How Access to Definitions Influences the Representations and Conceptual Insights Used in Student Arguments

A Dissertation
Presented in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
with a
Major in Mathematics
in the
College of Graduate Studies
University of Idaho
by
Annelise W. Nielsen

Approved by:
Major Professor: Robert Ely, Ph.D.
Committee Members: Julie Amador, Ph.D.; Chris Williams, Ph.D.; William Hall, Ph.D.
Department Administrator: Hirotachi Abo, Ph.D.

August 2023

Abstract

This study sought to explore whether access to definitions and general representations influences the construction of general direct arguments. Data was collected in college mathematics courses for prospective elementary school teachers. Participant arguments were analyzed along two variables: the generality of the representations and the viability of the conceptual insight they included. Participants were given one of three proving tasks. One task included no definitions, one task included definitions in the conceptual register, and the third task included definitions in the symbolic register. A randomized block design was used to explore the relationship between the definitions and the two variables. Qualitative methods were used to explore how participants intended their arguments. This study found that: 1) the inclusion of definitions on proving tasks does not have a substantial influence on the generality of the arguments or the viability of the conceptual insight used and 2) examples and algebraic representations were used as placeholders to demonstrate a procedure and to stand for the domain as a class of objects. The findings also indicate that the manner in which students are generalizing about the domain of the claim determines the structure of the domain they attend to, the conceptual insights available to them, and thus their ability to construct a viable argument. Future research is needed to connect the existing understanding of actions that support student generalization to supporting students in developing viable general direct proof.

Acknowledgments

Thanks to Rob Ely for his support and guidance finishing this project.

Dedication

I am so incredibly grateful to my family and friends for all their love and support. To my parents who are unwavering in their belief in me. To my friends who entertained my baby so I could write. To my sister, for her wisdom and humor. And to Dan, who kept the wheels on the bus that is our life.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

"This is trippy because I'm trying to find a way to prove it. There are infinite ways and it's messing with my head."

This was a statement made by Preston (pseudonym), a student asked to think aloud as he worked to develop an argument to the claim all terminating decimals are rational numbers. Preston used his understanding of place value to justify the truth of the claim for terminating decimals that end in the tens place, hundredths place, and thousands place individually and then he paused and was not sure how to proceed. He realized the domain of the claim[1] was infinite and was not sure how he could possibly address those "infinite decimal places."
In recent years, proof has been included in school mathematics classrooms across grade levels. The change in curriculum reflects the importance of proving activities to doing and understanding mathematics (Ball & Bass, 2000; Ball & Bass, 2003; Kitcher, 1984; Polya, 1981). Since proving activities are now more prominent in mathematics classrooms, research has sought to understand student experiences with proof and proving.

Fundamental to all explorations of student proving activities in response to a general claim is a focus on students' expression of generality. In mathematics, proving a general claim with an infinite domain with a direct proof requires some sort of description or representation of the domain of the claim that is general and can be operated upon using sequences of logical inferences that demonstrate that all objects in the domain of the claim have the properties of the conclusion of the claim. Key to developing a direct proof of a general claim is using representations and logical inferences that attend to all cases in the domain of the claim (Esty & Esty, 2009). In direct proofs, generality must be expressed in two ways: 1) in the representation of the domain of the claim and 2) in the logical inferences applied to the representations of the domain. Since generality of representations and logic is an essential requirement for general direct proof it is of interest to consider when students develop responses that are general and how they communicate generality.

[1] I define the domain of the claim to be the set of objects described by the conditions of the claim. In the case of Preston, the domain of the claim is the set of all terminating decimals.

It is well documented that students often develop empirical arguments in response to general claims (Coe & Ruthven, 1994; Goetting, 1995; Harel & Sowder, 1998; Healy & Hoyles, 2000; Sowder & Harel, 2003). These are non-general arguments that use a few examples as evidence that a general claim is true for a larger set. In much of the existing literature on proof, to understand student empiricism the theoretical assumption was made by researchers that what is taken by proof is equivalent to what the person finds convincing. Basically, expert mathematicians are convinced of the truth of a claim if there is a deductive proof that relies on logic and accepted prior results while students are convinced by a few conforming examples. In more recent literature, alternative explanations are emerging for why students may not create the deductive arguments the mathematical community craves. Among the more recent research is a study by Weber, Lew, and Mejía-Ramos (2020) which found that students and mathematicians consider multiple factors when deciding whether to produce a deductive or empirical argument. These factors include the perceived cost (which includes time) as well as their perceived likelihood of finding a deductive argument as factors when deciding whether to develop an empirical argument or a deductive one.

