reasoning. This deductive reasoning can be expressed by a chain of arguments or investigation on a specific example to emphasize the logic. It should be noted that, unlike the hierarchy of the framework of Balacheff (1988), these schemes are not mutually exclusive, and one can have more than one proof scheme (Harel & Sowder, 1998).

The proof scheme framework of Harel and Sowder (2007 & 1998) was considered to be one of the most used frameworks to investigate proof evaluations of teachers (e.g., Dalkılıç & Zeybek Şimşek, 2022) and middle school students (e.g., Heinze & Reiss, 2003; Liua et al., 2016). In addition, the proof scheme framework of Harel and Sowder (2007 & 1998) mentioned the external conviction proof scheme besides the experimental and deductive proof schemes, unlike the others. In addition, Harel and Sowder (1998) defined the proof scheme not only as a proving process but also as "*a cognitive stage, an intellectual ability, in students' mathematical development*"

(p. 244). From that perspective, this framework might be considered inclusive. Because of these reasons, in the current study, this framework was taken as basis. The framework guided the study to prepare the arguments and detect the participants' reasonings for justification.

2.2 Research on Teachers' Proof Evaluation

The educators are the one who leads the students to have deductive reasoning to reach mathematical proof (Hanna & de Villiers, 2008). Thus, their approaches to the proof are one of the essential elements to enhance proof concepts in a class. How teachers evaluate arguments that try to prove a statement has been one of the components of the proof evaluation. This section presents past studies regarding teachers' proof evaluations.

For instance, Ko and Hagen (2013) worked with 55 middle school mathematics teachers from nine schools; the participants attended an online survey that included a number-based, algebraic, and visual argument about the conjecture in numbers. The number-based argument included a few trials with numbers but not deductive generalization, the algebraic argument had deductive generalization, and the visual argument investigated on specific visual representation to focus the structure. The teachers were asked to rate these arguments in terms of being convincing and describe their criteria for these decisions. Based on the results, the majority of the teachers were not convinced by the numeric argument and mentioned the necessity of generalization for being convincing. Some considered the easiness of understanding as another criterion for convincingness. In addition, a few teachers suggested algebraic arguments to include some specific numeric examples. In terms of proof or not, more than half of the teachers considered algebraic argument as proof, and around half of them considered visual argument as proof. In many cases, having algebraic rules and mathematical symbols were the criteria for accepting as proof.

As another study on teacher proof evaluation, Boero (2006) showed a proof prepared by a brilliant 8th-grade student to 36 high school math teachers. The argument did not include an algebraic form. While 11 of the high school teachers accepted the student's proof since they thought the reasoning was substantial, other teachers did not accept it since they thought the expression was ambiguous or the argument needed to be more general and needed algebraic expression; examples did not represent proof. However, after the teachers were asked to prove the conjecture, some teachers changed their minds and accepted the student's argument because the algebraic proof was too difficult. Some teachers still considered the argument as unacceptable despite the difficulty of the proof. Therefore, while many teachers were looking for algebraic equations and generalizations, some teachers considered the argument without algebra as proof because of the difficulty of producing algebraic proof or substantial reasoning in the argument.

Unlike the majority in the studies of Ko and Hagen (2013) and Boero (2006), empirical arguments were found sufficient for some teachers. Knuth (2002b) interviewed 17 secondary mathematics teachers to understand their proof conceptions. In one part of the interview, researcher-constructed arguments were presented to be evaluated in terms of instructional appropriateness (i.e., whether they can convince the students). Based on the results, the proof conceptions were categorized under three groups: the formal, which refers to arguments under prescribed mathematical formats; the less formal, which is not necessarily in prescribed format but refers to the truth for all cases; and the informal, which are empirical-based arguments. All participants were aware of the limitations of such empirical-based arguments. However, most of them did not consider formal proof as central for secondary school because they did not think such proof was appropriate for all grades. Instead, they considered informal proofs as central to school mathematics.

The use of examples does not always make an argument inductive. It can be an argument from a transformational proof scheme. It can be generic proof where "*conjectures are interpreted in general terms, but their proof is expressed in a particular context*" (Harel & Sowder, 1998, p. 271). Raid and Vargas (2018) said that generic examples can be considered proof if they convince the reader with deductive reasoning and conform to the classroom. There are some studies that investigated the teachers' approaches towards generic examples. For instance, Dogan and Williams-Pierce (2021) reported from a study that included 12 in-service middle school teachers in terms of their approaches to generic examples in a course. The teachers seemed to be more convinced if generic examples included visual representation. Also, they thought the generic argument needed to include variables to be more convincing.

In terms of using generic examples in classrooms, teachers can consider them as convincing arguments for their elementary-grade students. For instance, Isler (2015) surveyed 80 elementary teachers and interviewed 20 of them to investigate their expectations about reasoning and proof in school mathematics. Based on the results, many teachers who evaluated several argument types considered the visual generic example argument the most convincing for their students. because they are helpful for not only understandability of the student but also showing why the statement was true was also the reason to be convincing for the students. Many classroom teachers saw examples unacceptable from students and would ask to show why or how. On the other hand, many teachers thought the students would use examples to make justification, and some teachers stated that the students would rely on authority. Moreover, like the teachers in Knuth (2002b), many teachers in the study of Isler thought the students would find the deductive argument and rule without backing the least convincing because the students would not understand the arguments.

The studies on teacher evaluation were also conducted with pre-service teachers. Morris (2007) interviewed 34 pre-service teachers. At the first stage of the study, the participants were randomly divided into two groups, and while one group read a transcript of 3rd grades lesson about arguments on "the sum of two odd numbers is always even", the other group read the same transcript but eliminated general deductive arguments. The pre-service teachers were asked to rank the best three arguments and decide the statements' generalizability, correctness, and understandability by the students. In the second stage, five arguments on "$n^2 + n$ *will always be even*" were presented (p. 488), and the participants were asked to decide their correctness. Based on the results, the pre-service teachers were found to be aware of which argument includes an explanation of the truth of the generalization. In addition, teachers who read transcripts with non-valid arguments tend to think the examples can prove the generalization.

In the Turkish context, the research usually focused on college students, including preservice teachers. For instance, one part of the analyses of the study of İmamoğlu and Yontar Toğrol (2015) aimed to investigate senior students enrolled in Mathematics, Secondary School Science and Mathematics Education, and Primary Education programs regarding proof evaluation. The result showed that the seniors were mostly aware of the limitations of using examples and were able to distinguish between inductive and deductive arguments.

In another research, Uygan et al. (2014) studied with three prospective mathematics teachers. As one part of the study, the participants evaluated arguments where one of them included examples of the statement on the dynamic geometry software, and the other one was in an axiomatic system but mistakenly used a result of the statement as an accepted statement. The participants mentioned that proof should include generalizability, convincingness, and understandability. While one of the participants emphasized the importance of deductive structure and the necessity of generalizability instead of giving examples, the same one did not notice the mistakenly accepted statement. Another participant noticed that disruptive element but accepted the empirical argument because of convincingness and understandability. The other one focused on similar proof in the courses instead of axiomatic structure. It can be said that prospective teachers might focus on different aspects regarding proof evaluations.

Like Uygan et al. (2014), Çontay and Duatepe Paksu (2019) studied with three prospective mathematics teachers. During the interviews, it was observed that one might have more than one proof scheme of the framework of Harel and Sowder (2007) based on the responses of the participants. However, the responses mostly belonged to the ritual proof schemes. In addition, the results showed that the empirical arguments did not mostly convince the participants. On the other hand, in the study of Dalkılıç and Zeybek Şimşek (2022), where teacher candidates and middle school (5 – 8th grade) teachers participated, the participants considered examples useful for classroom setting while they were convinced by analytical (deductive) level arguments, which was similar to Knuth (2002b). Moreover, they had a tendency to use examples when they struggled to create an argument. In addition, like the findings of Çontay and Duatepe Paksu (2019), the participants tended to rely on external factors such as mathematical symbols or methods.

2.3 Research on Students' Proof Evaluation

The students were the other component of the current study. Understanding which kinds of arguments convince students and why they are convinced can help to explore their proof conceptions. This section aims to present past studies about students' proof evaluations.

In one of the studies, Bieda and Lepak (2014) interviewed 25 middle school students about their evaluation of arguments in number theory to understand the difficulty of generating proof and see what was convincing for the students. The participants faced two categories of arguments to justify "the sum of two consecutive numbers is odd." One of the arguments was a general argument, and the other one was an example-based argument. The students were asked to evaluate in terms of convincingness and express ideas to improve the argument that they found less convincing. Based on the results of the study, a minority of the students regarded the general argument as convincing. The minority considered the argument should include more explanation instead of giving a bunch of examples. The majority of the students found the examples-based argument more convincing because they considered examples as essential for mathematical arguments and enhancing understanding. Moreover, in terms of improving the arguments, adding examples to “show” how it works and adding explanations to “tell” why it works were suggested. Thus, empirical arguments were found to be more convincing than general arguments by the students.

Even though empirical arguments may convince the students, they may also be aware of the limitations of these arguments. For instance, based on the interviews conducted with eleven 8th-grade students by Heinze and Reiss (2003), many students considered the empirical arguments as proof; however, most of them thought these arguments were not in a proof form. Similarly, the middle school students in the study of Healy and Hoyles (2000) were aware of the limitations of examples. Healy and Hoyles surveyed 2,459 fourteen-fifteen-year-old students in 90 schools and interviewed teachers and students in 10 schools. The aim was to investigate the characteristics of the arguments that were considered as proof, the reasons behind these evaluations, and the proof constructions of the students. Based on the results, the students considered examples beneficial to get convinced about the trueness of the statements, but they thought such empirical arguments were limited, not general. Despite the awareness of the limitation of examples, they mostly employed empirical arguments to construct their proofs. On the other hand, they considered such empirical arguments as possible to get the least mark from the teachers. Students believed the arguments using algebraic expressions would get a higher mark, while many teachers thought the logic of the argument would be essential.

Another study about student proof evaluations belonged to Liua et al. (2016). They interviewed eight 8th-grade students to evaluate the arguments’ correctness, showing always true, understanding, and supporting the conclusions. They determined that while the students were convinced by testing examples, some of them were aware of the limitations of the examples and found induction unreliable. On the other hand, only one student found algebraic deduction reliable. While some thought algebra was clear and concise, most of them did not see algebra as the representation of general cases. Instead, the transformation (being aware of the general property and connecting it with examples) and perceptual connection (visualization and intuition) were frequently found convincing as the mode of argumentation. In addition, the students preferred numerical and narrative arguments more than algebraic ones in terms of understandability. The students used easiness to understand as a criterion, but what made it easy was inconsistent. Easiness to understand could be easy examples and illustrations, familiarity with reasoning procedures, real-life scenarios, or not including complex procedures. But, in pictorial illustrations, while some students found them helpful for understanding, they could also be confusing based on images and diagrams. In addition, there were some students who did not consider the easiness of understanding. For instance, one student considered simple or straightforward arguments convincing and did not mention the easiness of understanding. Thus, there was variety among the evaluation ways of the students.

One of the factors that affect the criteria used by students to evaluate arguments is the teacher. Ozgur (2017) observed a course with thirty-one 9^{th} - 10^{th}-grade students and interviewed the teacher of the course and 18 of the students to explore the proof concepts of both the teacher and her students through investigating their proof descriptions, proof evaluations, and proof productions. In terms of proof evaluation, there were students who considered deductive arguments as proof and the students understood the deductive arguments but also accepted some empirical arguments as valid proofs. However, all students used the validity of the argument for all cases and explained why as criteria. In addition, the teacher of the course also gave importance to understanding why the statement was true and asked the students to seek why it is true in general. At the end of the study, Ozgur (2017) indicated that the students' proof conceptions were mostly similar to the proof conception of the teacher, and she mentioned that the teacher's instructional practices had an important impact on shaping the proof conceptions of the students.

In the Turkish context, there were limited studies with middle school students, and they focused on proof construction. For instance, Sen and Guler (2015) studied with 7th-grade students, and it was observed that one student might use different proof schemes at the same time. The results showed that the students' proofs were mostly from external conviction and empirical proof schemes. The axiomatic proof was not used at all, and transformational proof was not used so much. The students tend to use the trial-error method to prove statements instead of formal proof. Similarly, Aydoğdu İskenderoğlu (2003) investigated 5th to 9th grades' justifications through mathematical problems, and it was noted that the students used mostly external conviction and empirical proof schemes, and there were fewer deductive proof schemes. In addition, the students' proof schemes tended to turn from external conviction proof schemes to empirical and then to deductive proof schemes as their grades increased. As a proof evaluation, one part of the study of Arslan (2007) asked 6th to 8th grades to evaluate axiomatic, transformational, and empirical arguments that tried to prove one statement. While most students mentioned that they liked and understood the transformational argument and preferred less the axiomatic argument, they thought the axiomatic argument would get high points from the teacher.

To sum up, while some teachers considered the algebraic deductive arguments as proof and looked for algebraic generalization as a criterion (e.g., Boero, 2006; Ko & Hagen, 2013) and noticed the limitation of the empirical arguments (e.g., İmamoğlu & Yontar Toğrol, 2015), they might prefer the empirical arguments to convince their students regarding understandability (e.g., Dalkılıç & Zeybek Şimşek, 2022; Knuth, 2002b). In addition, some teachers looked for explaining why as another criterion and preferred generic examples in their classrooms (e.g., Isler, 2015). And they might look for visuals and variables to be more convinced even if it was a generic example (Dogan & Williams-Pierce, 2021). In addition, some teacher candidates thought proof should include generalizability and understanding (Uygan et al., 2014), and some were influenced by external factors such as mathematical symbols and methods (e.g., Çontay & Duatepe Paksu, 2019; Dalkılıç & Zeybek Şimşek, 2022).

In terms of proof evaluation of students, the studies showed that they could be convinced by testing some examples (e.g., Bieda & Lepak, 2014; Heinze & Reiss, 2003). Similarly, in the Turkish context, the students' making proofs were more external conviction and empirical proof scheme than the deductive proof scheme (Aydoğdu İskenderoğlu, 2003; Sen & Guler, 2015). On the other hand, students in national and international contexts might think the teacher would prefer the algebraic expression. (Arslan, 2007; Healy & Hoyles, 2000). In addition, some students were aware of the limitations of empirical arguments and looked for explaining why criteria like the teachers (Healy & Hoyles, 2000). Also, they might use criteria such as easiness for understanding, visuals, algebra, and being simple or straightforward to evaluate the arguments (e.g., Liua et al., 2016). In terms of the comparison, the literature had evidence for the similarity of the proof conception of the teacher and her students (Ozgur, 2017).

CHAPTER 3

METHODOLOGY

The study aimed to understand the approaches of both a teacher and his students to proofs in number and geometry learning areas at the middle school level. Thus, this study was conducted to investigate similarities and differences between a mathematics teacher's and his students' proof evaluation. The purpose was to explore criteria used to evaluate arguments that were made to show the truth of statements and investigate how these criteria are similar or different for the teacher and his students. To sum up, the following research question and its subquestions were investigated in this study:

What are the similarities and differences between the proof evaluation of a mathematics teacher and his 8th-grade students?

A. What criteria do the students use to evaluate an argument in terms of convincingness?

B. What criteria do the students use to evaluate an argument in terms of validity?

C. What criteria do the students expect their teacher to use for scoring an argument?

D. What criteria does the teacher use to evaluate an argument in terms of convincingness?

E. What criteria does the teacher use to evaluate an argument in terms of validity?

F. What criteria does the teacher use to score the arguments like they were responses of his students?

In this chapter, the details about how the study was conducted were given. The methodology of this study was presented under the titles: Design of the Study, Participants, Data Collection, Data Analysis, Reliability and Validity, Ethics, and Limitations.

3.1 Design of the Study

Qualitative studies are used to explore and understand issues and problems (Creswell, 2007). In the current study, the purpose was to seek to explore how similar or different a mathematics teacher and his students' proof evaluation are. In other words, the study sought "*the quality of the relations*" and situations, which implies qualitative research (Fraenkel et al. 2012, p. 426). Thus, the design of this study was suitably selected as qualitative research to get a deep insight into the proof evaluation.

Qualitative research can be conducted in different approaches, and one of them is the case study. Carswell (2007, p. 73) calls "*the study of an issue explored through one or more cases within a bounded system*" a case study. The case studies allow us to investigate a phenomenon in different data sources to get an understanding through a "*variety of lenses*" (Baxter et al., 2008, p. 544). Since this study aims to explore the proof evaluation of both students and their mathematics teacher, the case study is convenient for the design of the current study to get multiple facets of proof evaluation in a class. Although there were two components (i.e., the students and the teacher), proof evaluation was investigated in one class as a bounded system. The cases of the study can be one or more individuals/organizations (Fraenkel et al., 2012). In the current study, the class members formed one organization; so, it can be called a single case study. Yin (2009) mentioned that a single case study can include subunits to be analyzed, called an embedded case study. It can be said that the teacher and the students were the subunits of the single-class organization. As a result, the design of the study can be named as an embedded single case study.

3.2 Participants

Twenty 8th-grade students and their mathematics teacher participated in the study. They were invited to the study by purposive and convenience sampling methods. Fraenkel et al. (2012) call selecting participants based on prior information as purposive sampling, which is commonly used in qualitative studies. In the current study, the participants need to have basic knowledge included in the argument. Since students become more familiar with using algebra and geometry in the 8th grade, an 8th-grade class was purposely chosen. Moreover, it was aimed that the teacher and his students know each other for enough time to be able to evaluate arguments thinking from each other's perspectives. In the current study, the teacher had been the mathematics teacher of the class for three years. In other words, the participants were purposefully selected to fit these criteria and to have required previous knowledge.

In terms of convenience sampling, Stake (1995) suggests selecting easily accessible participants who may lead us to get an understanding. Also, Creswell (2007) suggests convenience sampling to save time and effort, which means selecting participants based on accessibility and easiness of data collection. Firstly, the teacher was selected from a school that was close to the researcher, and he and the researcher would know each other for 12 years. Then, one of his classes was chosen, whose lesson times were suitable for the researcher. The students were invited to the study, and they voluntarily participated. In short, the participants were conveniently selected based on their accessibility and voluntariness.

In the current study, the students first attended a questionnaire, and then five of them participated in the follow-up interviews. These interviews were conducted to get a deep understanding of the answers to the questionnaire. As Stake (1995, p. 4) suggests, the first criterion to choose was "*to maximize what we can learn*." So, the names of the students who are thought to have more to say in their written responses were listed. Then, these names were shown to the teacher (as an informant), and he chose names who were talkative and possibly would not hesitate to express an opinion. After this consultation, the students were invited to the follow-up interview, and two boys (i.e., students 12 and 13) and four girls (i.e., students 5, 9, 14, and 15) volunteered to join the follow-up semi-structured individual interview.

3.3 The Context of the Study

The study was conducted in a middle school in Gaziantep. The school had 75 teachers and around 1,500 students. The socioeconomic status of the parents was usually medium or low. The capacity of the classroom was around 30 to 40 students. The selected class had 17 girls and 18 boys (a total of 35 members), but only 9 girls and 11 boys (a total of 20 students) could participate in the study because the participation rate got low after the earthquake happened in April 2023. The study was conducted in the second semester of the 2022-2023 education year, and there had not been any educational activity in this year for two months because of the earthquake. Because of that, the study was conducted three months after the earthquake and after being sure that the students had readapted the school life routine.

Because the students were 8th-grade, they had been preparing for an exam to be accepted to a qualified high school (called high-school entrance system exam) throughout the education year. The students had been taking 5 hours of mathematics courses as compulsory and 2 hours of applications in the mathematics courses as selective per week. The mathematics teacher of both courses was the same person, and he mentioned that he had been teaching both courses as an ordinary compulsory mathematics course. In other words, the participating students had 7 lesson hours per week with the mathematics teacher. The teacher described the selected class as academically successful and highly participating in the lessons. In terms of teaching a lesson, the teacher claimed that he usually uses the direct instruction method and the smart board to reflect a textbook and show educational videos.

3.4 Role of The Researcher

The researcher played a role in creating arguments and questions in the written questionnaire. He also designed the interview protocols and implemented both data collection instruments for the teacher and his students. After the data collection and writing of transcripts, the researcher coded the data and reported the results.

The researcher was also a mathematics teacher during this time in the same district, but had not been to this school or knew the students before the data collection.