To Preston, not having a way to represent or describe all terminating decimals could be considered a cost as well as a perceived hurdle to his ability to produce a deductive argument. He continued to search for a way to objectify the set of all terminating decimals in a general manner but during the interview he was unable to find a way that allowed him access the structure of the domain that necessitates the conclusion of the claim. He finally settled on describing his approach to terminating decimals terminating in the tens, hundreds, and thousands place and saying “etcetera, etcetera” but was not able to identify why identifying the place value of the final digit in every terminating decimal was useful to demonstrating why the claim is true for all terminating decimals.

Preston’s search for a general representation highlights the importance of general representations in the development of deductive arguments. The general representation allows for all cases of the domain to be addressed and it gives access to describing and manipulating the set of objects for the purpose of demonstrating the conclusion is always true. This interaction led me to wonder how would Preston’s proof construction differ if there was a general representation available to him for the set of all terminating decimals? Would a representation have supported Preston to identify the mathematical structure of the set of all terminating decimals to write a general direct argument showing the truth of the claim? While there is a plethora of research determining the ways in which student proof differs from that of experts, a modest amount of research has been dedicated to identifying factors that influence student construction of general direct (Stylianides, Stylianides, & Weber, 2017). Existing studies have found that student skepticism of empirical data (Brown, 2014; Stylianides & Stylianides, 2009;), understanding of axiomatic systems (Jahnke & Wambach, 2013; Mariotti, 2000a, 2000b), as well as the practice of participating in self-explanation while writing proofs (Hodds, Alcock, & Inglis, 2014) improve student proving outcomes. Studies have yet to adequately study the role of definitions and general representations to mathematical objects in the claim influence argument construction. This study explores how access to definitions with general representations of objects in a mathematical claim influences proof writing by examining the generality of representations and viability of conceptual insights used across responses to similar tasks with variations in the definitions included. The research questions guiding the study are:

1. How does access to a general representation or general description of the mathematical objects in the claim influence the generality of a student's argument and the way the student represents the domain?
2. How does access to a general representation or general description of the mathematical objects in the claim influence the conceptual insights that are used in the argument?
3. How do students describe the representations they develop or choose to utilize in their arguments?

Chapter 2: Literature Review

2.1 Proof Construction and Validation

Proof has been centered in mathematics education as a fundamental part of learning and doing mathematics (Gokkurt, Soylu, & Sahin, 2014: Ingelis & Alcock, 2012; NCTM, 2000; Selden & Selden, 2003; Weber, 2008). Proving is viewed as an essential activity to deepen mathematical understanding (Hanna, 1990; Kitcher, 1984) and a necessary part of every student's mathematical education (Ball & Bass, 2003; Ball, Hoyles, Jahnke, & Movshovitz-Hadar, 2002; Hanna & Jahnke, 1996; Mariotti, 2006; NCTM, 2000; Stylianides, 2007, 2016; Yackle & Hanna, 2003).

There are many methods of proof. Some methods are only appropriate for certain types of claims. For example, consider the claim "there exists an integer x, such that x >3 and x <5." To address this claim, an existence proof is needed. An existence proof involves exhibiting a candidate and demonstrating that the candidate has the properties outlined in the claim (Esty & Esty, 2009). Other claims, called general claims, have some domain, often an infinitely large set, and assert that the set of objects shares some property or set of properties. For example, "all sums of three consecutive natural numbers are divisible by three." This claim asserts that every object in the domain shares the property "divisible by three" and as there are infinitely many sums of three consecutive natural numbers, the domain is an infinitely large set.

The proof methods used to demonstrate the truth of general claims fall into two classifications: direct proof and indirect proof. This study focuses on direct proof. For a general direct proof, the properties of the conditions are assumed. From the example above, I would assume the properties of sums of three consecutive natural numbers (Esty & Esty, 2009). Then prior results, definitions, and theorems are used to demonstrate that the properties of sums of three consecutive natural numbers necessitate that the sum is divisible by three. In contrast, indirect proof methods (contrapositive and contradiction) both begin with the negation of the conclusion and either demonstrate that this necessitates that the conditions are not true or leads to something that is false (Esty & Esty, 2009). In the case of the claim used in this study, students are unlikely to develop indirect proofs because they have likely not had instruction on how to develop these types of arguments and in the set "not divisible by three" is an unintuitive set to represent or describe in a general manner. Thus, the research described in my literature review is a subset of the proof research that is relevant to general direct proof and argument.

Two lines of inquiry have become prominent in the research on proof and proving: (1) student construction of proof and (2) student comprehension and validation of proof. Research has been drawn to these two lines of inquiry because students have persistently struggled with writing and reading proofs (Stylianides, Stylianides, & Weber, 2017). Proof construction and validation are connected because "one constructs a proof with an eye toward ultimately validating it" (Selden & Selden, 2003, p.6). Since validation and construction are connected results from studies exploring proof validation are relevant to my study. So while my study is situated within the proof construction, and I will primarily focus on the studies examining proof construction I will also touch on relevant findings from proof validation studies.

In explorations of how students read proofs and what they deem viable, researchers have found that students do not reliably make correct judgements about the viability of a proof. This has been found in studies of grade school students (Ahmadpour et al., 2019; Bieda & Lepak, 2014; Healy & Hoyles, 2000), undergraduate students (Alcock & Weber, 2005; Inglis & Alcock, 2012; Ko & Knuth, 2013; Selden & Selden, 2003; Weber, 2010), and even mathematicians have been found to disagree (Inglis & Alcock, 2012; Weber, 2010). In these studies, students were consistently focused on surface features such as the proof framework, or specific equations and manipulations of representations, rather than underlying mathematical structure. For example, Ahmadpour et al. (2019) found that students chose arguments with algebraic representations to be viable but when asked about the representation students did not connect the algebraic notation to the structure of the mathematical objects it represented.

Researchers have consistently shown that, across grade levels and through university, students are often not successful at writing proofs (Healy & Hoyles, 2000; Iannone & Inglis, 2010; Knuth, Choppin, & Bieda, 2009; Senk, 1989;). Researchers have hypothesized many sources for student difficulties including choosing an appropriate proof framework (Selden & Selden, 1995), making sense of the logical structures (Zandieh, Roh, & Knapp, 2014), lacking the means to communicate a proof (Mamona-Downs & Downs, 2009), mastery of the mathematical content (Azrou & Khelladi, 2019), understanding proof methods (Stylianides, Stylianides, & Pililppou, 2007), and transitioning from informal arguments that utilize diagrams and examples to form proofs (Pedemonte, 2007). Researchers have observed numerous ways that students fall short of the expectations of mathematicians when developing and validating proofs yet a very limited amount of research has explored methods of improving students' proof validation and construction (Stylianides et al. 2017).

Stylianides et al. (2017) summarize teaching intervention studies that were intended to explore ways of improving student proof outcomes. Within their synthesis of prominent intervention studies, progress has been made in promoting student skepticism of empirical evidence (Brown, 2014; Stylianides & Stylianides, 2009), understanding the axiomatic assumptions of a proof within the context of geometry (Jahnke & Wambach, 2013; Mariotti, 2000a, 2000b), and student understanding of induction (Brown, 2008; Harel, 2001; Ron & Dreyfus, 2004). Hodds, Alcock, and Inglis (2014) explored the influence of self-explanation training on student proof construction and validation. Self-explanation, explaining how the new information connects to what they know or the previous steps in the proof process, increased the quality of student-constructed proofs as well as student comprehension of written proofs. The intervention studies described above have demonstrated some factors that support student success in proof construction. My study contributes to the field by exploring factors that influence student construction of general direct arguments.