3.5 Data Collection

The data were collected through written questionnaires and semi-structured interviews, which are two of several data collection techniques of qualitative studies (Fraenkel et al., 2012). The purpose of the current study was to explore proof evaluations of the teacher and his students. So, there were two main sources of data: the teacher and the students. Firstly, 20 students participated in a written questionnaire. After collecting data from the written questionnaire, 5 of them (students 5, 9, 12, 13, and 15) took part in the interviews within two weeks. While selecting students for the interview, their voluntariness, their written responses to the questionnaire, and the teacher's suggestions about openness to speak were considered. After both the written questionnaire and follow-up interviews were compilated, an interview with the teacher was conducted.

3.5.1 Data Collection Tools

First, a written questionnaire was conducted on the students. The questionnaire included three tasks; two of them were from the numbers learning area, and one of them was from the geometry learning area. Each task included a mathematical statement and several arguments to show the truth of that statement. One of the statements in the number learning area was taken from Lovin et al. (2004), and the other two were created by the researcher based on the curriculum (MoNE, 2018). The statements used in the tasks are respectively given below.

Task 1 (Number Learning Area) – “The sum of any three consecutive whole numbers is equal to three times the middle number.” (Lovin et al., 2004, p. 3)

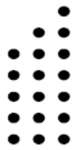
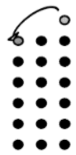
Task 2 (Geometry Learning Area) – “The sum of all exterior angles of any quadrilateral is always 360°.”

Task 3 (Number Learning Area) – Where the numbers x, y, a, b are natural numbers, $y \neq 0$ and $b \neq 0$; If there is an equation such that an $\frac{x}{y} = \frac{a}{b}$, we can always write an equation like “$x \cdot b = y \cdot a$”

Each task was designed to include arguments from each main proof scheme level of the framework of Harel and Sowder (2007), which are deductive, empirical, and external conviction proof schemes. For instance, Task 1 included two deductive, one empirical, and one external conviction argument. One of the deductive proofs was axiomatic, and the other one was transformational proof. The arguments and the argument categorizations can be seen in Table 3.1.

Table 3.1 The Arguments in Task 1

The Arguments	**The Proof Schemes**
Argument 1: I can take any 3 consecutive numbers. If n is the initial number, the middle number must be $n+1$, the last number must be $n+2$. If we sum 3 consecutive numbers; $n+(n+1)+(n+2)=3n+3$ Since the number $3n+3$ is 3 times the middle number $n+1$, the statement is correct.	Deductive Proof Scheme – Axiomatic Argument

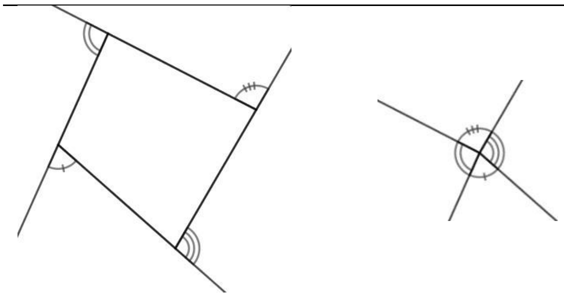
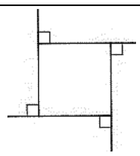
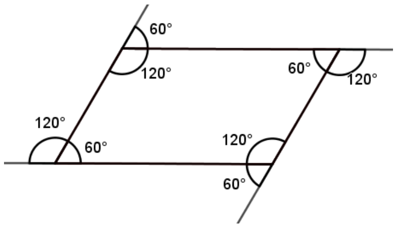
Argument 2: Imagine making three columns of dots representing any three consecutive numbers. The first column represents the first number, the second column represent the middle number, and the third column represent the last number.  I can take the top marble from the last column and move it to the first column. This makes the number of dots in each column the same as the number of dots in the middle column.  Since the total number of dots is always three times the number in the middle column, the statement is correct.	Deductive Proof Scheme – Transformational Argument
Argument 3: 5, 6, and 7 are three consecutive numbers, and 5 + 6 + 7 = 18, and 3 x 6 = 18. 7, 8, and 9 are three consecutive numbers, and 7 + 8 + 9 = 24, and 3 x 8 = 24. 569, 570, and 571 are three consecutive numbers, and 569 + 570 + 571 = 1710, and 3 x 570 = 1710. Since it works in these three examples, the statement is correct.	Empirical Conviction Proof Scheme – Inductive Argument
Argument 4: I can take any 3 consecutive numbers. If n is the initial number, the middle number must be $n + 1$, the last number must be $n + 2$. If we sum 3 consecutive numbers; $n + 1 + 2 = n + 3$ Since the number $n + 3$ is 3 times the middle number $+1$, the statement is correct.	External Conviction Proof Scheme – Symbolic Argument

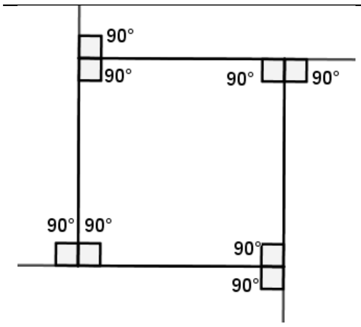
All the first three arguments were taken from Lovin et al. (2004, p. 3). The axiomatic argument represented the sum of the consecutive whole numbers with $(n-1)+n+(n+1)$ where n was the middle number. However, the argument was adapted to represent the sum of the consecutive whole numbers with $n+(n+1)+(n+2)$ where $n$wasthe smallest number because it might be harder to deal with positive and negative numbers at the same time. In addition, the arguments in Lovin et al. (2004) did not include an argument from an external conviction proof scheme. Therefore, argument 4 was created for the symbolic proof scheme. It was prepared to be a similar version of Argument 1 with incorrect algebraic operations and logic.

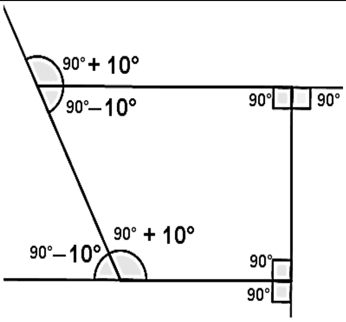
Task 2 included one external, two empirical, and two deductive arguments. Like Task 1, there were both axiomatic and transformational proofs. In addition, one empirical argument was perceptual, and the other one was inductive. Also, unlike Task 1, the external argument was an authoritarian argument, not a symbolic argument. All arguments were created by the researcher. The arguments and the argument categorization of Task 2 are presented in Table 3.2.

Table 3.2 The Arguments in Task 2

The Arguments	The Proof Schemes
Argument 1: In the textbook, it is written that the sum of all exterior angles of quadrilaterals is 360°. Thus, the above statement is true.	External Conviction Proof – Authoritarian Argument
Argument 2: Let’s draw any quadrilateral and its exterior angles.	Empirical Proof Scheme – Perceptual Argument

 If we cut and put together the exterior angles, a round angle occurs, which means it makes 360°.	
Argument 3: As an example, let’s take a square or a rectangle. Each of the exterior angles is 90°, and 4 right angles make 360° in total.  Or let's take a parallelogram in which one of the angles is 60°. The opposite angle must be 120°. Also, the exterior angles are 2 times 120° and 2 times 60°, so the sum of them makes 360° again.  The sum of the exterior angles of both a square and a parallelogram is 360°. Since it is correct for the examples I tried, the result will be 360° no matter which quadrilateral we try.	Empirical Proof Scheme – Inductive Argument
Argument 4: Let's take a rectangle. The sum of the interior angles of a rectangle is 360°, and the sum of the exterior angles is 360°. We can get all the other quadrilaterals by playing with the angles of the rectangle.	Deductive Proof Scheme – Transformational Argument





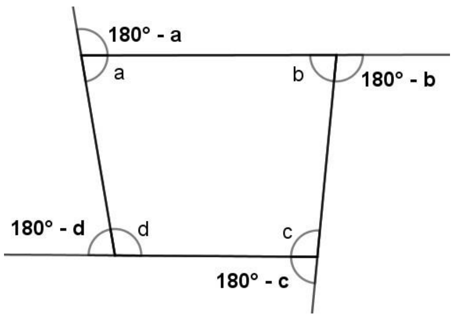
Let's increase one of its interior angles by 10°. Then the exterior angle of that angle must be decreased by 10°.

However, if we increase one interior angle, the other interior angles must also decrease by 10° as a total to ensure that the sum of the interior angles is 360°. Then, the exterior angles must increase by 10° in total.

So, no matter how much we play/change with the angles, the sum of the exterior angles will not change; it will be 360°.

Argument 5:

Let's draw any quadrilateral. Let's name the angles a, b, c, d. The sum of the interior angles = a + b + c + d.



Deductive Proof Scheme – Axiomatic Argument

The sum of the exterior angles =

= (180°– a) + (180°– b) + (180°– c) + (180°– d)

= 180° + 180° + 180° + 180° – a – b – c – d

= 720° – (a + b + c + d)

= 720° – (the sum of the interior angles)

We know that the sum of the interior angles of a quadrilateral is 360°.

= 720° – 360°
= 360°

There was one external, one empirical, and one deductive argument in Task 3. Like Task 2, it included authoritarian, inductive, and axiomatic proofs. All arguments were created by the researcher. The arguments and the argument categorization of Task 3 are presented in Table 3.3.

Table 3.3 The Arguments in Task 3

The Arguments	The Proof Schemes
Argument 1: Our teacher told us that if there is an equation like this, according to the cross-multiply, we can multiply the numerators with the denominator of the opposite number and make they must be equal.	External Conviction Proof Scheme – Authoritarian Argument
Argument 2: $\frac{4}{10} = \frac{2}{5} \quad => \quad 4 \cdot 5 = 2 \cdot 10$ $\frac{7}{3} = \frac{21}{9} \quad => \quad 7 \cdot 9 = 21 \cdot 3$ $\frac{5}{8} = \frac{500}{800} \quad => \quad 5 \cdot 800 = 500 \cdot 8$ Above, I tried multiple examples, and the statement is true for all of them. This means if $\frac{x}{y} = \frac{a}{b}$, we can write an equation as “$x \cdot b = a \cdot y$”.	Empirical Proof Scheme – Inductive Argument
Argument 3: $\frac{x}{y} = \frac{a}{b}$ The information on the left is given. Now let's make the denominators of these two rational numbers equal.	Deductive Proof Scheme – Axiomatic Argument

$$\frac{x}{y} = \frac{a}{b} \quad \rightarrow \quad \frac{x \cdot b}{y \cdot b} = \frac{a \cdot y}{b \cdot y}$$

(b) (y)

The denominator of two fractions is equal (y$\cdot b = b \cdot y$); therefore, their numerators must be equal, meaning it must be $x \cdot b = a \cdot y$.

As part of the written questionnaire, students were asked to evaluate these arguments in terms of convincingness, validity, and teacher expectancy. As the first question, the students were asked to score the argument in terms of convincingness by using a 10-point Likert scale and to explain their reasoning for each score. Thus, the first question of the survey was like the following sentence:

"*Score the given arguments out of 10 in terms of their convincingness. Explain why you score them like that. 10 points mean that I found it very convincing, and 1 point means that I did not find it convincing at all.*"

The "convincingness" term was used instead of the term proof, like Bieda and Lepak (2014) did, because the terms "proof" and "proving" are not included in the middle school curriculum. This question was designed to respond to the sub-research question A.

The second question was related to the validity of the arguments. The participants were asked to evaluate the arguments to decide whether the argument always works. In other words, the participants selected arguments considered valid arguments. This question served to get answers to the sub-research question B. The second question was like the following sentence:

"*Which student or students were able to demonstrate that the given statement always works? Explain your reasoning*."

The final question aimed to explore the students' expectancy for their teachers' scoring way of these arguments. This question was expected to help see the similarities and differences in proof evaluations between them and their teachers with their perspectives. In other words, the question aimed to find out the criteria that the student thought the teacher would use to evaluate these arguments. So, it served to sub-research question C. The third question was like the following sentence:

"*Please indicate how many points you think your teacher would give if he scored these arguments. Explain why you think your teacher would score them like that. 10 points are the maximum score, and 1 point is the minimum score*."

To conclude, the questionnaire included 3 tasks, which included mathematical statements in number and geometry learning areas. Several arguments were trying to show whether the statement was true or not. Four arguments were in Task 1, five arguments in Task 2, and three arguments were in Task 3. Each task in the written questionnaire had 3 questions that wanted to evaluate arguments in terms of convincingness, validity, and teacher expectancy (see Appendix B).

After collecting data through the written questionnaire, the names of students who gave interesting answers were noted for the follow-up interviews to explore their opinions deeply. And then, the teacher was consulted about whether these names were talkative and open to sharing their opinions. After the consulting, these students were invited, and 5 of them volunteered for the interview. During the interviews, voice recording was used. They were semi-structured interviews. Semi-structured interviews have base questions, and there are several optional or alternative questions to be asked based on given responses. The base interview questions were the same as the written questionnaire. The students were asked to go through their answers and provide more explanations of their evaluation ways and criteria. Then, if there was any marginal or unclear criterion, the student was asked to explain more.

Another subunit of analysis was the mathematics teacher of these students. Another interview protocol was prepared to collect data from the teacher. The same tasks (i.e., the same mathematical statements and the respective arguments) were presented to the teacher. He did not see the tasks before the interview. By starting from Task 1, the statements and the arguments were presented. Firstly, like the students, he was asked to evaluate arguments in terms of convincingness by using a 10-point Likert scale. This question aimed to look for the answer to research question D. Secondly, he was asked to select arguments that were able to show the statement was always true. In other words, he chose the arguments which he thought were valid. This question served to research question E. Then, he asked to score the arguments by imagining they were answers from his students. This question tried to investigate the research question F. The aim of this question was also to determine whether the evaluation criteria for his students would be different from the evaluation criteria for the convincingness and validity of the argument. To sum up, the teacher was asked to evaluate the arguments in terms of convincingness and validity and to score them like they were responses of his students.

After the teacher made an evaluation, the questions in the student-written questionnaire were mentioned to him. He was asked to rank arguments based on which argument he thought his students found the most or the least convincing. After the ranking, the actual rank from the written questionnaire results was provided, and the teacher made comments about why his rank and the actual rank were similar or not. Similarly, these rankings procedures were repeated for validity and teacher expectancy. The aim was to use the teacher's interpretation to see whether there was any unexpected possible reason for the relationship between the teacher and his students. The interview protocol for the teacher is given in Appendix C.

3.6 Pilot Study

A pilot study was conducted in the first semester of the 2022-2023 education year to test the understandability and appropriateness of the items. A mathematics teacher and her twenty 8^{th}-grade students participated in the pilot study. The school had similar socioeconomic and sociocultural backgrounds to the school in the main study, and its location was close to the school where the main study was conducted. The data collection procedure was the same. The students first attended to the written questionnaire, and then the students who might have had more to say were chosen. The names were shared with the teacher to select talkative ones to maximize what could be learned. The selected students were invited to follow-up interviews, and three of them volunteered to participate. After the data collection from the students was completed, an interview was conducted with the teacher.

While the data collection procedure was the same as the main study, the data collection tools changed based on the results of the pilot study and expert opinion. For instance, in the pilot study, there was an extra symbolic argument in Task 2. The interviews showed that the argument was not understandable. Moreover, the expert opinion was that the argument did not fit the intended proof scheme. Since there was already another argument from the external proof scheme, the symbolic argument in Task 2 was removed.

Like the symbolic argument in Task 2, the transformational argument in Task 3 was not understood by the students in the pilot study. Again, the feedback from the expert was that the argument was not suitable for a transformational proof scheme. In other words, this argument did not serve its purpose. Since there was already another argument from the deductive proof scheme, the transformational argument in Task 3 was removed.

Another critical change in the tasks was validity. When the students were asked to decide which argument always works, some of them tended to choose the arguments that included the word "always." Since this situation caused misleads about judgments on validity, the words "always" were deleted from the arguments. After revising the language of the arguments to be more understandable, the changes in the tasks were completed.

There were some changes in the questions in the written questionnaire, too. In the pilot study, the students made extra evaluations of the arguments in terms of liking. The aim was to determine whether there was any difference or similarity between liking and convincingness. However, the expert opinion said that the liking term was not clear. Moreover, the criteria used for liking were similar to the criteria for convincingness. Since this question did not give any meaningful result and was unclear, it was removed from the questionnaire.

In the pilot study, the students were asked to rank the arguments in terms of convincingness and teacher expectancy; however, by ranking, it could not be seen whether the students found an argument convincing or not; it just showed how they compared the arguments. A student could find an argument most convincing by comparing it with others, but s/he might still find it not so convincing, or vice versa. Thus, the ranking questions were adapted to scoring questions to see how an argument was evaluated individually. Similarly, there exist some studies about proof evaluation using Likert scales, which is a form of rating scale (e.g., Healy & Hoyles, 2000). In the current study, there is a scoring scale for each argument that asks about convincingness and teacher expectations.

The final change was to add a table to have students explain their reasons for each argument separately. In the pilot study, the students were asked to explain the reasonings for the ranking; however, many of them did not write any explanation, or their explanation did not include reasoning for each argument. A table that asks the reasoning for the scoring was added to direct students to examine and write the criteria for the arguments individually.

To sum up, by using results from the pilot study and expert opinion, the non-understandable arguments with unsuitable proof schemes were removed, and other arguments were revised to be clear and prevent any misleading. After making changes to the questions of the written questionnaire to direct students to examine each argument, the data collection tools took the final form.

3.7 Data Analysis

The data was analyzed by using inductive thematic analysis, which "*involves identifying and coding emergent themes within data*" (Guest et al., 2013, p. 9). The responses to the written questionnaire and follow-up interviews were investigated to determine what criteria emerged during the evaluations. Creswell (2007) mentioned that the exact words in the data can be used to label the codes, which are called in vivo codes. Firstly, all emerged criteria formed in vivo code by considering the exact words the participants used. Then, by checking the interrelationships, some codes regrouped under new or current codes. In that way, a coding scheme occurred like in Table 3.4.

Table 3.4 Codes Emerged in the Analysis

Codes	Definitions	Examples
Explanatory / Understanding	The participant states that s/he understands what the argument explains or not.	*It is easy to understand / It is not so understandable / I don't understand / It is explanatory.*
Detail	The participant uses the terms such as "detailed" or "not detailed" in the explanation.	*... it could be more detailed / because it's so detailed.*
Simplicity	The participant states that the argument uses simple explanations and avoids confusing elements.	*It is simple / It is complicated / It explains with simple language.*
Practicality	The participant states that the argument is or is not time-consuming and/or easy or hard to apply. When the participant just wrote "easy," that is also considered under this code.	*It is practical and easy / It saves time / It takes time / It is easy / It is hard.*
Example	The participant states that the argument includes examples or that the use of examples affects his or her scoring.	*It shows that it works by giving more than one example / It gives examples.*
Visual	The participant states that the argument includes visuals or that the use of visuals affects his or her scoring.	*It increases memorability by using visuals… / Because s/he did with the visual.*
Operations	The participant states that the argument includes operations or	*Because it is with operations… / Because*

	that the use of operations affects his or her scoring.	*s/he did a longer operation.*
Algebra	The participant states that using algebraic elements such as variables affects the scoring. Sometimes, participants may implicitly refer to variables.	*…he [the teacher] likes algebraic expression… / it works always whatever number we put there.*
Source of Explanation	The participant states that the explanation is or is not based on an authority figure such as a textbook or teacher.	*S/he gave the teacher's explanation, did not show his/her explanation / it is not his/her own opinion*
Length	The participant refers to the length of the argument.	*It is too long / It is a long explanation /It is short.*
Logic	The participant uses the terms "logical" or "illogical" in his or her explanation.	*It didn't make much sense / It is very illogical / It makes sense to me / The explanation is logical.*
Liking	The participant states she likes or does not like the argument.	*I like a lot / I don't like / It is beautiful / It is bad.*
Teacher Method	The participant states that the teacher might or might not use the argument to explain the statement.	*It is exactly our teacher's method. / …not sure our teacher uses this method.*
Correctness	The participant refers to the correctness of the operations or information in the argument.	*The operation is correct…/ n+1 is not 3 times +1 /… I think the result is wrong.*
Validity	The participant comments that the argument always works or not. This code also is used when the participant refers to the correctness of the method to show the truth of the statement.	*… it doesn't show it always works / I think it always works / … I think s/he used a wrong way of explanation.*
Convincingness	The participant states that the convincingness of the argument affects the evaluation.	*It is not convincing / I think it won't convince him (the teacher).*

Scientific	The participant states that the argument is scientific or uses similar words.	*Because s/he scientifically explains directly.*
Rule	The participant states that there is a "rule" in the argument.	*Because s/he set a rule… / Because they all have certain rules.*
Provability	The participant uses the term "provability" in the explanation.	*… there is nothing in terms of provability / it is provable.*
Same With Me	The participant states that the teacher would think the same as him/her.	*I think my teacher would agree with me.*

In Table 3.4, the codes and their definitions with examples were given. To code a participant's response, each word that may represent one of the codes was noted. In other words, instead of whole sentences, the components in the sentences were coded. For instance, student 8 wrote, "*It explains the subject well, but it could have been explained with given examples.*" to evaluate the nonreferential symbolic argument in Task 1. In this example, since the sentence includes comments about both how well the argument was explained and how it could use examples, the codes were "explanatory/ understanding" and "example." As another example, student 12 wrote, "*Explaining with visuals makes it more logical, and this becomes more provable*." As can be noticed, the comment includes visual, logical, and provable words; thus, it was coded with "visual, logic, and provability" codes.

Although many of the criteria appeared in all questions of the written questionnaire (i.e., convincingness, validity, and teacher expectancy), some codes emerged in specific questions. For instance, a student might have considered an argument convincing because she thought the argument was valid. This would be coded as "validity." However, if she stated an argument valid for the "validity question" of the questionnaire, that does not show any criterion. In other words, the elements about validity were not coded as "validity" in the validity question.

Besides the above codes, there were answers that did not fit into determined categories, and some students did not give an answer. These were coded as "no response" and "other" codes, respectively.

To sum up, the coding scheme was formed by responses to the written questionnaire and interviews. Each term in a response was coded if it affected the evaluation. The exact words, in vivo coding, in the comments were used to label the criteria. Finally, each component in the sentence was coded, so a sentence could have more than one code.

3.8 Reliability and Validity

Fraenkel et al. (2012) mentioned that reliability and validity are usually expressed as "credibility" in qualitative studies. One of the ways to support the credibility of the study is triangulation, which refers to the benefit of the variety of approaches used in the study. Korstjens and Moser (2007) describe methodological triangulation as using varied instruments to collect data. In the current study, there was more than one instrument to determine the criteria of the students. First, the students participated in the written questionnaire, and then follow-up interviews were conducted as a part of the methodological triangulation.

As another way to ensure credibility, Korstjens and Moser (2007) suggest investigator triangulation, which means the cooperation of researchers and analysis of data by several researchers. Similarly, O'Connor and Joffe (2020) mention the necessity of two or more independent coders to ensure intercoder reliability. They also mention that a subset (around 10% - 25%) of the whole data can be randomly chosen to ensure inter-coding reliability. In the current study, the second coder, who was an expert mathematics education researcher, coded the randomly selected 20% of the data, and the agreement between the two coders was more than 80%. Moreover, after this comparison, two coders interacted for each suspense code in the remaining data to agree on the codes. In that way, it was tried to establish investigator triangulation.

During the development of the statements, arguments, and questions in the written questionnaire, expert opinions were taken from the researchers studying proof in mathematics education. It was consulted about whether the language, grade level and the proof schemes were appropriate to the purpose of the study. Moreover, as mentioned in data collection, a pilot study was conducted to be sure whether the questionnaire and interview protocol could give the intended data. By editing based on the result of the pilot study and expert opinions, the data collection instruments took the final forms.

To ensure the trustworthiness of a case study, another component is transferability (Baxter, 2008). Even though qualitative studies do not aim to generalize the findings, a thick description may enable other researchers to make judgments about whether the findings may be transferable to their studies (Korstjens & Moser, 2007). The features of the school and the class were detailly described in the context of the study, and this thick description may allow further researchers to make a judgment about transferable parts to their studies.

3.9 Ethics

Frankel et al. (2007) consider protecting participants from physical and psychological harm as one of the main responsibilities of the researcher. And in the current study, any possible discomfort situations were aimed to be avoided. Firstly, the study was conducted after getting the necessary permissions from the Middle East Technical University Ethics Committee and Gaziantep Provincial Directorate for National Education (see Appendix A). Moreover, the researcher took verbal consent from the school principal about the availability to conduct the study at the school. Then, the consent forms were distributed to inform the parents about the aim of the study and the procedure of data collection. The study was conducted with the students who had his or her parent's consent. Before implementing written questionnaires and interviews, the students were informed that they could leave any time they wanted. The study was conducted after their verbal consent. Similarly, the consent form and verbal consent were obtained from the teacher. To protect the privacy of the participants, written responses and voice records were only accessed by the researcher and his supervisor. And finally, while reporting the data, the actual names of the participants were not used.

3.10 Limitations

The design of the study is a qualitative study, so this study had no aim to generalize findings because of the nature of the research design. The data was collected by a teacher and a limited number of students in the described class. In addition, it should be noted that the participants had experienced an earthquake with a big impact, although the data was collected after adequate time after the disaster.

Besides the design of the study, this study was limited by the data collection tools and procedures used in the study. The data was collected by a written questionnaire and interviews. In addition, the tasks were limited to the number and geometry learning areas. Also, the study investigated the proof evaluation with only convincingness, validity, and teacher expectancy aspects. Moreover, it should be noted that there were no incorrect statements and no arguments to disprove the statement. The arguments were designed based on the proof scheme framework of Harel and Sowder (2007), but there was no argument created fitting the ritual proof scheme because the students had not been faced with the "proof" term to have a ritual to prove something. Furthermore, in terms of the inductive proof scheme, Hanna and de Villiers (2008) considered empirical argumentation valid if the conjecture implies a finite set and all possible cases are tested or if the proof arguments try to show a counterexample or prove the existence of such a case. For these conditions, empirical proof schemes are valid proofs; however, empirical proof arguments were less desirable for the current study because the tasks were limited with the statements implying an infinite set and there were no incorrect statements, as another limitation of the study.

CHAPTER 4

FINDINGS

The findings are presented under two main sections: the findings from the students and the findings from the teacher. The first section presented the descriptive results of the questionnaire, the qualitative results from the questionnaire, and the insights from the follow-up interviews. In each subtitle, the findings were investigated in terms of convincingness, validity, and teacher expectancy. The second section includes the findings from the interviews conducted with the teacher. And finally, in the end, the comparison between the findings of the students and the teacher was stated.

4.1 The Findings from the Students

The students were one of two embedded units of the current single case study. Firstly, the descriptive findings will be presented to report the numeric results of how the students evaluated the arguments. Secondly, the codes that emerged in the written parts of the questionnaires will be presented to investigate the criteria the students used. Thirdly, the insights from the follow-up interviews will be used to investigate selected students' proof evaluations. Finally, the summary of the findings from the students regarding both the written questionnaires and the interviews was presented.

4.1.1 The Descriptive Findings from the Written Questionnaire

As a reminder, the questionnaire had 3 tasks, and each task had one mathematical statement and several arguments. The statements in Tasks 1 and 3 were from the numbers learning area, and the statement in Task 2 was from the geometry learning area. Since the statements in Tasks 2 and 3 were prepared based on the curriculum, the students were familiar with the statements from 7th grade, but they faced the statement in Task 1 for the first time. The arguments were prepared for the proof schemes of Harel and Sowder (2007). Tasks 1, 2, and 3 had respectively 4, 5, and 3 arguments.

The written questionnaire asked the students to score the arguments in terms of convincingness and teacher expectancy and select the argument/s they found valid. For scoring, 1 was the lowest score, and 10 was the highest score they could give. The students repeated these evaluation steps for each task of the written questionnaire. The results of these scorings and selecting were analyzed under convincingness, validity, and teacher expectancy titles.

4.1.1.1 Convincingness

This section presents findings related to the convincingness scores of the arguments which the students gave. All scores for each argument were summed and divided by 20 to find the averages.

Table 4.1 The Average of The Scores for Convincingness in Task 1

	Axiomatic Argument	Transformational Argument	Inductive Argument	Symbolic Argument
Averages of The Scores	6.8	4.7	7.55	5.85

The averages of the scores in Task 1 are presented in Table 4.1. Based on the results, the average of the scores of inductive arguments was the highest. In other words, in the current case, the inductive argument was the most convincing. And axiomatic and symbolic ones, respectively, followed it. And the average transformational argument was the lowest one.

Table 4.2 The Average of The Scores for Convincingness in Task 2

	Authoritarian Argument	Perceptual Argument	Inductive Argument	Transformational Argument	Axiomatic Argument
Averages of The Scores	3.7	4.9	7.6	5.4	7.5

Similarly, Table 4.2 presents the average of the scores in Task 2. Based on the average scores, the highest average belonged to the inductive argument. Also, the average of the axiomatic argument was close to the average of the inductive argument. Then, the average of the transformational, perceptual, and authoritarian arguments respectively followed them.

Table 4.3 The Average of The Scores for Convincingness in Task 3

	Authoritarian Argument	Inductive Argument	Axiomatic Argument
Averages of The Scores	5.37	8.21	6.68

For Task 3, the average of the scores of the arguments are given in Table 4.3. The average score of the inductive argument was the highest average as it was in Tasks 1 and 2. Similarly, the average score of the axiomatic argument was the second highest. Like it was in Task 2, the authoritarian argument was considered the least convincing argument in Task 3.

To sum up, the inductive and axiomatic arguments were found to be convincing arguments, and the inductive arguments were the most convincing ones for each task. On the other hand, the transformational argument in Task 1 and the authoritarian arguments in Tasks 2 and 3 were found to be the least convincing arguments.

4.1.1.2 Validity

The students were asked to select arguments that show that the statement always works to understand their validity conceptions. It is important to note that the students were free to select none or more than one argument as valid. There were no students who selected all arguments as valid, but one student did not give a response to select valid arguments in Tasks 2 and 3 (So, the size was 19 students for these tasks). Based on the responses, Table 4.4 shows the frequencies of selected arguments for validity.

Table 4.4 The Frequencies of Selected Arguments for The Validity

Proof Schemes of The Arguments	Frequency of Selection in Task 1	Frequency of Selection in Task 2	Frequency of Selection in Task 3
Authoritarian	-	4	7
Symbolic	4	-	-
Inductive	13	11	14
Perceptual	-	2	-
Transformational	5	4	-
Axiomatic	7	9	6

In Task 1, the inductive argument was found to be valid by 65% of the students. The frequencies of axiomatic and transformational arguments, respectively, followed the inductive argument. The symbolic argument, which was the least selected, was found valid by 20% of the students.

In Task 2, the inductive argument was the most selected as a valid argument, and the axiomatic argument followed it in the second line. Both the authoritarian and transformational arguments were equally selected, and they were at the third line. Finally, as the least selected argument, the perceptual argument was found valid by around 10.5% of the students.

In Task 3, most students found the inductive argument valid. Around 36.8% of the students considered the authoritarian argument valid, and around 31.6% selected the axiomatic argument.

The results show that more than half of the students selected inductive arguments as valid arguments for each task. As the students found them convincing, inductive arguments were also the most selected arguments regarding validity.

The axiomatic argument was the second-most selected argument in Task 1 and 2, but it was the least selected in Task 3. Moreover, only in Task 2, around half of the students thought the axiomatic argument could show that the statement always works. While the axiomatic arguments are a valid way to show "the statement always works," more than half of the students in any task did not see them as valid arguments.

As the least selected arguments, the symbolic argument (from the external proof scheme) in Task 1, the perceptual argument (from the empirical proof scheme) in Task 2, and the axiomatic argument (from the deductive proof scheme) in Task 3 had the lowest selection frequencies. It can be said that while the most selected arguments for the validity were inductive in each task, the least selected argument changed from task to task; it did not matter whether the argument was from external, empirical, or deductive proof schemes.

4.1.1.3 Teacher Expectancy

The students were asked to score the arguments based on how many points the teacher would give to these arguments to understand the students' expectations of the teacher. Like convincingness, the scores were summed and divided by 20 to find the average. However, it should be noted that the sum was divided by 19 for Task 2 because one of the students did not give any score.

Table 4.5 The Average of The Scores for The Teacher Expectancy in Task 1

	Axiomatic Argument	Transformational Argument	Inductive Argument	Symbolic Argument
Averages of The Scores	7.90	5.20	7.35	5.55

The results from Task 1 are presented in Table 4.5. Based on the table, the highest average belonged to the axiomatic argument. The average of the inductive argument was close to the axiomatic argument. In other words, the students thought their teacher would give high points to axiomatic and inductive arguments. On the other hand, as it did in the convincingness, the transformational argument had the lowest average. Since the averages of transformational and symbolic arguments were close to each other, it can be said that the students thought the teacher would not prefer transformational and symbolic arguments as much as the others.

Table 4.6 The Average of The Scores for The Teacher Expectancy in Task 2

	Authoritarian Argument	Perceptual Argument	Inductive Argument	Transformational Argument	Axiomatic Argument
Averages of The Scores	3.74	6.16	7.53	5.84	7.58

Similarly, Table 4.6 was formed by the averages of the scores of the arguments in Task 2 based on how many points the teacher would give to these arguments. The results show that the students thought the teacher would give high scores to the axiomatic and inductive arguments. The averages of the perceptual and transformational arguments respectively followed them. Finally, based on the average, the teacher might not prefer the authoritarian argument from the students' perspective.

Table 4.7 The Average of The Scores for The Teacher Expectancy in Task 3

	Authoritarian Argument	Inductive Argument	Axiomatic Argument
Averages of The Scores	5.25	8.25	6.85

Table 4.7 was formed by the results of Task 3. The highest average belonged to the inductive argument. The average of the axiomatic argument followed it, and the average of the authoritarian argument was the lowest. So, it can be said that the students thought the teacher would give high points to the inductive argument while the authoritarian argument would be preferred less by the teacher, similar to Task 2.

To sum up, the students generally thought the teacher would prefer inductive and axiomatic arguments in each task. On the other hand, the averages showed that the teacher would not give high points to the symbolic, authoritarian, and transformational arguments as many as to the other.

To conclude the descriptive findings from the written questionnaire, the students considered the inductive and axiomatic arguments convincing and preferable by the teacher in each task. On the other hand, in each task, while more than half of the students found the inductive arguments as a valid way to show the truth of the statement, more than half of the students thought the axiomatic arguments were invalid. Still, it should be noted that there were close to 30% of the students considered the axiomatic arguments as valid, in each task.

4.1.2 The Qualitative Findings from the Written Questionnaire

The open-ended responses to the written questionnaire were investigated to explore students’ evaluation criteria. This title is related to the criteria used to evaluate an argument in terms of convincingness, validity, and teacher expectancy for each task. The frequencies of the use of the codes for convincingness, validity, and teacher expectancy were respectively presented.

4.1.2.1 Qualitative Findings from the Written Questionnaire Regarding the Convincingness

Table 4.8 is made based on the number of students who used the mentioned code to evaluate the convincingness of the arguments in each task. Twenty students evaluated the arguments with several criteria, and the most used criteria were reported in Table 4.8.

Table 4.8 The Frequencies of The Codes Used to Evaluate Convincingness

Code Names	**Task 1**	**Task 2**	**Task3**
Explanatory / Understanding	14	11	8
Detail	1	2	1
Simplicity	5	10	1
Practicality	3	2	5
Example	5	4	6
Visual	3	1	0
Operations	2	2	2
Algebra	1	0	1
Source of Explanation	0	7	6
Length	1	5	1
Logic	9	9	6
Liking	7	8	8
Correctness	5	2	2
Validity	3	4	4
Rule	1	0	0
Provability	1	2	1

Other	6	9	5
No Response	0	0	0

The results show that more than half of the students used explanatory/understanding criteria to evaluate the convincingness of the arguments in Task 1. As the second and third most emerged code, 45% of the students used logic criteria, and 35% of the students used liking criteria. In short, the students paid attention to whether the argument was understandable, logical, and likable in Task 1.

Like in Task 1, explanatory/understanding code was the most emerged criterion in Task 2. It was used by 55% of the students. Also, like in Task 1, logic and liking arguments were among the most used criteria in Task 2. On the other hand, unlike in the other task, the second most used criterion was simplicity in Task 2 (used by 50% of the students). In addition, as another frequently used criterion, the source of the explanation code emerged in 35% of the answers because there was an authoritarian argument in Task 2, unlike in Task 1.

In Task 3, there were no criteria used by more than half of the students. Liking and explanatory/understanding codes were the most used criteria, with 40%. As the second most emerged code, 30% of the students used examples and sources of explanation codes. It should be noted that the source of the explanation code was frequently used in Tasks 2 and 3 but not in Task 1 because the source of the explanation code emerged in only the authoritarian arguments.

In further investigation of codes, a minority of the answers that were coded with the source of explanation criteria mentioned that the statement was convincing if the authority said so. In terms of using the source of explanation code, 5 of 7 students in Task 2 and 4 of the 6 students in Task 3 were less convinced by the authoritarian arguments.

To sum up, the explanatory/understanding code was the most used criterion in each task. So, the students gave more attention to the argument’s being explanatory or understanding, no matter whether it was from the number or geometry learning area. Also, many students used the liking and logic codes to evaluate the convincingness of arguments in each task. On the other hand, in terms of differences between tasks, the simplicity code was the second most used criterion in only Task 2. In addition, the source of the explanation code took part among frequently emerged code in only Tasks 2 and 3 because Task 1 did not include an authoritarian argument.

4.1.2.2 Qualitative Findings from the Written Questionnaire Regarding the Validity

The codes listed with the number of students who used them to evaluate the validity of the arguments in each task are provided in Table 4.9. However, one student did not give a response to the validity question in Tasks 2 and 3, so the size of participants was 19 for these tasks.

Table 4.9 The Frequencies of The Codes Used to Evaluate Validity

Code Names	**Task 1**	**Task 2**	**Task3**
Explanatory / Understanding	8	4	5
Detail	0	1	0
Simplicity	3	1	0
Practicality	3	1	3
Example	8	3	5
Visual	2	2	1
Operations	0	0	3
Algebra	1	0	0
Source of Explanation	0	2	2
Length	0	1	2
Logic	1	4	3

Liking	5	4	2
Rule	1	1	0
Convincingness	2	1	1
Other	1	4	1
No Response	0	1	1

In Task 1, the explanatory/understanding and example codes were the most used criteria for validity evaluation. The second most emerged code was liking, but its percentage was still low: 25%. In addition, unlike convincingness, logic was one of the codes with the lowest frequency for validity. On the other hand, in Task 2, logic, liking, and explanatory/understanding codes were equally the most emerged codes. However, it should be noted that their percentages were only around 21.05%. Similarly, while examples and explanatory/understanding codes were the most used codes in Task 3, their percentages were around 26.32%. Furthermore, none of the codes was used by more than 50% of the students, so there was no dominant criterion to evaluate the validity.

Unlike convincingness, the example code emerged as many as the explanatory/ understanding code. It was the most used code in Tasks 1 and 3 and the second most used code in Task 2, even though its percentages were not so high. This was expected because the most selected argument in terms of validity was inductive. Many students thought it was a valid way to show that a statement is always true if the statement works in examples.

To sum up, there were no dominant criteria to evaluate the validity but explanatory/ understanding and example criteria were the most used criteria. Like convincingness, the students frequently selected the arguments they understood better as valid arguments, and they considered the use of examples as one of the ways to show validity.

4.1.2.3 Qualitative Findings from the Written Questionnaire Regarding the Teacher Expectancy

The students were asked to score the arguments based on how many points between 1 – 10 they thought their teacher would give to the argument. And they were asked for their reasoning for the scoring. Table 4.10 lists the frequencies of the criteria that they think their teacher would use to evaluate the arguments based on the number of students.