2.2 Proof Classification

When exploring the proofs students construct, it becomes apparent that there is a spectrum of types of responses that students develop that are not proof as is known and accepted by mathematicians. Classification schemes were developed to organize the types of arguments students develop based on varying criteria for the purpose of examining student progress towards deductive proof (Balacheff, 1988; Harel and Sowder, 1998). Balacheff's Hierarchy of Student Proofs and Harel and Sowder's Classification of Students' Proof Schemes hierarchy provide the foundation for how student proofs are discussed and classified.

Balacheff's Hierarchy of Student Proofs consists of four categories: Naïve Experience, the Crucial Experiment, the Generic Example, and the Thought Experiment (1988). Each category is distinguished by a different level of awareness and execution of generality. The first two categories include no generality. Naïve Empiricism classifies arguments that rely solely on examples. Either the student does not engage in the proving process and gives only conforming examples, or the argument suggests that a selection of conforming examples is sufficient to form a valid general proof. While the Crucial Experiment does not involve generality, it does mark a transition from pure empiricism because here students will seek specific examples that they see as extreme cases. By testing cases that are perceived by the student as extreme the student gains the belief that the claim is true for all cases (Knuth & Elliott, 1998). If the domain is integers, a Crucial Experiment may involve confirming that the claim holds for a really large number.

The final two categories include an awareness of the need for generality: the Generic Example and the Thought Experiment. In Balacheff's original description, Generic Example is defined as "making explicit the reasons for the truth of the conjecture by means of actions on an object which is not there in its own right, but as a characteristic representative of its class" (Balacheff, 1891, p.7). The Thought Experiment is then distinguished from the generic example as involving internalizing the reasoning and detaching the proof from a particular representation.

On first consideration it is not clear how a student can use an object as a representative of a class and not have internalized the properties of the class, leading to the necessity of the conclusion. Knuth and Elliot clarify this distinction by describing the Generic Example as relying on inductive reasoning, i.e., an observation is made based on empirical data and assumed to be true for a larger set. Hence the representation is representing a class of objects but the shared properties of those objects that are used in the transformation are derived from observation. In contrast the Thought Experiment uses deductive reasoning and thus the properties defining the class stem from the definition of the domain (Knuth & Elliot, 1998). When considering work produced by students, the distinction then between Generic Example and Thought Experiment may be as subtle as the student stating that a transformation can always be applied because everything in the domain has a certain observed structure versus connecting the structure to the definition of the objects.

Elements of Balacheff's hierarchy laid the groundwork for the classification of student proof. It is important to note that the definition of Generic Example is not consistent across the literature. The manner in which it is defined determines whether it is considered an to be part of a non-general proof (Balacheff, 1988; Leron & Zaslavsky, 2013) or a way for students to develop general proof (Yopp & Ely, 2016). Various definitions of Generic Example are described and compared in section 2.2.2.

While Balacheff's hierarchy distinguished student work by generality, Harel and Sowder classified student proof by the source of the student's conviction (1998). In their work, Harel and Sowder found that students draw conviction from a variety of sources, including empirical evidence, authority, or analytical systems. These sources of conviction then form the categories of a hierarchy, each with subcategories formed by distinctions found within the given source of conviction.

Harel and Sowder's taxonomy relies on the assumption that conviction and validity are intertwined—if an argument convinces the reader of the truth, then the argument is valid and vice versa. If students produce an empirical argument, they are considered to have used an empirical proof scheme which means they derive conviction from the section of confirming cases they include in their argument.

Generic examples and the notion that conviction and validity are intertwined are two aspects from these frameworks that are particularly relevant to my study and thus will be further explored in the following sections. Balcheff's hierarchy is useful as it identified features of student proving activity that is between empiricism and formal deductive proof. Primarily, the concept of Generic Example Proof is of interest and useful to me in this study as generic examples can be used by students to represent an infinite domain of a claim. The key assumption to Harel and Sowder's classification system—that the source of student conviction determines the type of proofs they create and choose, shaped the manner in which student empiricism was understood and researched. The findings from studies searching for answers to why students write empirical proofs motivated my study.