Table 4.10 The Frequencies of The Codes Used Regarding Teacher Expectancy

Code Names	Task 1	Task 2	Task3
Explanatory / Understanding	7	5	6
Detail	0	1	1
Simplicity	3	5	0
Practicality	3	2	3
Example	2	2	3
Visual	3	0	0
Operations	0	1	2
Algebra	1	1	0
Source of Explanation	0	5	4
Length	0	3	2
Logic	4	6	4
Liking	8	8	10
Teacher Method	2	1	0
Correctness	1	1	1
Validity	4	2	3
Scientific Way	0	1	0
Rule	1	0	1

Provability	0	1	0
Same With Me	3	4	0
Convincingness	5	3	4
Other	7	7	3
No Response	2	2	1

Based on the results, the liking code was the most used criterion in Task 1. Then, the explanatory/understanding code took the second line. As the third most emerged code, 25% of the students thought the teacher would consider the convincingness of the argument to score. In Task 2, the liking code was the most used criterion, and the logic code followed it. In addition, 25% of the students' answers were coded as the source of explanation and explanatory/understanding. In Task 3, half of the students' answers were coded as liking. Then, 30% of the students thought the explanatory/understanding would be the evaluation criteria of the teacher.

In Task 2, 2 of 5 students who used the source of explanation criterion thought that the teacher would give high points if the argument presented an explanation of an authority. Similarly, in Task 3, 3 of 4 students who used that criterion thought the authoritarian arguments would get high scores from the teacher. On the other hand, the other 3 of 5 students in Task 2, and 1 of 4 students in Task 3 thought that these arguments would not be so preferable to the teacher. In short, there were students who thought that the teacher would give high scores and low scores to explanations of an authority.

In terms of teacher expectancy, new codes emerged different from convincingness and validity. For instance, in particular to teacher expectancy, the "same with me" code emerged in Tasks 1 and 2. This code represents the students who directly said, "*The teacher would think like me*." In other words, some students thought the way of their evaluation could be the same as the teacher's.

As a new code, the teacher method code emerged, particularly in teacher expectancy. Some students thought the axiomatic arguments in Tasks 1 and 2 and the inductive argument in Task 2 were similar to the teacher's way of teaching; they considered these similarities might affect the teacher's scoring.

To sum up, some students thought the teacher's score would be affected by similarities of his teaching method, or the teacher's evaluation way would be similar to theirs. Also, while some students who used a source of explanation code considered the explanation of authority would receive high points, some of them disagreed. Furthermore, unlike convincingness and validity, the liking code was the most used criterion for teacher expectancy in each task. But still, the explanatory/understanding code was one of the most used criteria. On the other hand, it should be noted that there was not a dominant code since no code emerged in the answers of more than half of the students.

4.1.3 Findings Regarding the Student Follow-Up Interviews

Follow-up interviews were conducted to get a deeper understanding of how the presented criteria affected the proof evaluation of the students. This section will present the findings related to the follow-up student interviews, which included each participant's responses to written questionnaires. It should be noted that the most important themes in the interviews were reported, and the findings were presented for each student respectively.

4.1.3.1 Student 5

Table 4.11 shows the points of convincingness for the arguments and the criteria used by Student 5 based on the questionnaire. Her questionnaire showed that she was convinced by the inductive arguments in Tasks 1 and 3. In Task 2, she was convinced by the axiomatic argument. During the follow-up interview, why she made such scoring and how the criteria affected her scoring were investigated.

Table 4.11 Convincingness Evaluation of Student 5 In the Written Questionnaire

Types of the Arguments	Task 1	Task 2	Task 3
Authoritarian	-	1 Logic Length Exp./Understanding Source of Explanation	1 Practicality Source of Explanation Validity
Symbolic	4 Logic	-	-
Inductive	10 Exp./Understanding	4 Liking	10 Logic Liking Exp./Understanding
Perceptual	-	3 Liking	-
Transformational	1 Simplicity Exp./Understanding	2 Logic Length	-
Axiomatic	2 Exp./Understanding	10 Logic Exp./Understanding	5 Exp./Understanding

Firstly, it can be said that the source of the explanation criteria had an essential effect on her evaluation of convincingness. In the written questionnaire, she wrote, "*So illogical, short, nonexplanatory. I could get this answer by looking at the book*" in Task 2, and "*it was not his/her answer*" in Task 3. She did not prefer an authority's explanation. The interesting finding from the interview was that she also found the axiomatic argument in Task 1 as an authority's explanation. She said, "*It seemed as if she had read it from somewhere and pasted the same thing as if she did not know it either*." Thus, she focused not only on whether the argument claimed it was from an authority but also on whether the argument looked like an authority's explanation. In short words, not only what the argument claimed but also what the argument looked like might be essential in terms of her using the source of the explanation criteria.

There were example codes in the interview that did not exist in her written questionnaire. She said, "*She/he gave a lot of examples*" to the question "*What did* *you find explanatory*" for the inductive argument in Task 1. Moreover, for the inductive argument in Task 3, she said, "*I like examples because they help me understand*." In addition, she emphasized the examples make the argument convincing in Task 2. As a result, the use of examples makes the argument explanatory and convincing for Student 5.

Although the students claimed that the use of examples was the essential criterion, her concept of "example" was problematic. For instance, in Task 1, she said, "*There is only one example, and I think it is not explanatory enough*" for the axiomatic argument. She considered the algebraic operation in Task 1 as an example, and this criterion was repeated for the axiomatic argument in Task 3. Similarly, she said, "*Because there were a lot of examples, I liked it*" for the algebra part of the axiomatic argument in Task 2. When the student was asked to explain the axiomatic argument, she said, "*Wrote one by one,*" and added, "*Even if it is not an example, it is logical for me to write one by one*." Thus, the example might not be the actual criteria, but the appearance or the understandability of the argument might be. The appearances of the arguments that got 10 points from Student 5 are presented in Figure 4.1 to explore the situation.

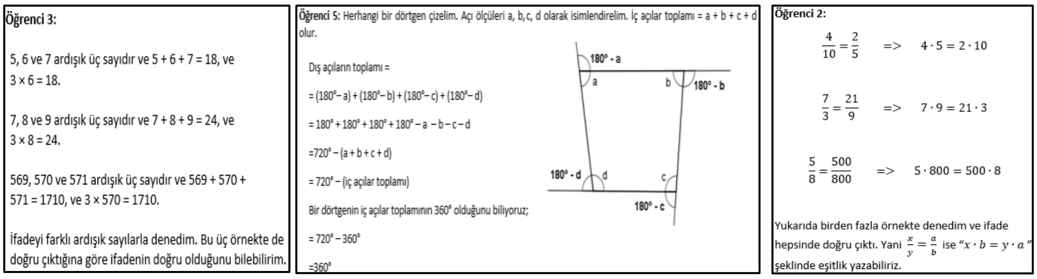


Figure 4.1 The Most Convincing Arguments for Student 5

While the student found the inductive arguments most convincing in Task 1 and 3, she found the axiomatic argument the most convincing in Task 2. As can be noticed in Figure 4.1, all arguments looked like a list form. In addition, she considered the writing one by one as an example, which makes the argument understandable and convincing for her.

In terms of the validity question, Table 4.12 presents the selected arguments and the codes for the reasoning in responses to the written questionnaire for Student 5. The arguments selected as valid showed parallelism to the most convincing arguments. The codes for validity were similar to the codes for convincing, but she was also assigned the code "example" in Task 1.

Table 4.12 Validity Evaluation of Student 5 In the Written Questionnaire

	Task 1	Task 2	Task 3
The Argument Type	Inductive	Axiomatic	Inductive
The Code for The Reasoning	Exp./Understanding Example	Exp./Understanding Logic	Exp./Understanding

She did not give an extra code for validity in the interview; she just confirmed the codes from the written questionnaire.

In terms of teacher expectancy, Table 4.13 shows the scores that the student thought the teacher would give and the codes for her reasonings. Based on the table, she thought the inductive arguments were preferable for the teacher in Tasks 1 and 3. This was a similar case with the convincingness score. However, in Task 2, the highest teacher expectancy score belonged to the perceptual argument, while the highest convincingness score belonged to the axiomatic argument.

Table 4.13 Teacher Expectancy Evaluation of Student 5 In the Written Questionnaire

Types of the Arguments	Task 1	Task 2	Task 3
Authoritarian	-	1 Liking Source of Explanation	2 Source of Explanation
Symbolic	3 Convincingness	-	-
Inductive	8 Exp./Understanding Example	2 Liking	10 Exp./Understanding Logic
Perceptual	-	7 Liking	-

		Length Simplicity	
Transformational	1 Simplicity Visual	1 Example	-
Axiomatic	5 Simplicity	5 Length	4 Exp./Understanding Length

The difference between her convincingness and teacher expectancy scores in Task 2 might be caused by the length of the argument. Firstly, it was asked why the axiomatic argument in Task 1 got 5 points while its convincingness score was 2. She said, "*..., he might like it because it was short and simple*." Similarly, she found the perceptual argument preferable to the teacher because she thought the teacher liked short and simple answers, while she criticized the axiomatic argument in terms of being too long for the teacher. Thus, the length might have made the perceptual argument preferable to the axiomatic argument in Task 2 for teacher expectancy although it was vice versa for convincingness.

The length code was frequently used for teacher expectancy; however, the meaning of the length code was not always consistently used. To evaluate the axiomatic argument in Task 3 regarding teacher expectancy, she said, "*It was short but not explanatory.*" It was asked what an argument needs to be explanatory, she responded, "*Not long, not short either. In a way that won't kill my time, but explanatory.*" In that case, the length might refer to the practicality and being explanatory.

Finally, Student 5 thought the teacher would not like explanations from an authority. She claimed that the teacher would give low points to the authoritarian arguments in Tasks 2 and 3 because of the direct quotation from the teacher or the book. Similar to her criteria for convincingness, she thought the teacher would consider the symbolic argument (which had quite a similar look to the axiomatic argument) in Task 1 as an explanation from an authority.

4.1.3.2 Student 9

Table 4.14 presents the points of convincingness for the arguments and the criteria used by Student 9. Based on the table, the student found the inductive arguments convincing in each task, like many of the class. Also, the symbolic argument in Task 1 and the axiomatic arguments in Tasks 2 and 3 were found convincing. In addition, the transformational argument in Task 1 and the authoritarian arguments in Tasks 2 and 3 were the least convincing arguments for her.

Table 4.14 Convincingness Evaluation of Student 9 In the Written Questionnaire

Types of the Arguments	Task 1	Task 2	Task 3
Authoritarian	-	4 Logic Provability	6 Practicality Source of Explanation Validity
Symbolic	8 Practicality	-	-
Inductive	7 Practicality Simplicity Validity	9 Logic Provability Practicality Validity	7 Provability Validity
Perceptual	-	3 Logic Exp./Understanding	-
Transformational	4 Logic Practicality Visual	10 Logic Provability Practicality	-
Axiomatic	5 Simplicity Validity	7 Logic Practicality Validity	8 Logic Provability

Practicality was one of the essential criteria for Student 9. She criticized the arguments in terms of saving time. For instance, Student 9 criticized the visuals in the transformational argument in Task 1, she said "*We waste time by drawing these… but it is logical to understand better*." As another example, in Task 2, she considered cutting and pasting steps in the perceptual argument as wasting time. Moreover, she claimed that the axiomatic argument in Task 2 was logical "*because it is done more quickly on paper (and) because it is practical.*" Thus, practicality was an essential criterion to be convinced for her.

Student 9 was one of the students who used the provability code. This code was used for Tasks 2 and 3. For instance, in the questionnaire, for the authoritarian argument in Task 2, she wrote, "*Actually, it is logical but nothing in terms of provability*." When the meaning of the provability was asked, she answered, "*It did not give examples. Maybe it's wrong in the textbook. I mean, it does not, but... it could be.*" Thus, it can be said that the provability code referred the use of examples. However, it should be noted that the term provability was not related to validity because she claimed, "*There is a provability, but it may not always be true*" for the inductive argument in Task 3 (she said the prime numbers may not work). Therefore, the proof term in the provability code did not hold for "always work," it did for "examples where the statement works."

Another essential finding from the interview with Student 9 was that she found the inductive argument in Task 1 convincing despite the validity. The inductive argument checks the statement with several consecutive numbers, and the students claimed that the set of "1, 2, 3" does not hold for the statement. She said, "*In fact, the most convincing one was student 3 (the inductive argument), but 1, 2, 3 does not work*." On the other hand, she gave 7 points in terms of convincingness. Thus, even if the statement does not work in some examples, the argument might still be convincing.

Examples were elements increasing convincingness, but also, the use of numbers in the example may impact the student. For instance, she found the transformational argument more convincing than the inductive argument in Task 2; however, the responses of the student did not include any sentence referring to deductive reasoning in the transformational argument; and she said, "*Here, all of them are 90°. That's why… because it's easier to sum.*" Therefore, the way of using examples might affect her being convinced of the arguments.

Student 9 might misinterpret the algebraic expressions. Like Student 5, Student 9 considered the algebraic operations of the axiomatic argument in Task 3 as “an example.” She claimed that that axiomatic argument has a low possibility for provability because “*It just gave an example. If we give a few more examples, it may be (more convincing).*” On the other hand, this situation was not observed in other tasks. In Task 2, she said that she found the axiomatic argument practical, but its algebraic expressions were confusing; however, she didn’t claim these algebraic expressions were examples. In Task 1, Student 9 found the symbolic argument to be more convincing than the axiomatic argument because she thought the axiomatic argument used the distributive property, and the parentheses confused her mind. However, there was no use of distributive property in the axiomatic argument. So, the algebra code affected her being convinced, but the way of using algebra might have different effects on the convincingness from task to task.

Table 4.15 shows the arguments selected by Student 9 as valid arguments and her criteria for making this selection. Based on the table, the inductive arguments were valid in each task. Also, transformational arguments in Tasks 1 and 2, and the axiomatic argument in Task 2 were accepted as valid. In addition, in terms of validity, the example code emerged in Tasks 1 and 3 and the rule code emerged in Task 2.

Table 4.15 Validity Evaluation of Student 9 In the Written Questionnaire

	Task 1	Task 2	Task 3
The Argument Type	Inductive Transformational	Inductive Transformational Axiomatic	Inductive
The Code for The Reasoning	Example	Rule	Example

As a first impression from the table, the validity might not have been related to the convincingness for Student 9. For instance, while three arguments with high convincingness points in Task 2 (see Table 4.14) were selected as valid, the arguments with the highest convincingness points in Tasks 1 and 3 were not selected. In addition, in Task 1, she found the transformational the least convincing (see Table

4.14) but valid based on Table 4.15. So, there were differences between the convincingness and validity evaluation of Student 9 in Tasks 1 and 3.

As the criteria of validity evaluation, in the written questionnaire, the student wrote that she selected these arguments "*because it turns out true every time it's repeated*" in Task 1, and "*because it demonstrated with multiple examples*" in Task 3. On the other hand, in Task 2, she wrote, "*because they all have certain rules*." The meaning of the rule was asked during the interview, and she said, "*They look very similar to each other, I mean almost… Based on a rule… Because the sum of the exterior angles is 360°.*" When it was asked whether that rule exists in the authoritarian and perceptual argument, she said, "*But this does not prove*." Unfortunately, she did not give more explanation about the "rule." However, it was noticed that she used the "proof" term to refer to examples in the convincingness evaluation.

Similar to the convincingness evaluation, the student did not see examples where the statement doesn't work as an obstacle to validity. In Task 1, she made a calculation mistake and thought 1, 2, and 3 (as consecutive numbers) didn't hold for the statement, but still, she accepted the inductive argument as valid. In Task 3, she had doubts about whether the statement would work with the prime numbers, but she still selected the inductive argument. Thus, for Student 9, a statement could be valid despite the existence of contrary examples.

In terms of teacher expectancy, Table 4.16 presents the scores that the student thought the teacher would give. Student 9 thought the teacher would give high points to all arguments in Task 1. In Task 2, the transformational and inductive arguments got the highest scores, while the authoritarian argument got the lowest score. Similarly, the student thought the teacher would lower points to the authoritarian argument in Task 3. The inductive and axiomatic arguments got high scores in Task 3.

Table 4.16 Teacher Expectancy Evaluation of Student 9 In the Written Questionnaire

Types of the Arguments	Task 1	Task 2	Task 3
Authoritarian	-	1 Logic	4 Logic
Symbolic	10 Practicality	-	-
Inductive	8 Practicality Validity	10 Logic Provability	8 Practicality
Perceptual	-	7 Exp./Understanding	-
Transformational	10 Exp./Understanding	10 Logic	-
Axiomatic	10 Liking	8 Logic Simplicity	10 Logic

As seen in Table 4.16, the student thought the teacher would give a low point to the authoritarian arguments in Tasks 2 and 3 because she found these arguments illogical. Moreover, she claimed that the authoritarian argument in Task 2 was illogical for a mathematician "*Because there is no operation*." She added, "*…it must be provable*" to be logical. And "*It could give examples*" to be provable. To sum up, the authoritarian arguments did not convince the student, she thought these arguments were less preferable to the teacher.

In Task 3, while the axiomatic argument received 10 points, the inductive argument received 8 points because she thought the several examples in the inductive argument were impractical. On the other hand, the student said that she was less convinced by the axiomatic argument because she thought it was a single example. When the researcher asked whether one single example would be enough for the teacher, she said, "*No, but he can experiment himself*." Thus, if the teacher could test the argument with other examples, this might increase the point that the argument would get, in her opinion. Even though one argument might not be enough for the teacher, it could get higher points because of the testability of the argument for her.

From Student 9's perspective, the teacher would give high scores to complex arguments. She wrote "*it is the best way for a mathematician*" for the axiomatic argument and "*it is the easiest way for a mathematician*" for the symbolic argument in Task 1. When the meanings of the best and easiest ways were asked, she said, "*Because our teacher loves complexity… I guess…*" She found the axiomatic argument complex and the transformational argument impractical in terms of convincingness, but she thought the teacher would like this complexity.

In Student 9's opinion, the teacher could give high points to understandable arguments. In Task 1, she added, *"… my teacher likes hard stuff, but he would show students this tactic (inductive argument) … He would also show Student 2 (transformational argument) because (he would think) they (students) understand more with the visual…*" Similarly, in Task 2, the perceptual argument got a high score in terms of teacher expectancy, unlike in convincingness; and she said *"(The teacher) may think the students can understand better*" as the reason of the high score. As a result, she thought the teacher might give high points to the arguments that he could show the students to understand better.

4.1.3.3 Student 12

Table 4.17 was formed by the scores and criteria in Student 12's written questionnaire in terms of convincingness. The inductive arguments were the most convincing in each task, like most of the students. In addition, the axiomatic argument in Task 3 and the transformational argument in Tasks 1 and 2 followed them at the second line. On the other hand, the arguments from the external conviction proof schemes were found to be the least convincing by Student 12. Also, it can be seen that the example code frequently emerged based on the table.

Table 4.17 Convincingness Evaluation of Student 12 In the Written Questionnaire

Types of the Arguments	Task 1	Task 2	Task 3
Authoritarian	-	1	1

		Source of Explanation	Source of Explanation
Symbolic	2 Correctness	-	-
Inductive	9 Example Exp./Understanding	7 Example Exp./Understanding	9 Example Validity
Perceptual	-	3 Example Provability	-
Transformational	8 Logic Provability Visual	5 Simplicity	-
Axiomatic	5 Algebra	4 Other	8 Example

Student 12 was mostly convinced by the inductive argument in each task. He mentioned that seeing the statement in numbers made the argument explanatory and provable. When the meaning of provability was asked in Task 3, he said, "*(It) gave numbers. It gave more than one number*." So, like Student 9, he uses the provability code to refer to examples.

As another argument from empirical proof schemes, Student 12 found the perceptual argument less convincing because he said, "*I didn't see how to cut, how to put it together. I couldn't understand because I couldn't imagine it in my mind.*" He mentioned that it would be more convincing if he saw someone cut and paste the paper in the real world.

Student 12 criticized the deductive proof schemes' arguments in terms of simplicity and practicality. For instance, in Task 2, he wrote that the transformational argument was confusing. He considered the transformational argument in Task 1 as explaining numbers with images. He claimed that explaining with numbers or visuals was the same, but said, "*We need to understand visuals… (but) we can check whether it is true or not by only adding and subtracting (with numbers)*." Thus, numbers might be more practical, in his opinion. In addition, he mentioned that we can check whether the axiomatic argument was true by giving numbers to the variables.

However, he also said, “*I saw that it was true, but we have to struggle with it. Instead of giving a few numbers to ‘n,’ we can give a direct number and see directly that it is correct.*” Therefore, it can be said that the student was not aware of the generalization power of the deductive arguments.