2.2.1 Generic Example Argument

Early definitions of generic example proof (sometimes called generic proof or generic example argument) include the following: 1) an example featuring a particular object in the domain, 2) a transformation or procedure applied to the object transforming the object to demonstrate that the conclusion holds, and 3) the property that the transformation or procedure can be applied to any object in the domain (e.g., Tall, 1979; Mason & Pimm, 1984; Balacheff, 1988). The generic examples and generic example proof are relevant to the discussion of student arguments and the classroom proving experience because they are often used in instruction, they have been found to have pedagogical benefits in the teaching of proof and are commonly seen in student arguments. In the literature, the definition of generic example is not consistent and thus there have been different theoretical interpretations of whether generic examples can be used to form a proof of viable argument.

One of the reasons generic examples are relevant to a discussion about proving is that they have been determined to be a valuable pedagogical tool when teaching proof. Researchers have encouraged educators to use generic examples as an argumentation method when teaching (Rowland, 1998; Russell et al., 2011). In his study, Tall (1979) found that students reported understanding the generic example argument better than the formal general proof. Tall's work suggested that generic proofs have more explanatory power for beginning students than general proofs. Generic examples have been found to have the power to convince and explain in all levels of mathematics (Rowland, 1998). Some researchers have suggested that generic example proofs are more accessible to learners as they do not require formal notation (Balacheff, 1988; Dreyfus et al., 2012; Stylianides, 2008). Others have concluded that generic examples can be a tool to express generality when formal mathematics notation may not be available (Harel & Sowder, 1996; Russell et al., 2011). As such, they have been viewed as a tool to bridge the gap between empirical arguments and formal arguments (Dreyfus et al., 2012; Leron & Zaslavsky, 2014).

In the literature, the definition of generic example and generic example proof has evolved over the last forty years. In the early definitions found in Tall (1979), Mason and Pimm (1984), and Balacheff (1988a) a common thread in the definitions is that the generality of the example is perceived by the reader but not directly stated by the author of the argument.

> [The proof] is generic in the sense that it contains within it a complete spectrum of proofs for all square roots of non-squares. It shows why no square of a rational equals 2 and the same proof readily adapts to 5/8 or any other non-square (Tall, 1979, p. 5).

> The generic proof, although given in terms of a particular number, nowhere relies on any specific properties of that number (Mason & Pimm, 1984, p. 284).

> The generic example involves making explicit the reasons for the truth of an assertion by means of operations or transformations on an object that is not there in its own right, but as a characteristic representative of the class. The transparent presentation of the example is such that analogy with other instances is readily achieved, and their truth is thereby made manifest. Ultimately the audience can conceive of no possible instance in which the analogy could not be achieved (Balacheff, 1988, p. 219).

In these early descriptions and examples of generic examples and generic proof there is no criteria that works to include an explicit statement addressing how the example is representative of its class or an explanation of why the reasoning used on the example can be applied to all objects of its type.

Without the requirement to be explicit about the generality and reasoning at hand it is not surprising that there is dispute in the literature about whether a generic example argument can be taken as proof. In early classifications of student argument, generic example proofs are not considered to be proof. Harel and Sowder (1996) assert that generic example proofs reflect a “students’ inability to express their justification in general terms” (p. 43). Similarly, in Balacheff’s classification of student work, generic proof was not considered full or proper proof. This perception persisted and is present in Leron and Zaslavsky’s (2013) reflection on generic example and its role in pedagogy and proof. In contrast, Rowland (1998) disagreed with what he referred to as the deficit view of generic proof taken by researchers, stating in some cases the generic example “adequately bears the intended generality and is fully sufficient for purposes of conviction and explanation” (p. 70). At the heart of these two opinions is a difference in opinion about who is responsible for seeing the generality of the example—author or reader. Those who say that generic proof is not proof using early definitions seem to be saying that the author is responsible for communicating why their example is general. Those who say generic example proof is proof are saying that if the reader can perceive the generality of the example, then the generic proof is proof. However, if the goal is assessment of student understanding, a generic example presented as defined in this manner leads to challenges because how would the teacher or researcher know whether the example was intended as generic?