In Task 2, he criticized the “*We know the sum of interior angles of a quadrilateral is 360°*” expression which was included in the axiomatic argument. When the reasoning for the low score was asked, he said, “We *can't see where it comes from or how it's done,*” and emphasized the lack of explanation. In other words, his response could be coded as explanatory/understanding.

Finally, unlike Students 5 and 9, Student 12 noticed the incorrect algebraic expressions in the symbolic argument in Task 1 by giving several numbers to “n,” and found the argument unconvincing because of the incorrectness. In addition, he found the authoritarian arguments unconvincing in Tasks 2 and 3. He added, “*I have to test it. In other words, he has never tested it himself and accepted this thesis directly*” for Task 2, “*I need to see what our teacher said,*” for Task 3. Thus, it can be said that the arguments from the external conviction proof scheme did not convince Student 9.

In terms of validity, the selected arguments and the criteria in Student 12’s written questionnaire were presented in Table 4.18. Like convincingness, the inductive arguments were selected as valid in each task. Also, the transformational argument in Task 1 and the axiomatic argument in Task 3 were considered valid arguments. While the example codes emerged in Tasks 1 and 3, the other code was used in Task 2 because his response did not include any criteria.

Table 4.18 Validity Evaluation of Student 12 In the Written Questionnaire

	Task 1	Task 2	Task 3
The Argument Type	Transformational Inductive	Inductive	Inductive Axiomatic
The Code for The Reasoning	Visual Example	Other	Example

The student wrote, "*...the visuals and more than one example made me think it is always true*" in Task 1. Similarly, in Task 3, he mentioned several examples of inductive arguments as "*more than one proof.*" In addition, he considered the axiomatic argument in Task 3 the same as the inductive argument without numbers. He said, "*The things done here (the inductive argument) are explained with variables (the axiomatic argument), without numbers*." Because of that, he also considered the axiomatic argument as valid. During the interview, he mentioned the inductive argument as logical as the reason to select a valid argument, but there was no further comment.

In terms of teacher expectancy, the scores that Student 12 thought the teacher would give and the criteria are listed in Table 4.19. Based on the table, the highest points belonged to the axiomatic arguments in Tasks 1 and 3 and the perceptual argument in Task 2. The inductive arguments in Tasks 1 and 3 and the axiomatic argument in Task 2 followed them in the second line. The arguments from the external conviction proof scheme got the least points in each task.

Table 4.19 Teacher Expectancy Evaluation of Student 12 In the Written Questionnaire

Types of the Arguments	Task 1	Task 2	Task 3
Authoritarian	-	1 Source of Explanation	2 Liking
Symbolic	1 Other	-	-
Inductive	3 Simplicity	6 Liking	8 Example
Perceptual	-	8 Other	-
Transformational	5 Logic Visual	3 Liking	-
Axiomatic	8 Algebra Liking	7 Algebra Liking	10 Validity

The criteria used to evaluate convincingness were similar to the criteria to evaluate teacher expectancy of the arguments from the external conviction proof schemes. In Tasks 2 and 3, he thought the teacher would not like an explanation of authority. For instance, he said, "*(The teacher) would like (the student) to do it himself. He would like the students to make the argument to be accepted by giving several examples.*" in Task 3. Thus, the student thought the teacher would prefer testing with examples to give an explanation of authority. In addition, in terms of criticizing the symbolic argument in Task 1, he wrote, "*(The teacher) would prefer the objective one*," which was coded as the other. Based on the interview, by saying "objective one," the student meant the teacher would not prefer the argument with incorrect explanations.

It can be noticed that the scores of the axiomatic arguments were higher than the inductive arguments. The student thought the teacher liked the algebraic equations, and that is why the axiomatic arguments were preferable according to him. On the other hand, in Task 2, the perceptual argument had the highest score. As the reason, he said, "*I thought the teacher would explain it by cutting a paper since he would want us to see*." In other words, the teacher could test the argument by cutting a real paper. Thus, he thought the argument would get a high score because of the "testability."

4.1.3.4 Student 13

Table 4.20 presents scores and criteria to evaluate the convincingness of the arguments in Student 13's written questionnaire. In Task 1, he found the axiomatic arguments as the most convincing, the transformational argument as the second most convincing, and the symbolic argument as the least convincing. In Task 2, the authoritarian argument was the most convincing, the axiomatic argument was the second most convincing, and the perceptual argument was the least convincing. In task 3, the inductive argument was the most convincing, and the authoritarian argument was the least convincing. Thus, it can be said that there was inconsistency in terms of evaluating the most and least convincing arguments.

Table 4.20 Convincingness Evaluation of Student 13 In the Written Questionnaire

Types of the Arguments	Task 1	Task 2	Task 3
Authoritarian	-	10 Length	5 Exp./Understanding
Symbolic	1 Correctness	-	-
Inductive	5 Logic Exp./Understanding	5 Validity	10 Liking Logic Practicality
Perceptual	-	1 Logic	-
Transformational	7 Logic	3 Simplicity	-
Axiomatic	9 Logic Practicality	8 Length Liking	6 Simplicity

Some opinions of the student changed during the interviews, and it was observed that the student was convinced by the explanations of the authority. For instance, in Task 2, he found the authoritarian argument practical. While he mentioned that he did not understand completely in the written questionnaire, the student said that he would be convinced if the teacher said there was such a rule in Task 3. So, the student was convinced by the authoritarian arguments in both tasks; he just did not understand the text in Task 3.

The student found the inductive argument in Task 3 practical because it directly checked the statement with numbers. However, he did not understand the inductive argument in Task 1. He wrote, "*It did not say which number we would multiply around three numbers.*" in the written questionnaire. When he noticed that the statement said to multiply the middle number, he claimed that the inductive argument was more convincing than the axiomatic argument. On the other hand, he did not find the inductive argument as convincing as the axiomatic argument in Task 2, and he added, "*It could have been a different angle… will it always turn out to be true… it may not be true*." Therefore, while he found the inductive arguments in Tasks 1 and 3 convincing, he criticized the inductive argument in Task 2 in terms of its validity.

To evaluate the axiomatic argument, he claimed that algebra makes the argument practical in Task 1. In Task 2, he considered the axiomatic argument as the way the teacher would teach. But, when he compared the axiomatic argument and the authoritarian argument in Task 2, he said the authoritarian argument was more practical, but the axiomatic argument was more logical because it taught how to find external angles. In Task 3, the use of algebra in the axiomatic argument was confusing for Student 13. He said, "*It would be more logical if it showed by giving numbers*." As a result, evaluating axiomatic arguments could change from task to task.

In terms of validity, Table 4.21 was formed by the findings of Student 13's written questionnaire. The axiomatic argument in Task 1, the authoritarian argument in Task 2, and the inductive argument in Task 3 were selected as valid arguments. While the argument types were changing from task to task, the practicality code emerged in each task, and the logic code emerged in Tasks 2 and 3.

Table 4.21 Validity Evaluation of Student 13 In the Written Questionnaire

	Task 1	Task 2	Task 3
The Argument Type	Axiomatic	Authoritarian	Inductive
The Code for The Reasoning	Practicality	Logic Practicality	Logic Practicality

As mentioned, practicality was one of the key criteria for validity from his perspective. For instance, at the beginning, the student did not understand the inductive argument in Task 1. After he understood, he said that the inductive argument could be counted as valid because it might provide practicality and easiness with long and large numbers. As another example, in Task 2, he said, "*It would save time in exams*" for the authoritarian argument in Task 2. In short, Student 13 gave importance to the practicality while evaluating the validity of the arguments.

Table 4.22 is prepared to express the findings regarding teacher expectancy of Student 13's written questionnaire. The results showed that the axiomatic argument in Task 1, both the authoritarian and axiomatic arguments in Task 2, and the inductive argument in Task 3 got the highest scores. On the other hand, the inductive argument in Task 1, the perceptual argument in Task 2, and the authoritarian argument in Task 3 got the lowest scores based on the result from the written questionnaire of Student 13.

Table 4.22 Teacher Expectancy Evaluation of Student 13 In the Written Questionnaire

Types of the Arguments	Task 1	Task 2	Task 3
Authoritarian	-	9 Practicality	1 Liking
Symbolic	7 Logic	-	-
Inductive	4 Same With Me	4 Teacher Method	10 Same With Me
Perceptual	-	1 Logic	-
Transformational	5 Liking	2 Liking Teacher Method	-
Axiomatic	10 Teacher Method Liking	9 Liking Logic Teacher Method	7 Liking Practicality

The teacher method was one of the essential criteria for Student 13. For instance, the student said the authoritarian argument gets high points because "*(The teacher) would say that open, read the book or repeat the textbook, then you will remember*." So, to evaluate the teacher expectancy, the student focused on what the teacher would say, and he thought the teacher would prefer ways to help them remember. For example, in Task 2, he thought the teacher would not prefer the perceptual argument since the students might easily forget that way. In addition, the teacher would select practical ways to teach, in his opinion. He thought the teacher would prefer the axiomatic argument in Task 1 to teach since the teacher thinks the students learn quickly in that way. In short, the student considered the methods the teacher would use to teach, and in his opinion, the teacher focused on memorability and practicality of the arguments.

Student 13 considered what the teacher likes to evaluate arguments regarding teacher expectancy points. He thought the teacher liked algebra. The student mentioned the teacher would like the axiomatic arguments and give high scores. Besides, although the student was aware of incorrect algebraic expressions in the symbolic argument in Task 1, he thought the teacher would still give a high point because the teacher liked algebraic expressions, and the teacher might find it logical. Thus, algebra was an essential criterion for teacher expectancy according to him.

4.1.3.5 Student 15

Table 4.23 presents the findings regarding convincingness in Student 15's written questionnaire. Based on the results, the student found the authoritarian arguments in Tasks 2 and 3 most convincing. In addition, she found the axiomatic argument in Task 1 the most convincing but the axiomatic arguments in Tasks 2 and 3 the least convincing. Finally, she considered the symbolic argument in Task 1 the least convincing because of its incorrectness.

Table 4.23 Convincingness Evaluation of Student 15 In the Written Questionnaire

Types of the Arguments	Task 1	Task 2	Task 3
Authoritarian	-	10 Source of Explanation	10 Source of Explanation
Symbolic	1 Correctness	-	-
Inductive	4 Validity Example	6 Validity	5 Validity
Perceptual	-	6 Validity Example	-
Transformational	2 Logic	7 Logic	-

	Exp./Understanding		
Axiomatic	10 Rule Validity	3 Other	2 Logic Validity

During the interview, the rule code frequently emerged. She said, "*There are rules in mathematics*," and she claimed that she made scoring based on these rules. At the beginning, she mentioned the algebraic expressions as rules in Task 1. She said, "*There is an equation there, and no matter what number you put in this equation, it works*." Similarly, the transformational argument in Task 2 emphasized that it won't change even if the angles change, and the students mentioned that that argument included a rule. On the other hand, in Tasks 2 and 3, she called the authoritarian arguments as rules. She considered the facts that were taught in the school as rules, and these arguments were quite convincing to her. In short, from her perspective, the rules in mathematics were the accepted facts and structures in the school.

Although the axiomatic arguments could be a rule for the student, the essential criterion was understanding, not just deductive reasoning. For instance, she did not find the axiomatic argument in Task 2 convincing because she thought there was a need for more verbal explanation. Similarly, she did not consider the transformational argument in Task 1 and the axiomatic argument in Task 3 because she did not understand and found it illogical. She said that it would be more convincing if the axiomatic argument in Task 3 included more verbal explanation or gave numbers instead of variables.

Another essential criterion was about accepted statements. This criterion emerged in Task 2. The transformational and axiomatic argument used the statement "the sum of the interior angles of a quadrilateral is 360°" to prove "the sum of the exterior angles of a quadrilateral is 360°." The student criticized adding another fact that was not given, and she questioned why we did not know the sum of the exterior angles if we knew the sum of interior angles. Therefore, giving extra accepted statements played a role in reducing the convincingness points from her perspective.

Unlike the other students who were interviewed, Student 15 was aware of the limitations of the example use. For instance, she said that the inductive argument got 4 points in Task 1 "*Because there are only 3 attempts. This could be a coincidence. There must be a certain rule*." Similarly, she emphasized that the perceptual and the inductive arguments in Task 2 and the inductive argument in Task 3 were not sufficient to be convincing because she thought the statements could not be explained with limited examples.

In terms of validity, Table 4.24 was prepared. The axiomatic argument in Task 1 and the authoritarian arguments in Tasks 2 and 3 were selected as valid arguments. The rule and the source of explanation codes emerged as criteria based on the responses in the written questionnaire.

Table 4.24 Validity Evaluation of Student 15 In the Written Questionnaire

	Task 1	Task 2	Task 3
The Argument Type	Axiomatic	Authoritarian	Authoritarian
The Code for The Reasoning	Rule	Source of Explanation	Source of Explanation

As can be noticed, the validity criteria were the same as the convincingness criteria (see Table 4.23). The student selected the valid arguments based on the rule. In the convincingness part, she mentioned that both the axiomatic argument and the authoritarian arguments included a rule. And based on this, the valid arguments and the most convincing arguments became the same.

Finally, Table 4.25 is formed by the teacher expectancy scores and criteria in Student 15's written questionnaire. The table shows that the axiomatic argument in Task 1 and the authoritarian arguments in Tasks 2 and 3 got the highest scores. On the other hand, the symbolic argument in Task 1 and the axiomatic arguments in Tasks 2 and 3 got the lowest scores in terms of teacher expectancy.

Table 4.25 Teacher Expectancy Evaluation of Student 15 In the Written Questionnaire

Types of the Arguments	Task 1	Task 2	Task 3
Authoritarian	-	10 Source of Explanation	10 Source of Explanation Rule
Symbolic	1 Validity	-	-
Inductive	6 Validity Example	7 Other	6 Validity
Perceptual	-	7 Validity Example	-
Transformational	3 Exp./Understanding Rule	6 Logic	-
Axiomatic	10 Logic	3 Other	3 Logic

The ranks of the arguments from the highest to lowest teacher expectancy score were the same with the ranks of the arguments from the most convincing to the least convincing. In addition, both the convincingness and teacher expectancy criteria were quite similar (see Table 4.23).

Based on the interview, she thought the teacher would criticize the examples in terms of validity as she did. In addition, she thought the teacher would give high points to the arguments that fit the rule concept of the student. In addition, the understandability of the arguments and use of accepted statements could affect the teachers' scoring, in her opinion. To sum up, there was a similarity between her evaluations of the convincingness of the arguments and the ones she thought the teacher would prefer.

4.1.4 The Summary of The Findings from the Students

The first sub-research question aimed to determine the criteria that the students use to evaluate an argument in terms of convincingness. The written questionnaires showed that the students generally found inductive arguments in each task convincing and the axiomatic arguments second convincing. On the other hand, the transformational argument in Task 1 and the authoritarian arguments in Tasks 2 and 3 were found to be the least convincing arguments. In addition, the explanatory/understanding, liking, and logic criteria frequently emerged in the written responses for evaluating convincingness, and the most used one was the explanatory/understanding code. On the other hand, what makes the argument explanatory or understandable was unclear. Using this code might not refer to understanding why the argument is true; instead, it might be understanding how it is true. For instance, students 5 and 12 mentioned that the examples make the argument explanatory. The students might evaluate the convincingness of the arguments based on whether the argument can show how the statement is true. They might look for the reflection of the statements on examples.

During the interviews, one of the students (Student 15) mentioned that giving examples satisfying the statement could not show the statement is always true and emphasized the validity of the argument as another criterion for convincingness. For instance, she chose the axiomatic argument in Task 1 as the most convincing argument because it included an equation that works for any number attained and criticized the inductive argument in terms of the possibility of coincidence.

In terms of using variables, the students had different views. For example, while Student 15 considered the axiomatic argument in Task 1 convincing, she did not find other axiomatic arguments convincing because she did not understand the arguments. Student 12 preferred numeric examples instead of using variables because he thought assigning numbers to variables was a struggle for him. He was not aware of the generalizability power of algebra. As another example, student 9 considered the axiomatic argument as a single example. In addition, while some students noticed the incorrect algebraic expression by assigning numbers, the number of these students was limited. Thus, the students had different perspectives on using variables to evaluate the arguments.

During the interviews, the practicality code also emerged. For example, the use of algebra was evaluated from different perspectives. While Student 13 thought algebra made arguments practical, Student 12 thought he should assign numbers to variables, and this was impractical. In addition, the visuals in the transformational argument in Task 1 were criticized in terms of practicality. Student 9 considered drawing visuals to be a waste of time, and Student 12 mentioned the extra work to understand the visuals. Thus, practicality was also one of the essential criteria regarding the interviews.

Besides emerged codes, Student 5's answers created a suspicion about whether she evaluated the arguments based on their outlooks. While she was convinced by inductive arguments in Tasks 1 and 3, she said the one-by-one explanation in the axiomatic argument in Task 2 was convincing, and she thought each step was an example. Moreover, each convincing argument was in list form. Thus, the appearance of the argument or the understandability of the argument, which was facilitated by the list form, might be another criterion for evaluation, instead of using examples.

Besides the explanatory/understanding, liking, and logic criteria, the source of explanation criterion was another frequently emerged code in Tasks 2 and 3. This code refers to whether the explanation in the argument is stated by an authority or not. The averages of convincingness of authoritarian arguments were the lowest ones, so it can be said that the students did not generally find explanations from authority as convincing as other arguments. Similarly, during the interviews, some students, such as Students 9 and 12, thought these arguments needed effort and extra work instead of giving an explanation from authority. On the other hand, some students, such as Students 13 and 15, thought it must be true if this information was in the book and taught to the students. Therefore, there were students who considered the authoritarian arguments convincing or unconvincing in the class.

As the second sub-research question, exploring the criteria the students use to evaluate an argument in terms of validity was the aim. The most frequently chosen valid argument belonged to the inductive arguments in each task. The symbolic argument in Task 1, the perceptual argument in Task 2, and the axiomatic argument in Task 3 had the lowest selection frequencies. Although there was no dominantly emerged code, explanatory/ understanding and liking codes frequently emerged as criteria like they were in the convincingness. In addition, unlike convincingness, the example code emerged as many as the explanatory / understanding code. Similar to the evaluating convincingness of the arguments, the students considered using examples as a way to show that the statement was always valid. Moreover, two students used provability code in their written answers: Students 9 and 12. During the interviews, both mentioned that this code stood for showing example cases for the statements.

Although no statements required a counterexample, Student 9 mentioned that the consecutive numbers “1, 2, and 3” do not hold for the statement in Task 1 because she made a calculation mistake. On the other hand, despite thinking there is a case where the statement does not work, she still accepted the argument as valid. So, in her opinion, a counterexample was not an obstacle to the validity.

The third sub-research question was to explore the criteria the students expect their teacher to use for scoring an argument. Based on the averages, the students thought the teacher would give the highest scores to the axiomatic arguments in Tasks 1 and 2 and the inductive argument in Task 3. They thought the second-highest scores would belong to the inductive arguments in Tasks 1 and 2 and the axiomatic arguments in Task 3. Thus, the students generally thought the teacher would prefer inductive and axiomatic arguments in each task. Like it was in the validity evaluation, there were no dominant emerged codes to evaluate teacher expectancy. The explanatory/ understanding and liking codes were frequently used criteria.

During the interviews, the comments of the students regarding the teacher expectancy were generally similar to the convincingness. For example, the students who considered the authoritarian arguments convincing thought the teacher would give high scores to these arguments, or vice versa. Similarly, the students who thought examples made the argument explanatory and convincing thought the teacher would give high scores to inductive arguments. On the other hand, there were also changes. For instance, the use of algebra affected the teacher expectancy evaluation. Student 12 thought the teacher would give higher scores to axiomatic arguments than he gave for their convincingness. Moreover, while Student 13 gave a low score to the symbolic argument in Task 1 because of incorrect expression, he thought the teacher would give a higher score to the argument, and he thought he might not understand the argument.

Another interesting criterion for teacher expectancy evaluation was practicality versus complexity. While Students 5 and 13 thought the teacher would prefer short, simple, and practical arguments, Student 9 thought the teacher would give higher scores to complex arguments. For instance, although Student 9 mentioned that the visuals in the transformational argument in Task 1 were impractical and found the axiomatic argument in Task 1 complex, she thought the teacher would like complexity. Despite this dissensus, it was mentioned on both sides that the teacher would prefer an argument that the students can easily understand.

4.2 The Findings Regarding the Interview with the Teacher

The same arguments in the written questionnaire were presented to the teacher, and he was asked to evaluate and give points to the arguments in terms of convincingness and select valid arguments by providing reasons. Also, he was asked to score the arguments, thinking as if they were presented by his students. In addition, in each task, the ranks for the average of the scores given by the students to convincingness were shared with the teacher, and it was asked why such rank might have been observed. The same process was repeated regarding the validity and the scoring of the arguments like they were responses of his students. The results were presented under the titles: convincingness, validity, and scoring of the arguments like they were responses of his students.

4.2.1 Convincingness

Based on the findings of the interview, Table 4.26 was formed by how many points the teacher gave to the arguments and which criteria were used in terms of convincingness. The results show that the axiomatic arguments were the most convincing in each task. The perceptual argument in Task 2 and the inductive arguments in Tasks 1 and 3 were the second most convincing. On the other hand, the symbolic argument in Task 1 and the authoritarian arguments in Tasks 2 and 3 were the least convincing arguments. Also, based on the scores, it can be said that all arguments except the arguments from the external conviction proof schemes convinced the teacher.

Table 4.26 The Teacher's Evaluation for The Convincingness of The Arguments

Types of the Arguments	Task 1	Task 2	Task 3
Authoritarian	-	2 Source of Explanation	2 Source of Explanation
Symbolic	2 Correctness Exp./Understanding	-	-
Inductive	8 Example Logic	6 Example	9 Example Algebra Validity
Perceptual	-	8 Validity Visual	-
Transformational	8 Example	7 Example	-
Axiomatic	10	10	10

	Algebra Validity	Algebra Validity	Algebra Validity Practicality Rule

In Task 1, the teacher claimed that the axiomatic argument was the most convincing because it makes a generalization. "*You always see the result come out (true)*," he said, and he added, "*It is nice to use algebra there for me; it makes generalizations*." Thus, he gave a full point to the axiomatic argument because of algebra use and its help in making generalizations.

The teacher was aware of the limitation of the inductive argument in Task 1 and said, "*It gave three examples. Maybe it won't come out true somewhere, ... Well, that's also a logical solution but…*" Despite the limitation of the inductive argument in generalizing, he thought the argument was convincing and it could prove and show the statement was true. When the meaning of the proof term he used was asked, he said that while it was using a general expression such as "n" which represents any number for the axiomatic argument, it was proving by taking an "arbitrary number" for the inductive argument. Thus, from the teacher's perspective, arbitrarily selected examples could also be ways to convince that the argument was true.

He gave the same points to the inductive and transformational arguments in Task 1 because he considered them similar arguments. He said, "*I would say this (the inductive) just used numbers, this (the transformational) just used points... I mean, they are the same…*" Thus, it can be said that the teacher was not aware of the structure of the transformational argument and accepted it as a single example.

In the symbolic argument in Task 1, the teacher noticed that the algebraic expression "$n + 1 + 2$" did not represent the sum of the consecutive whole numbers, and "$n + 3$" was not 3 times "$+1$." On the other hand, he could not claim it was incorrect; he just said that he could not understand, "*I must be missing something.*" After a while, he admitted, "*I thought it was true because you prepared it*." Then, he changed the score from 5 to 2. Therefore, although there was no authoritarian argument in Task

1, he considered the researcher as an authority and had doubts about the incorrectness of the argument.

In Task 2, the teacher criticized the authoritarian argument because it does not explain "why." He said that we cannot understand why just because it is written, and it could be written incorrectly. Thus, the authoritarian argument was not convincing for the teacher.

In terms of the most convincing argument in Task 2, the teacher selected the axiomatic argument because "*It is certain… using algebra, it will always come out true (for) whatever the quadrilateral it draws.*" Similarly, the teacher thought the perceptual argument in Task 2 showed that it will always be 360°. He said, "*It shows with the visual… draw any shape with 4 sides, cut them, and put them together, it will always get 360°.*" Thus, the perceptual argument was the second most convincing for the teacher.

The teacher considered the inductive and perceptual arguments as similar arguments and said, "*This (the perceptual) showed with visual here, this (the inductive) will show with numbers…*". In addition, unlike many students, the teacher emphasized the similarity of the inductive arguments in Tasks 1 and 2. And, because of the examples with several quadrilaterals, he found the inductive argument convincing.

Unlike Task 1, the teacher noticed that the example in the transformational argument in Task 2 could generate another example. By using the transformational argument, he gave another example: "*It tries to reach: If we change one of the angles, will the sum of the exterior angles change?... If I increase it by 20°, I have to decrease it by 20° to balance it out*" (it was 10° in the argument). He thought the transformational and inductive arguments were similar, but he likes the transformational argument more because "*It is like showing with one number here (the inductive), but here (the transformational) you can do things by changing it differently*." Thus, in Task 2, the teacher was aware of the transformability of the example in the transformational argument.

In Task 3, the teacher found the axiomatic argument the most convincing because "*It put a general rule… by using algebra*." Since he described the generalization as "coming out true always" at Task 1, these criteria were coded as both algebra and validity.

The teacher found the inductive argument in Task 3 the second most convincing at the beginning because it showed examples, like in Tasks 1 and 3. Then, when he compared the inductive arguments in Tasks 1 and 3, he thought the algebraic expression stated at the end of the argument in Task 3 was a general expression. On the other hand, although both inductive and axiomatic arguments had a general statement at the end, the teacher thought the axiomatic argument was more practical because it directly reached the generalization instead of struggling with some examples and then coming up with the general rule.

Like in Task 2, the teacher found the authoritarian argument in Task 3 the least convincing because the teacher said that these arguments looked like knowing because they heard from somewhere. And, he said such information which was heard from somewhere can be wrong.

To sum up, the teacher was convinced by the axiomatic arguments in each task because they used algebra to reach a generalization. He was aware that the examples might not show the statement coming out always true, but still, he found these arguments convincing because of selecting arbitrary example cases. Finally, the teacher was not convinced by the arguments from the external conviction proof schemes because the symbolic argument included incorrect statements, and the authoritarian arguments had the possibility of including incorrect statements.

4.2.1.1 Expectancies from the students

The teacher was asked what he thought would convince the students, and then findings regarding the arguments students were convinced were shared by the teacher. In Task 1, the teacher guessed which arguments would convince his students. He thought the students would find the axiomatic argument the most convincing, then the inductive argument would be in the second line, and the transformational argument would be in the third line because he thought the students would understand easier with the inductive argument, and the transformational argument would make the students struggle if they chose bigger numbers.

After the teacher shared his expectancies in Task 1, the rank of the convincing points was shared as "the inductive > the axiomatic > the symbolic > the transformational." He said, *"(It might have occurred) Because they better understand with numbers…"* and thought making a general rule in the axiomatic could be hard for the students. He had no idea why the students preferred the symbolic to the transformational argument at the beginning; then, he thought the reason might be the hardness to show bigger numbers with drawing points.

In Task 2, in terms of being convinced, his expectancies from the students were like "the axiomatic > the perceptual > the inductive > the transformational > the authoritarian." It was similar to the rank for his convincingness concept, but he thought the students would understand the inductive argument easier than the transformational argument. In addition, although the authoritarian argument was at the end of the rank, the teacher mentioned this rank might change based on the situation. For instance, the teacher thought an average student in his class might be convinced by the authoritarian argument; however, if the student is interested in the lesson, s/he might question whether the argument is true and not convinced by the authoritarian argument. In other words, he thought the students without interest might be more convinced than the students with interest.

After the teacher shared his expectancies for the students in Task 2, the findings regarding the rank of the convincing points were shared as "the inductive ≈ the axiomatic > the transformational > the perceptual > the authoritarian." He thought his way of teaching might affect the students to be convinced by inductive and axiomatic arguments. The teacher mentioned that he starts with algebraic expressions and then gives some examples with numbers in the lessons. He said that the students might find the inductive arguments easier to understand. He did not have a certain idea about why the perceptual argument was not convincing for the students. He just said the students might have not noticed that the argument included a generalization.

In Task 3, the teacher considered the students could be most convinced by the inductive argument and the least convinced by the authoritarian argument. He thought the students would understand better with examples, and so the students would be more convinced by the inductive arguments. His expectation was consistent with the rank regarding the average scores of the convincingness. However, he mentioned that he got this idea because of seeing ranks in Task 1 and 2. In his opinion, he would think the axiomatic argument would be more convincing because it did not need an example.

4.2.2 Validity

Table 4.27 presents the arguments that were selected as valid by the teacher and the criteria for this selection. Based on the table, the teacher accepted the axiomatic arguments in each task as valid arguments.

Table 4.27 The Teacher's Evaluation for The Validity of The Arguments

	Task 1	Task 2	Task 3
The Argument Type	Axiomatic	Axiomatic Perceptual	Axiomatic Inductive
The Code for The Reasoning	Other	Algebra	Algebra

In Task 1, the teacher selected the axiomatic argument as valid but did not mention any criteria. On the other hand, regarding the convincingness, he had already said that the axiomatic argument included a general rule by using algebra, which shows the statement always works. At the validity part, he said that the inductive argument needed to try all numbers to be valid even though it was convincing. And, since it is not possible to try all numbers, we cannot know if there is any number that does not hold for the statement.

In Task 2, the teacher selected the axiomatic arguments because of the use of algebra, and he said, "*Since this is my favorite solution (the axiomatic argument), I always compare (other arguments with this one)*." For instance, he considered the perceptual argument valid because it included angle symbols instead of any numbers. He said that the size of the angles could be increased or decreased in the perceptual argument; it will always be 360°. On the other hand, although he noticed that the example in the transformational argument could generate other examples, he did not consider the argument valid because he said, "*Since it would take too long to show by using all numbers… it is limited*." Therefore, the teacher considered the transformational argument invalid because he thought there was a need to check all numbers. On the other hand, he accepted the perceptual argument as a valid argument because it did not include any number like the axiomatic argument, and he thought this situation ensured generalization.

In Task 3, the teacher directly selected the axiomatic argument, then he added the inductive argument among the valid arguments because *"…it used an algebraic expression at the end, it created something like a formula again… it reached a generalization…*" he considered the use of algebra at the end as reaching a generalization.

To sum up, the teacher considered the arguments using algebra to make generalizations as valid arguments. The axiomatic arguments were selected for each task. The teacher also selected the perceptual argument in Task 2 because it was free from the numbers and selected the inductive argument in Task 3 because it included algebraic expression at the end of the argument.

4.2.2.1 Expectancies from the students

The teacher was asked to guess the rank of the arguments regarding the frequencies of the selection by the students. In Task 1, he thought most students might have selected the inductive argument because it was easy to understand like they did for the convincingness. The teacher also said, "*I think they did not doubt whether it would work if we did (try) it for all numbers*." Then, the teacher thought the frequency respectively continued like the axiomatic, the transformational, and the symbolic arguments. The teacher was informed that the frequency for Task 1 was the same as his expectancy. The teacher said he usually showed the general rule, where it comes from, and what the formula might be, and gave examples with numbers to make it understandable; and he thought his way of teaching might have influenced the students.

The teacher's expectation about the frequencies for Task 2 was like "the axiomatic > the inductive > the perceptual > the transformational > the authoritarian." The teacher put the inductive arguments in the second line while he found the perceptual argument valid. As the reason why he expected the inductive argument more in terms of validity, he said that he gave numeric examples after an algebraic explanation. In other words, the teacher thought his way of teaching might have influenced the students.

The teacher was informed about the frequencies for Task 2: the frequencies of the axiomatic and the inductive arguments were high as he expected, the frequencies of the transformational and the authoritarian arguments were equal in the third line, and the frequency of the perceptual argument was the lowest. As the reason for the difference, the teacher thought the students might not be able to think when there was no number in the perceptual argument.

In Task 3, the teacher guessed the rank of the frequencies would be "the inductive > the axiomatic > the authoritarian." He said that the teacher would understand the inductive argument since it started with numbers and the students understand better by seeing numbers. The rank of the frequencies was presented: the inductive argument was selected most; the axiomatic and authoritarian arguments were selected almost equally. He was surprised and could not make a comment about why the frequencies of the axiomatic and authoritarian arguments were close to each other.

4.2.3 Scoring of the Arguments Like They Were Responses of His Students

The teacher was asked to score the arguments like they were responses of his students to get insights about his evaluation of the arguments in a class. Based on the scores the teacher would give and the criteria the teacher would use, Table 4.28 was formed. The results showed that the teacher would give the highest scores to the axiomatic arguments in each task. Also, he gave the transformational arguments high scores. In addition to the axiomatic arguments, the teacher gave also 10 points to the perceptual argument in Task 2 and the inductive argument in Task 3.

Table 4.28 The Teacher's Scoring of the Arguments Like They Were Responses of His Students

Types of the Arguments	Task 1	Task 2	Task 3
Authoritarian	-	2 Exp./Understanding Source of Explanation	2 -
Symbolic	2 -	-	-
Inductive	8 Example	6 -	10 Example
Perceptual	-	10 -	-
Transformational	8 Example	7 Liking	-
Axiomatic	10 Validity	10 -	10 Algebra Liking Practical

In Task 1, the teacher confidently gave the maximum point to the axiomatic argument because it showed that the statement always comes true. He also gave the inductive and transformational arguments high scores because he said, "*They make generalization by here (by using examples), it is okay, but exceptions may occur, as I said*." The teacher thought these arguments showed the statement was true, but he had stated that he was aware of the limitations of the example in Task 1. In terms of the symbolic argument, he did not give any criteria regarding validity; he criticized the symbolic argument in terms of incorrectness when asked about the convincingness of the arguments.

In Task 2, the teacher gave a low score to the authoritarian argument because he thought the argument did not include any explanation and was just a statement from an authority. In addition, he said that he liked the transformational argument more than the inductive argument. The teacher did not give more criteria for scoring the arguments, but he said his convincingness inevitably affected his scores.

In Task 3, the teacher gave the maximum score to the axiomatic argument, again. The teacher said that he liked it because it directly reached the conclusion in a practical way. He considered using algebra to create a general rule as making the argument practical. Then, the teacher gave the inductive argument the maximum point, too. He said the argument used examples and reached a generalization. He considered both arguments the same in terms of reaching generalization.

The teacher said the inductive argument was like a form of the axiomatic argument, adding examples in Task 3. Also, he mentioned that he would see the examples as extra and give 10 points if a student came up with an inductive argument. It can be said that the teacher did not distinguish between inductive and deductive reasoning in Task 3. In addition, he criticized the inductive argument in terms of the practicality during the convincingness evaluation; however, he did not use this criterion to evaluate his students' arguments, and he just focused on whether the argument includes a statement that it always works.

4.2.3.1 Expectancies from the students

As a reminder, the students were asked to score the arguments regarding how many points they thought the teacher would give. In task 1, the students scored like "the axiomatic > the inductive > the symbolic ≈ the transformational." Again, the teacher thought the reason for the high scores of the axiomatic and inductive arguments might be related to the way of his teaching. However, in terms of other arguments, the teacher said that he had expected that the transformational argument would get higher.

In Task 2, he said, "*The students would know that I would give the highest score to Student 5 (the axiomatic argument)*." He thought the students would consider the validity of the arguments. Then, he was undecided about whether the students thought the teacher would give a higher score to the inductive or the perceptual argument. He tended to select the perceptual argument. However, he thought the student might not understand the perceptual argument since, previously, they did not give a higher score in terms of convincingness. He thought the students would also consider whether there are examples or not to reach a generalization. Then, the transformational and the authoritarian arguments would be, respectively, in the fourth and fifth lines, in his opinion. The teacher's expectations about the scores were consistent with the current situation. Again, the teaching method he used was the teacher's interpretation of these similarities between his expectancy and the rank of the real averages.

In Task 3, the teacher expected the students to score the arguments in terms of teacher expectancy, such as "the axiomatic > the inductive > the authoritarian. To score the teaching expectancy, he expected the students to consider whether the argument made a generalization like a rule or a formula. When the teacher was informed that the rank was "the inductive > the axiomatic > the authoritarian," He said "*I think they thought for themselves. It's easier to understand this way (inductive). Because it's nice to them (the students like the inductive argument)*."

4.2.4 The Summary of The Findings from the Teacher

As the findings regarding a sub-research question related to the teacher side of the study, the teacher found the axiomatic arguments most convincing in each task and looked for "generalizability" of the arguments to evaluate convincingness. He emphasized the "always" term in the statements and mentioned the arguments should include a generalization and show the statement is always true. Thus, the "validity" for all cases was also a criterion for the teacher. Moreover, the teacher considered variables as the components that make an argument general and valid for all cases. For instance, he evaluated the inductive argument in Task 3 as a valid argument because he thought the argument reached a generalization with the algebraic expressions. However, he criticized the argument in terms of "practicality" and gave a high but not max point to the convincingness of the argument. In his opinion, while the axiomatic arguments could directly show the truth of the statement, the inductive argument used extra examples.

From the teacher's perspective, the variables were not limited to algebraic letters. He considered the perceptual argument in Task 2 valid because he thought the measurement of angles could be any number, and they were variables since the angle symbols did not include any specific numbers. In other words, from his perspective, not only axiomatic arguments were valid, but also inductive and perceptual arguments can be valid if they include variables. In that way, for the teacher, the "variables" were the criterion for evaluating the arguments' validity, which is the finding regarding the fifth sub-research question.

Although validity was a criterion for convincingness, not all convincing arguments were valid in the teacher's opinion. The teacher mentioned that the "examples" were a way to prove the truth of the statement and a logical solution, but he emphasized the limitation of the examples in terms of being true for all cases by saying, "*Maybe it won't come out true somewhere*." On the other hand, the proving with arbitrarily selected examples convinced him. Despite the validity, both axiomatic, inductive, and transformational arguments were convincing, and the "examples" were also a criterion to evaluate the convincingness of the arguments for the teacher.

Regarding the last sub-research question, which investigated the criteria the teacher used to score the arguments like they were responses of his students, the teacher gave the highest points to the argument which he considered "valid" and mentioned that these arguments included "general" rules. Even though the teacher criticized the inductive argument in Task 3 in terms of practicality and saw the examples as extra, he gave 10 points to it because of its validity. In short, the "validity" for all cases was a criterion the teacher used. In addition, like the convincingness, the teacher gave high scores to inductive and transformational arguments. He saw examples as a way to show the statements were true but did not give 10 points because he again emphasized the limitations of examples to show validity for all cases.

When the averages of students' scores were presented to the teacher, he thought his teaching method might affect these averages. He said that he gave a general algebraic form and then gave example cases to ensure better understanding in his lessons, and he thought this might be a reason for the high averages of axiomatic and inductive arguments.

Another finding from the teacher was that he was not convinced by the arguments from the external conviction proof schemes. For the authoritarian arguments, he said, "*How do we know it's true?*" and emphasized that they were not a valid way to show the truth and there was a possibility of misremembering or incorrect information in a book. For the symbolic argument in Task 1, he noticed the incorrect algebraic expressions. At first, he hesitated about whether he missed a part, but then he noticed that the task might include an incorrect argument and reduced his score for convincingness. Still, his first and last scores for the symbolic argument were the lowest scores for convincingness. Parallelly, he did not select the symbolic argument as valid and gave the lowest score if it was an answer of his students.

It was observed that the teacher saw the transformational arguments in Tasks 1 and 2 as a form of inductive argument. For instance, he gave the same scores to the transformational argument and inductive argument in Task 1. On the other hand, in Task 2, he noticed that the transformational argument was a generic example, and giving another number would not change the result. Because of that, he gave it 1 point higher than the inductive argument. However, he thought it was not enough for validity. He mentioned that it was limited and took too long to try other numbers. In short, the transformational arguments were example cases for the teacher, but the argument in Task 2 was a generic example for him even if he did not consider it valid.

4.3 The Comparisons Between the Findings from the Teacher and His Students

First, the findings from the students were investigated based on the numeric scores, the emerged codes, and the insight from the interviews. Second, the interview findings from the teacher were presented. This section aims to present the similarities and differences between the findings of the teacher and the students. They were presented below without regard to the order of importance.

1. Based on the averages, while the inductive arguments were the most convincing in each task for the students, they were the axiomatic arguments for the teacher. On the other hand, it can be said that both inductive and axiomatic arguments were found convincing for both the teacher and students.

2. In Task 1, the least convincing argument was the transformational argument, while it was the symbolic argument for the teacher. On the other hand, in other tasks, the least convincing argument was the same for both the teacher and students: the authoritarian one.

3. In terms of validity, the teacher selected the axiomatic arguments in each task; however, the axiomatic arguments in each task were not considered valid by around half of the students. Similarly, the inductive arguments in

each task were selected by more than half of the students; the teacher considered the inductive argument valid in only Task 3 because there was a general algebraic expression at the end of the argument.

4. The teacher thought the perceptual argument in Task 2 was a valid argument. On the other hand, 90% of the students did not consider the perceptual argument valid. Thus, the evaluation way for the perceptional argument seems different for the teacher and the students.

5. There was a similarity in terms of teacher expectancy. Students thought the teacher would give high scores to the axiomatic arguments, and similarly, the teacher gave the maximum score to the axiomatic arguments in each task.

6. Based on the averages, the students thought the authoritarian arguments in Tasks 2 and 3 would get the lowest scores from the teacher; similarly, the teacher gave the lowest scores to these arguments.

7. In terms of the emerged codes, while the students mostly focused on the arguments' being explanatory/understanding for each theme (i.e., convincingness, validity, and teacher expectancy), the teacher focused on the validity and using algebra. It should be noted that the validity code referred to explaining why the statement always worked, while it was observed that the explanatory/understanding code usually referred to "what" the argument says, not "why" the argument said so.

8. While some of the students thought the teacher would use the practicality criterion to score arguments, the teacher did not pay attention to the practicality when he scored the arguments when thinking like they were responses of his students.

9. The way to show an incorrect statement in an argument was similar for the teacher and some of the students. Like the teacher, there were students who were aware of the incorrectness of the symbolic argument in Task 1. They noticed that "$n + 3$" was not 3 times "$+1$." And both some students and the

teacher gave some numbers to show that incorrectness. On the other hand, the students did not emphasize that "$n + 1 + 2$" was not the algebraic form of the sum of consecutive whole numbers, unlike the teacher did. Thus, while the students just focused on "$n + 3 \neq 3 \cdot (+1)$," the teacher noticed the incorrect match between algebraic and verbal expressions.

10. The teacher's evaluation of the transformational and inductive argument in Task 1 was similar to some of the students' evaluations. Both considered these two arguments similar, except that one of them included numbers and the other included drawing points. In addition, some students thought drawing these dots might waste time, as the teacher expected. So, the teacher's opinions about these two arguments were similar to those of some students.

11. The transformational argument in Task 2 was confusing for many students. Some students did not understand the explanation, and some students understood but did not notice the example in the argument was the generic example, while the teacher noticed that the transformational argument in Task 2 implied several example cases. However, still, both the teacher and 80% of the students did not consider the transformational argument in Task 2 as valid.

CHAPTER 5

DISCUSSION AND IMPLICATIONS

The current study aimed to investigate the proof evaluations of 8^{th}-grade students and their teacher. The participants evaluated the arguments in terms of convincingness and validity. In addition, the teacher was asked to score the arguments like they were the responses of his students, and the students were asked to score the arguments based on how many points they thought their teacher would give. The study also tried to compare the findings to explore students' and teacher's proof evaluations. The findings from the teacher, and the students, and the comparison of their findings will be respectively discussed. Then, the implications and suggestions for future research will be presented.

5.1 The Proof Evaluation of the Teacher

It was noted that the teacher found the axiomatic arguments convincing because he was looking for whether the arguments were general and in algebraic form. The teacher used algebraic forms and mathematical symbols as criteria to prove the statements because he considered algebra as a way to represent a general case. This can be expected because past researchers (e.g., Boero, 2006; Ko & Hagen, 2013; Uygan et al., 2014) mentioned that many teachers look for generality and algebraic expression to consider an argument as proof.

The teacher mentioned that "showing always works" was also a reason for being convinced by the axiomatic arguments. He used "validity for all cases" as a criterion. Similarly, many of the teachers in past research (e.g., Boero, 2006; Ko & Hagen, 2013) emphasized the proof should be general. However, deductive reasoning may not be required for the teacher to show validity. For instance, he claimed that the inductive argument in Task 3, which included a general algebraic statement at the end, could show that the statement always works because he interpreted that as the argument reaching a generalization. As mentioned, the "use of algebra" had a big impact on convincing the teacher.

Like the findings from Boero (2006), İmamoğlu and Yontar Toğrol (2015), and Ko and Hagen (2013), the teacher was aware of the limitations of the use of examples to make generalizations. He emphasized that there could be cases that may not hold for the general statement in empirical arguments. On the other hand, unlike the majority of teachers in the study of Ko and Hagen (2013), the teacher considered the inductive (empirical) arguments as convincing arguments even if the argument did not include any algebraic form.

Another finding was the teacher found the perceptual argument in Task 2 convincing. He emphasized that there were no specific numbers on the angle symbols, and one could assign any numbers there like one can assign any numbers to algebraic variables. In other words, angle symbols without numbers were considered a powerful tool playing the same role as algebraic variables for the teacher. Dogan and Williams-Pierce (2021) indicated that the teachers look for variables in generic examples. Even though it was not a generic example, the teacher thought the example used in the perceptual argument could generate other cases. So, this might lead the teacher to accept the perceptual argument as valid for all cases.

Knuth (2002b) found that the teachers thought the empirical arguments were more suitable for the early grades than algebraic arguments. Similarly, the teacher mentioned that the students might have difficulty understanding algebra, and the examples with numbers might be more comprehensible for the students. He emphasized the usefulness of examples for his students, like the participants in the study of Dalkılıç and Zeybek Şimşek (2022). However, he did not mention that the axiomatic argument was not suitable for his students. During the interview, the teacher mentioned that, in his lessons, he tended to start with the algebraic generalization and then moved with examples to make it more comprehensible in his lessons. It was not clear whether the teacher used deductive reasoning while justifying algebraic statements, but he stated that he used algebra in his instruction.

In the study of Morris (2007), the preservice elementary teachers were able to distinguish valid and invalid arguments, and the teachers tended to think empirical arguments could prove the general statement if the discussion did not include valid arguments. In the current study, it can be said that the teacher was aware that an inductive argument could not show the validity for all cases. However, although there were both arguments with deductive and inductive arguments, the teacher also considered that arguments with inductive reasoning could stand for validity if they had variable statements at the end.

Ozgur (2017) underlined the similarities between the proof conception of the teacher and the students and emphasized that the instructional activities might have affected students' proof conceptions. Similarly, the teacher thought his teaching way might have influenced the students' evaluations. The teacher believed that his way of teaching was a possible reason why the inductive and axiomatic arguments were convincing for the students. In addition, many students', and the teacher's proof evaluation for authoritarian arguments were similar. In that way, this finding was parallel to the findings of Ozgur (2017).

In the studies of Çontay and Duatepe Paksu (2019) and Dalkılıç and Zeybek Şimşek (2022), it was observed that the responses of teachers and teacher candidates could be categorized according to the external proof scheme. A teacher might be affected by proof rituals in the college or mathematical symbols. In the current study, this case was not observed. The teacher emphasized the unreliability of authoritarian arguments and gave low scores to arguments with incorrect algebraic expressions.

Harel and Sowder (1998, 2007) mentioned that one can have multiple proof schemes. Similarly, Çontay and Duatepe Paksu (2019) determined more than one proof scheme in the responses of the participants. In the current study, the teacher consistently preferred the axiomatic arguments in each task at most and did not prefer the arguments from the external conviction proof schemes. In addition, he was consistent with the inductive arguments in each task. He thought the inductive arguments were a way to show the truth of the statement. On the other hand, in terms of the transformational arguments, he had different comments in Tasks 1 and 2. Although he considered the transformational arguments as a form of an example, he noticed the generic example in Task 2. This might be caused by the familiarity with generic examples. While he mentioned the ways he possibly used in his lessons, there was no emphasis on generic examples. The teacher might not be familiar with generic examples, and that might have caused different evaluations of the transformational arguments.

5.2 The Proof Evaluation of the Students

In some past studies (e.g., Heinze & Reiss, 2003; Liua et al., 2016), it was indicated that the empirical arguments convinced the students because they thought such arguments could lead to understanding how the statement was true. In the current study, the inductive arguments, which were example-based, were the most convincing in each task. In addition, the explanatory/understanding code was one of the most frequently emerged criteria. Also, the interviews showed that the students found the inductive arguments understandable and convincing. On the other hand, some students were aware of the limitations of the use of examples, like the findings of Healy and Hoyles (2000) and Ozgur (2017). For instance, Student 15 said there could be some cases that do not hold for the statement, similar to the explanation of the teacher. However, based on the questionnaire results, it can be said that the students who were aware of the limitation of the example were the minority.

In the studies of Heinze and Reiss (2003) and Healy and Hoyles (2000), the students thought empirical arguments were not a form of proof. However, in this study, it seems that there were a limited number of students who were aware of the limitation of examples, based on the averages from the questionnaire. On the other hand, many students thought algebraic arguments were preferable for teachers in past research (e.g., Arslan, 2007; Healy & Hoyles, 2000). This might be coherent with the current study because the students considered the axiomatic arguments, which were in algebraic form, would get higher points from the teacher than the inductive arguments in Tasks 1 and 2. This might be especially due to the presence of algebra in those statements as found out in the interviews.

Although the axiomatic arguments in Tasks 1 and 2 got the highest points in terms of teacher expectancy, the inductive arguments were also considered as the arguments that would get high points from the teacher in each task. It should be remembered that the teacher gave inductive arguments 10 points like the axiomatic arguments in each task. Therefore, it can be said that he accepted the examples as a way to justification in his class, unlike the classroom teachers in the study of Isler (2015). In other words, the teacher’s teaching preferences might have influenced the student’s evaluation of the teacher’s expectancy.

Despite minority cases, the inductive arguments generally convinced the class, and more than half of the students considered them valid. Also, in the study of Aydoğdu İskenderoğlu (2003) and Sen and Guler (2015), students’ answers were mostly from the inductive proof schemes. Moreover, the responses from the external conviction proof schemes were also widespread in these studies. In the current study, there were students who were convinced and also unconvinced by the authoritarian arguments and symbolic arguments.

The averages showed that there were a limited number of students who noticed the incorrect algebraic expressions. This might be caused by the lack of understanding of algebra. During the observation, it was observed that student 9 misinterpreted the algebraic expressions, which caused her to be convinced by the symbolic argument. Understanding algebra better might be helpful to notice incorrect expressions and not be convinced by just the mathematical symbols. In addition, there were students who considered algebraic expressions as examples. They were not aware of how using variables might help justify a statement in all cases. Thus, such focus might be beneficial for the students to be aware of the generalizability power of algebra and the validity of the axiomatic arguments.

In terms of transformational arguments, Arslan (2007) had a similar task to Task 1 of the current study, and her result showed that the students liked and understood the visual transformational argument. On the other hand, in the study of Sen and Guler (2015), few students' responses were categorized under the transformational proof scheme. Similarly, in the current study, the students mostly were not convinced by the transformational arguments in Tasks 2 and 3 and did not notice the generic example in the arguments. The long explanation of these arguments might have decreased the understandability of the students. Or, the generic examples were not around the teaching preferences of the teacher, unlike the study of Isler (2015), and those preferences might have led the student to pay less attention to the deductive structures in a transformational argument.

During the interviews with students, the "explain why" criteria were not among the frequently emerged criteria. Although explanatory/understanding criteria was one of the most used codes, this code may not be related to why the statements were true. Similar to the findings of Liua et al. (2016), the easiness of understanding stood for examples, illustrations, and easy procedures. The students mostly focused on understanding "how" the statement was true, and some students focused on "validity for all cases" of the argument in their evaluation, but there was slight interest in understanding "why" the statement was true.

In terms of visual criteria used by the students, the findings of the current study showed similarity to the findings of Liua et al. (2016). Some students found visual elements comprehensive. No student found the visuals confusing in the current study, but some students criticized the visuals in terms of practicality. Like in Liua et al. (2016), practicality (or being straightforward) emerged during the interviews. Even though the practicality did not frequently emerge in the questionnaire, the students emphasized the practicality when asked about convincingness and teacher expectancy during the interviews. The situation might be caused by being in 8^{th} grade in Turkey. The students were being prepared for an exam to enroll in qualified high schools, and this exam marathon might have affected the lessons and the priority of the students for their evaluations. In a high-stakes examination with multiple-choice questions, practicality could be more essential than understanding why a mathematical statement is true. This situation might direct the students to understand what the statement claims with examples of cases and the practicality of an argument instead of the focus on generalizability and the reasoning of the arguments.

5.3 The Similarities and Differences Between the Students and the Teacher's Proof Evaluations

Although the most convincing argument was the axiomatic for the teacher and the inductive was for the students, the teacher and the students were convinced by both axiomatic and inductive arguments. This finding was coherent with the findings of Ozgur (2017), who emphasized the similarity of the proof conception of the teacher and the student.

In Isler (2015), the classroom teachers thought the students would rely on authority; however, the teacher in the current study predicted that authoritarian arguments would not convince the students. As he predicted, the authoritarian arguments were the least convincing for the students and the teacher based on the averages from the findings of the questionnaire. On the other hand, it should be noted that this similarity did not occur in the symbolic argument, which was another type of argument from the external conviction proof scheme. Many students did not notice the incorrect algebraic expression in the symbolic argument, contrary to the prediction of the teacher.

Students in Liua et al. (2016) considered the transformational and perceptual arguments convincing. In the current case, while the teacher found these arguments convincing, the students found them confusing or lack of explanation. On the other hand, despite the difference in convincingness, the teacher and the students did not notice that the transformational arguments could show the statement was always true. This might be caused by the instructional activities in the lesson. Unlike many teachers in the study of Isler (2015), the teacher did not mention that he had used generic examples. In addition, while the teacher noticed the generic example in the transformational argument in Task 2, he considered the transformational argument in Task 1 as a single example. Thus, the teacher's approaches to the generic examples in the transformational argument were inconsistent.

Another difference was observed in the evaluation of validity. While the teacher considered the axiomatic arguments as valid arguments, more than half of the students did not see them as valid. This can be caused by the fact that the students might not be aware of the generalizability power of algebra, like some students in Liua et al. (2016). While some students knew that they could write any number to the variables and it had generalizability power, some students did not give any such comments. In addition, the frequency of the algebra code for validity in the questionnaire was very low. This situation also supports the idea that the majority did not use algebra as a criterion for validity.

Finally, as Uygan et al. (2014) mentioned, the teacher in the current study used generalizability and understandability as some of the main criteria. In terms of the frequently used criteria, it can be said the teacher mostly focused on examples, validity for all cases, and using algebra like some teachers in past research (e.g., Boero, 2006; Knuth, 2002; Ko & Hagen 2013). The students mostly focused on examples and understandability of the argument, like some students in past research (e.g., Bieda & Lepak, 2014; Liua et al., 2016). Moreover, there was a slight focus on why the statement was true for both teachers' and students' evaluations.

To sum up, there were both similarities and differences between the proof evaluation of the teacher and his students. Both used examples as one of the main criteria to evaluate arguments and were convinced by inductive arguments. On the other hand, while the teacher focused on generalizability and accepted the arguments with variables as valid arguments, the students mostly considered the inductive arguments valid because they could show how the statement was true. The reason might be that many students were not aware of the generalizability power of the variables. However, it should be noted that the teacher accepted the inductive argument with algebraic generalization in Task 3 as valid even if it did not include any deductive justification. Thus, there might be less focus on the deductive justifications by both sides. And finally, the teacher said that he would give high points to the inductive and axiomatic arguments as the students expected.

5.4 Implications and Recommendations

The study showed the possible effect of the teacher's way of teaching on the students' proof evaluation. A similar connection was indicated in Ozgur (2017). The teacher can play an essential role in drawing boundaries about which kind of arguments can be accepted as proof. Regarding this situation, it would be beneficial for the teacher to underline the limitations of the empirical arguments, to question why the statement is true, and to emphasize the importance of deductive reasoning. On the other hand, it was observed that the students can easily see how the statement is true with inductive arguments instead of axiomatic arguments. At that point, the transformational arguments can be beneficial because these arguments investigate deductive reasoning in a specific case. However, mathematics educators should be more careful about the presentation of transformational arguments. In the study, it was observed that the students could find these arguments confusing or the same as the empirical arguments. Educators should avoid showing a generic example like a single example and emphasize the deductive reasoning in it. In addition, educators should try to use simple explanations in teaching activities to ease following the reasoning.

The results of the study may have implications for teacher education. For instance, the teacher accepted the inductive argument (example-based) as valid for all cases if it included an algebraic generalization at the end. That meant the teacher did not notice that empirical arguments were invalid for statements that cover an infinite set. Although the mathematical courses in college aim to help teacher candidates gain valid reasoning, the candidates should learn to notice students' reasoning and guide them to develop their reasoning in mathematics lessons in all grades. For example, the generic examples may help the students to gain deductive reasoning without using algebraic expressions; however, while the teacher noticed the generic example in one of the transformational arguments, he did not notice the other. Such activities and courses in the teacher education programs and professional development programs might help pre-and in-service teachers notice the deductive structures in the arguments and help enhance the activities that improve the reasoning of the students. Therefore, even though the curriculum does not include the term "proof," adding activities and courses that aim to teach reasoning and proof in early grades and/or improving current courses in that sense can enhance the proof conceptions of the teachers, and so help their students to notice and use deductive reasoning.

Although there were similarities between the proof evaluation of the teacher and the students, there were also differences. Unlike the expectation of the teacher, many students did not notice incorrect algebraic expressions and did not focus on the generalizability power of the algebra while evaluating the validity of the arguments. Therefore, giving more attention to the generalizability power of algebra might help to enhance the proof concept of the students.

Most of the research instruments in the study were developed by the researcher. The arguments and the questions in the questionnaire can be used for future research because the instrument suitably helped to get the intended kind of information. The instrument can be reused by developing or adapting to future research.

The study was limited to the number and geometry learning areas. In future research, the learning areas can be extended or specified. Also, having only correct mathematical statements to be justified in each task was another limitation. Thus, in future research, the research instrument can be developed such that it includes incorrect statements as well as correct ones. Like the statements and arguments, the questions can be extended. The questions in this study were limited to convincingness, validity, and teacher expectancy. They can be revised to include other evaluation components. Moreover, the data was collected by written questionnaire and follow-up interviews. Adding extra ways to collect data, such as observation, can enhance a variety of the data in future research. Furthermore, this study put possible similarities and differences, but these results cannot be generalized to all classroom settings because of the nature of the qualitative study. In future research, a quantitative study can be conducted to investigate general trends in Turkey.

REFERENCES

Arslan, Ç. (2007). *The development of elementary school students on their reasoning and proof ideas*. (Doctoral dissertation) Bursa Uludağ University.

Aydoğdu İskenderoğlu, T. (2003). *Middle-grade students' justifications for mathematics problems.* (Master dissertation) Bolu Abant İzzet Baysal University.

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., Jahnke, 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.

Baxter, P., & Jack, S. (2008). Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers. *The Qualitative Report*, *13*(4), 544-559. Retrieved from https://nsuworks.nova.edu/tqr/vol13/iss4/2

Boero, P. (2006). Habermas' theory of rationality as a comprehensive frame for conjecturing and proving in school. *In Proceedings of the 30th Conference of the International Group for the Psychology of Mathematics Education* 2, 185-192).

Bieda, K., & Lepak, J. (2014). Are you convinced? Middle-grade students' evaluations of mathematical arguments. *School Science and Mathematics, 114*(4), 166–177.

Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five traditions* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Çontay, E. G., & Duatepe Paksu, A. (2019). The proof schemes of Preservice Middle School Mathematics teachers and investigating the expressions revealing these schemes. *Turkish Journal of Computer and Mathematics Education, 10*(1), 59–100. https://doi.org/10.16949/turkbilmat.397109

Dalkılıç, T., & Zeybek Şimşek, Z. (2022). Investigating Mathematics and Prospective Mathematics Teachers' Argument Construction and Evaluation Processes. *Pamukkale University Journal of Education, 54*, 357–384. https://doi.org/10.9779/pauefd.814059

Dogan, M. F., & Williams-Pierce, C. (2020). *The role of generic examples in teachers' proving activities. Educational Studies in Mathematics*, 106(1), 133–150. https://doi.org/10.1007/s10649-020-10002-3

Ellis, A. B., Bieda, K., & Knuth, E. J. (2012). Proof and Proving: The Big Ideas and Essential Understandings. *In Developing essential understanding of proof and proving for teaching mathematics in grades 9-12* (pp. 7–10). essay, The National Council of Teachers of Mathematics, Inc.

Ellis, A. B., Ozgur, Z., Vinsonhaler, R., Dogan, M. F., Carolan, T., Lockwood, E., Lynch, A., Sabouri, P., Knuth, E., &; Zaslavsky, O. (2019). Student thinking with examples: The criteria-affordances-purposes-strategies framework. *The Journal of Mathematical Behavior, 53*, 263–283. https://doi.org/10.1016/j.jmathb.2017.06.003

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in Education* (8th ed.). McGraw Hill.

Guest, G., Namey, E. E., & Mitchell, M. L. (2013). *Collecting Qualitative Data: A Field Manual for Applied Research*. https://doi.org/10.4135/9781506374680

Hanna, G., & de Villiers, M. (2008). ICMI study 19: Proof and proving in Mathematics Education. *ZDM*, *40*(2), 329–336. https://doi.org/10.1007/s11858-008-0073-4

Hanna, G. (2000). Proof, Explanation and Exploration: An Overview. *Educational Studies in Mathematics,* 44(1/2), 5–23. https://doi.org/10.1023/a:1012737223465

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

Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. *Research in Collegiate Mathematics Education. III*, 234–283. https://doi.org/10.1090/cbmath/007/07

Heinze, A., & Reiss, K. (2003). Reasoning and proof: Methodological knowledge as a component of proof competence. *In Proceedings of the Third Conference of the European Society for Research in Mathematics Education, Bellaria, Italy*.

İmamoğlu, Y., & Yontar Togrol, A. (2015). Proof construction and evaluation practices of prospective mathematics educators. *European Journal of Science and Mathematics Education, 3*(2), 130–144. https://doi.org/10.30935/scimath/9427

Isler, I. (2015). *An Investigation of Elementary Teachers' Proving Eyes and Ears*. (Unpublished Doctoral dissertation) University of Wisconsin - Madison.

Kline, M. (1970). Logic versus pedagogy. *The American Mathematical Monthly, 77*(3), 264–282. https://doi.org/10.2307/2317710

Knuth, E. J. (2002a). Proof as a tool for learning mathematics. *The Mathematics Teacher, 95*(7), 486–490. https://doi.org/10.5951/mt.95.7.0486

Knuth, E. J. (2002b). Teachers' conceptions of proof in the context of secondary school mathematics. *Journal of mathematics teacher education*, 5(1), 61-88.

Ko, Y. Y., & Hagen, C. J. (2013). Conviction and validity: Middle school mathematics teachers' proof evaluations. *In 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).

Korstjens, I., & Moser, A. (2017). Series: Practical guidance to qualitative research. part 4: Trustworthiness and publishing. *European Journal of General Practice*, *24*(1), 120–124. https://doi.org/10.1080/13814788.2017.1375092

Lin, F. L., Yang, K. L., Lee, K. H., Tabach, M., & Stylianides, G. (2012). Principles of task design for conjecturing and proving. *Proof and proving in mathematics education*, 305-325.

Liu, Y., Tague, J., & Somayajulu, R. (2016). What Do Eighth Grade Students Look for When Determining If a Mathematical Argument Is Convincing? *IEJME — Mathematics Education, 11(7), 2373–2401.*

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.), *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.

Millî Eğitim Bakanlığı (MEB) (2018). Matematik dersi öğretim programı 1-8. sınıflar. Retrieved July 25, 2020, from http://mufredat.meb.gov.tr/ProgramDetay.aspx?PID=329

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 [NCTM]. (2000). Principles and Standards for School mathematics. Reston, VA: Author.

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods, 19,* 160940691989922. https://doi.org/10.1177/1609406919899220

Ozgur, Z. (2015). *Relationships Between Students' Conceptions of Proof and Classroom Factors*. (Unpublished Doctoral dissertation) University of Wisconsin - Madison.

Pfeiffer, K. (2011). *Features and purposes of mathematical proofs in the view of novice students: Observations from proof validation and evaluation performances*. (Doctoral dissertation) National University of Ireland, Galway.

Reid, D., & Vallejo Vargas, E. (2018). When is a generic argument a proof? *ICME-13 Monographs,* 239–251. https://doi.org/10.1007/978-3-319-70996-3_17

Sen, C., & Guler, G. (2015). Examination of secondary school seventh graders' proof skills and proof schemes. *Universal Journal of Educational Research, 3*(9), 617–631. https://doi.org/10.13189/ujer.2015.030906

Stake, R. E. (1995). *The art of case study research*. Sage.

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

Stylianides, A. J., Bieda, K., & Morselli, F. (2016). Proof and argumentation in mathematics education research. In Á. Gutiérrez, G. C. Leder, & P. Boero (Eds.), *The second handbook of research on the psychology of mathematics education* (pp. 315-351). Rotterdam/Boston/Taipei: Sense Publishers.

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). National Council of Teachers of Mathematics.

Uygan, C., Tanışlı, D., & Köse, N. Y. (2014). Research of pre-service elementary mathematics teachers' beliefs in proof, proving processes and proof evaluation processes. *Turkish Journal of Computer and Mathematics Education, 5*(2), 137–157. https://doi.org/10.16949/turcomat.33155

Yackel, E., & Hanna, G. (2003). Reasoning and proof. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 227-236). Reston, VA: The National Council of Teachers of Mathematics.

Yin, R. K. (2009). *Case study research: Design and methods* (4th ed., Vol. 5). Sage.

APPENDICES

A. APPROVAL OF THE UNIVERSITY HUMAN SUBJECTS ETHICS COMMITTEE

UYGULAMALI ETİK ARAŞTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

DUMLUPINAR BULVARI 06800
ÇANKAYA ANKARA/TURKEY
T: +90 312 210 22 91
F: +90 312 210 79 59
ueam@metu.edu.tr
www.ueam.metu.edu.tr

Konu: Değerlendirme Sonucu

13 EYLÜL 2022

Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

İlgi: İnsan Araştırmaları Etik Kurulu Başvurusu

Sayın Işıl İşler BAYKAL

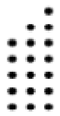
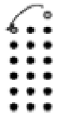
Danışmanlığını yürüttüğünüz Halil İbrahim ŞİMŞEK'in "**8. Sınıf Matematik Öğretmeninin ve Öğrencilerinin İspata Yönelik Argümanları Değerlendirmelerinin İncelenmesi**" başlıklı araştırması İnsan Araştırmaları Etik Kurulu tarafından uygun görülerek gerekli onay **0479-ODTUİAEK-2022** protokol numarası ile onaylanmıştır.

Bilgilerinize saygılarımla sunarım.

B. WRITTEN SURVEY

Dört öğrenciden aşağıdaki ifadenin doğru veya yanlış olduğunu nedenlerini açıklayarak göstermeleri istenmiştir.

"Ardışık 3 tam sayının toplamı her zaman ortadaki sayının 3 katıdır."

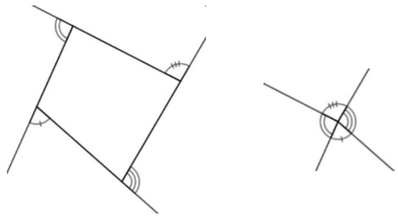
Öğrenci 1: Herhangi ardışık 3 sayıyı alabilirim. Eğer baştaki sayı "n" ise ortadaki sayı "$n+1$", son sayı "$n+2$" olur. O zaman ardışık 3 sayıyı toplarsak; $n+(n+1)+(n+2)=3n+3$ $3n+3$ sayısı ortadaki $n+1$ sayısının 3 katı olduğuna göre ifade doğrudur.	**Öğrenci 2:** Yandaki üç nokta sütununun ardışık üç sayıyı gösterdiğini hayal edelim. İlk sütun birinci sayıyı, ikinci sütun ikinci ve üçüncü sütun da son sayıyı göstersin.  Son sütunun en üstteki noktasını alıp ilk sütuna koyabiliriz. Böylelikle her sütunda, ortadaki sütundaki nokta sayısı kadar sütun olur.  Toplam nokta sayısı orta sütundaki nokta sayısının üç katı olduğuna göre ifade doğrudur.
Öğrenci 3: 5, 6 ve 7 ardışık üç sayıdır ve 5 + 6 + 7 = 18, ve 3 × 6 = 18. 7, 8 ve 9 ardışık üç sayıdır ve 7 + 8 + 9 = 24, ve 3 × 8 = 24. 569, 570 ve 571 ardışık üç sayıdır ve 569 + 570 + 571 = 1710, ve 3 × 570 = 1710. İfadeyi farklı ardışık sayılarla denedim. Bu üç örnekte de doğru çıktığına göre ifadenin doğru olduğunu bilebilirim.	**Öğrenci 4:** Herhangi ardışık 3 sayıyı alabilirim. Eğer baştaki sayı "n" ise ortadaki sayı "$n+1$", son sayı "$n+2$" olur. O zaman ardışık 3 sayıyı toplarsak; $n+1+2=n+3$ $n+3$ sayısı ortadaki +1 sayısının 3 katı olduğuna göre ifade doğrudur.

Beş öğrenciden aşağıdaki ifadenin doğru veya yanlış olduğunu nedenlerini açıklayarak göstermeleri istenmiştir.

"Herhangi bir dörtgenin dış açıları toplamı her zaman 360°'dir."

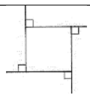
Öğrenci 1: Ders kitabında dörtgenlerin dış açıları toplamının 360° olduğu yazıyor. Bu yüzden yukarıdaki ifade doğrudur.

Öğrenci 2: Herhangi bir dörtgen ve dış açılarını çizelim.

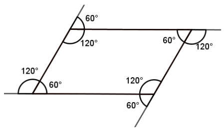


Bu dış açıları kesip bir araya getirirsek bir tam açı oluşturur, yani 360° yapar.

Öğrenci 3:
Örnek olarak kare veya dikdörtgeni ele alalım. Bütün dış açıları 90°'dir ve 4 tane dik açıtoplam 360° yapar.



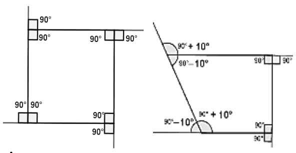
Veya bir açısı 60° olan bir paralelkenarı alalım. Karşısındaki açı 120° olur. Dış açıları da 2 tane 120° ve 2 tane 60° olur yani yine toplamları 360° yapar.



Hem karede hem paralelkenarda dış açılar toplamı 360°.

Denediğim örnekler de doğru olduğuna göre hangi dörtgende denersek deneyelim sonuç 360° olacaktır.

Öğrenci 4: Bir dikdörtgeni ele alalım. Dikdörtgenin iç açıları toplamı 360° ve dış açıları toplamı 360°. Dikdörtgenin açılarıyla oynayarak diğer tüm dörtgenleri elde edebiliriz.



İç açılarından birini 10° arttıralım. O zaman o açıya ait dış açının da 10° azalmış olması lazım.

Ancak bir iç açıyı arttırsak iç açılar toplamının 360° olması için diğer iç açılardan da toplam 10° azalması lazım. O zaman da dış açılardan toplam 10° artacaktır.

Yani açılarla ne kadar oynarsak oynayalım dış açılar toplamı değişmeyecek, 360° olacak.

Öğrenci 5: Herhangi bir dörtgen çizelim. Açı ölçüleri a, b,c, d olarak isimlendirelim. İç açılar toplamı = a + b + c + d olur.

Dış açıların toplamı =

= (180°– a) + (180°– b) + (180°– c) + (180°– d)

= 180° + 180° + 180° + 180° – a – b – c – d

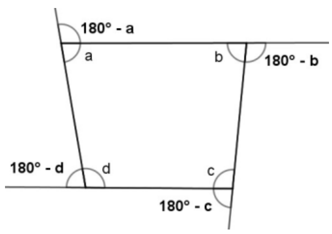
=720° – (a + b + c + d)

= 720° – (iç açılar toplamı)

Bir dörtgenin iç açılar toplamının 360° olduğunu biliyoruz;

= 720° – 360°

=360°



Üç öğrenciden aşağıdaki ifadenin doğru veya yanlış olduğunu nedenlerini açıklayarak göstermeleri istenmiştir.

"x, y, a, b sayıları birer doğal sayı, y $\neq 0$ ve b $\neq 0$ olmak üzere; olmak üzere;

Eğer $\frac{\mathbf{x}}{\mathbf{y}} = \frac{\mathbf{a}}{\mathbf{b}}$ eşitliği var ise her zaman "$\mathbf{x} \cdot \mathbf{b} = \mathbf{y} \cdot \mathbf{a}$" şeklinde bir eşitlik yazabiliriz.

<table>
<tr>
<td>Öğrenci 1: İçler dışlar çarpımı kuralına göre bu şekilde bir eşitlik varsa payları karşıdaki sayının paydasıyla çarpıp eşitleyebileceğimizi öğretmenimiz söylemişti.</td>
<td>Öğrenci 2:
$\frac{4}{10} = \frac{2}{5}$ => $4 \cdot 5 = 2 \cdot 10$
$\frac{7}{3} = \frac{21}{9}$ => $7 \cdot 9 = 21 \cdot 3$
$\frac{5}{8} = \frac{500}{800}$ => $5 \cdot 800 = 500 \cdot 8$
Yukarıda birden fazla örnekte denedim ve ifade hepsinde doğru çıktı. Yani $\frac{x}{y} = \frac{a}{b}$ ise "$x \cdot b = y \cdot a$" şeklinde eşitlik yazabiliriz.</td>
</tr>
<tr>
<td colspan="2">Öğrenci 3:
$\frac{x}{y} = \frac{a}{b}$ Yandaki bilgi verilmiş. Şimdi bu iki rasyonel sayının paydalarını eşitleyelim.
$\underset{(b)}{\frac{x}{y}} = \underset{(y)}{\frac{a}{b}} \rightarrow \frac{x \cdot b}{y \cdot b} = \frac{a \cdot y}{b \cdot y}$
İki kesrin paydası eşit $(y \cdot b = b \cdot y)$; dolayısıyla payları da eşit olmalı, yani $x \cdot b = y \cdot a$ olmalı.</td>
</tr>
</table>

1. Öğrencilerin sunmuş olduğu argümanları ikna edicilik yönünden 10 puan üzerinden puanlayınız. Puanlamanızı neden bu şekilde yaptığınızı açıklayınız.
   10 puan çok ikna edici buldum; 1 puan hiç ikna edici bulmadım anlamına gelmek üzere;

	İkna Edicilik Puanım	Neden Bu Puanı Verdim?
Öğrenci 1		
Öğrenci 2		
Öğrenci 3		

2. Hangi öğrenci veya öğrenciler verilen İfadenin *her zaman* işe yaradığını gösterebilmiştir? Nedeninizi açıklayınız.

3. Öğretmeniniz bu argümanları puanlasaydı kaçar puan vereceğini düşündüğünüzü belirtiniz. Öğretmeninizin neden bu puanı vereceğini düşündüğünüzü açıklayınız.
   10 en yüksek, 1 en düşük puan olmak üzere;

	Öğretmenimin vereceği puan	Öğretmenim Neden Bu Puanı Verirdi?
Öğrenci 1		
Öğrenci 2		
Öğrenci 3		

C. INTERVIEW PROTOCOL

1. Öğrencilerin sunmuş olduğu argümanların ikna ediciliğini puanlayacak olsaydınız 10 üzerinden;

- Öğrenci 1’e kaç puan verirdiniz? Neden?
- Öğrenci 2’ye kaç puan verirdiniz? Neden?
- ….
- Ne olsaydı puanınız yükselirdi? (Düşük puan verilen öğrenciler için sorulacak)

2. Kendi öğrencilerinizi göz önüne alarak, sizce öğrencileriniz bu argümanların ikna ediciliğine 10 üzerinden ortalama kaç puan vermiştir? Neden böyle düşündüğünüzü açıklayınız.

- Öğrenci 1’e kaç puan vermişlerdir? Neden?
- Öğrenci 2’ye kaç puan vermişlerdir? Neden?
- ….
- Öğrencileriniz sizce argümanların ikna ediciliğini puanlarken nelere dikkat etmişlerdir?

➢ Sayılar: Öğrencileriniz verilen argümanların ikna ediciliği için “Öğrenci 3 > Öğrenci 1 > Öğrenci 4 > Öğrenci 2” şeklinde bir puanlama yapmış. Sizce neden böyle bir sıralama yapmışlar?

➢ Geometri: Öğrencileriniz verilen argümanların ikna ediciliği için “Öğrenci 3 ≈ Öğrenci 5 > Öğrenci 4 > Öğrenci 2 > Öğrenci 1” şeklinde bir puanlama yapmış. Sizce neden böyle bir sıralama yapmışlar?

➢ Cebir: Öğrencileriniz verilen argümanların ikna ediciliği için “Öğrenci 2 > Öğrenci 3 > Öğrenci 1”şeklinde bir puanlama yapmış. Sizce neden böyle bir sıralama yapmışlar?

3. Hangi öğrenci veya öğrenciler verilen ifadenin her zaman doğru olduğunu gösterebilmiştir? Nedeninizi açıklayınız.

4. Öğrencilerinize de hangi argüman veya argümanların verilen ifadenin her zaman doğru olduğunu gösterebildiğini sormuştuk. Sizce gelen cevaplarda en çok hangi argüman seçilmiştir? En çok seçilenden en az seçilene doğru bir sıralama yapsanız nasıl sıralardınız? Nedeninizi açıklayınız.

- Öğrencileriniz sizce genel olarak argümanın her zaman doğru olduğunu gösterebilip gösterememesine nasıl karar veriyorlar? Nelere dikkat ediyorlar?
- Sayılar: Öğrencilerinize sorduğumuz "hangi öğrenci veya öğrencilerin verilen ifadenin her zaman doğru olduğunu gösterebilmiştir" soruna verdikleri cevaplarda en çok gelen yanıttan en az gelen yanıta doğru şöyle bir sıralama var: Öğrenci 3 > Öğrenci 1 > Öğrenci 2 > Öğrenci 4. Sizce böyle bir sıralama çıkmasının nedeni ne olabilir?
- Geometri: Öğrencilerinize sorduğumuz "hangi öğrenci veya öğrencilerin verilen ifadenin her zaman doğru olduğunu gösterebilmiştir" soruna verdikleri cevaplarda en çok gelen yanıttan en az gelen yanıta doğru şöyle bir sıralama var: Öğrenci 3 > Öğrenci 5 > Öğrenci 4 = Öğrenci 1 > Öğrenci 2. Sizce böyle bir sıralama çıkmasının nedeni ne olabilir?
- Cebir: Öğrencilerinize sorduğumuz "hangi öğrenci veya öğrencilerin verilen ifadenin her zaman doğru olduğunu gösterebilmiştir" soruna verdikleri cevaplarda en çok gelen yanıttan en az gelen yanıta doğru şöyle bir sıralama var: Öğrenci 2 > Öğrenci 1 = Öğrenci 3. Sizce böyle bir sıralama çıkmasının nedeni ne olabilir?

5. Öğrencilerinize sizin bu argümanlara kaçar puan verebileceğinizi sormuştuk. Sizce bu puanlamayı nasıl yaptığınızı düşünmüşlerdir? Nedeninizi açıklayınız.

- Öğrencileriniz bu puanlamayı yaparken sizin nelere dikkat ettiğinizi düşünüyor olabilirler?
- Sayılar: Öğrencileriniz sizin "Öğrenci 1 > Öğrenci 3 > Öğrenci 4 > Öğrenci 2" şeklinde bir puanlama yapacağınızı düşünmüşler. Sizce neden böyle bir sıralama yapmışlar?
- Geometri: Öğrencileriniz sizin "Öğrenci 3 ≈ Öğrenci 5 > Öğrenci 2 ≈ Öğrenci 4 > Öğrenci 1" şeklinde bir puanlama yapacağınızı düşünmüşler. Sizce neden böyle bir sıralama yapmışlar?
- Cebir: Öğrencileriniz sizin "Öğrenci 2 > Öğrenci 3 > Öğrenci 1" şeklinde bir puanlama yapacağınızı düşünmüşler. Sizce neden böyle bir sıralama yapmışlar?