Kemper and Biehler (2015) present a solutsion to this challenge by emphasizing that the inclusion of narrative reasoning is what makes the generic proof a valid general argument. They do not explicitly define “narrative reasoning” but in their example the narrative reasoning describes the structure of the example that is general to all objects of the domain and why that structure allows for the procedure or transformation that is completed to demonstrate the truth of the claim.

$1 + 2 + 3 = (2 - 1) + 2 + (2 + 1) = 3 \times 2$

$4 + 5 + 6 = (5 - 1) + 5 + (5 + 1) = 3 \times 5$

You can always write the sum of three consecutive numbers as: (“number in the middle“-1) + “number in the middle” + (“number in the middle“+1). Since this sum equals three times the „number in the middle“, the sum is always divisible by three.

Figure 1: Example of a generic example in a valid argument according to Kempen and Biehlers’ criteria (2015, p. 136-137)

In their handling of generic example, Yopp and Ely (2016) state, "…in a generic example argument, the generality lies not in the representation but in the way the example is appealed to." In contrast, "formal proof often uses general representations, such as quantified variables or symbolic placeholders" (p.41). Yopp and Ely present a framework for evaluating whether an example used in an argument is a generic example and thus they argue can be part of a viable argument. Their framework requires that at each step in the example the argument appeals to a general property of the example. In their paper, they highlight that in the previous literature the generality and hence viability of a generic example argument has been determined by the readers' assumptions and that this does not allow for researchers and teachers to assess student work without making assumptions that may mischaracterize the student's intended reasoning.

Reid and Vallejo Vargas sought to navigate the challenge of sorting authors' intent from readers' assumptions by defining generic proof in terms of two factors: psychology of the readers and the social conventions of the context.

> Psychologically, for a generic argument to be a proof it must result in a convincing deductive reasoning process occurring in the mind of the reader. Socially, for a generic argument to be a proof it must conform to the social conventions of the context (Reid & Vallejo Vargas, 2018, p. 239).

They acknowledge that determining the psychology of a student can be challenging and suggest sharing two criteria with students to support students in writing arguments that give the teacher sufficient evidence to determine whether the example is used by the student generically or empirically. The criteria are: 1) evidence of awareness of generality and 2) mathematical evidence of reasoning (p. 247).

Reid and Vallejo Vargas (2010) interpret phrases such as "the same reasoning can be used for the other cases", "it also applies to the other cases involved", or "this is true for all [objects in the domain]" as evidence of awareness of generality (p. 247). Reid and Vallejo Vargas state the "main reason for considering this as relevant evidence is the need to be sure whether or not the students are aware that they are not only dealing with empirical evidence, but that their work shows general structures through the use of their examples" (p.247). They explain "[mathematical evidence of reasoning] mainly points to the mathematical reasons for why the same structure can be extrapolated for other cases from the example(s) given, and it is based not only on the conditions of the problem given but also on the ground knowledge the community shares at that point (the social aspect)" (p.247). Figure 2 shows an example they give of a generic example proof satisfying their criteria. The proof is in response to the statement "Prove that the sum of the first *n* natural numbers is $\frac{n(n+1)}{2}$."

Consider the sum $1+2+3+4+5+6+7+8+9+10$. Write this sum, and the reverse, and add them:

$$\begin{array}{l} 1+2+3+4+5+6+7+8+9+10 \\ \underline{10+9+8+7+6+5+4+3+2+1} \\ 11+11+11+11+11+11+11+11+11+11 = 10 \times 11 \end{array}$$

Adding down always gives 11 (one more than the last term in the sum) because in the first case we are adding $10+1$ (the last term plus 1), and then we are adding a number that is one more (1 becomes 2) to a number that is one less (10 becomes 9). The numbers are consecutive, so the increase in the top row is the same as the decrease in the bottom row. There is one sum adding down for every number in the top row, which is 10 in this case. So we multiply the highest number in the sum (10) by one more than the highest number (11). The product is two times bigger than it should be, because we added $(1+2+\cdots+9+10)$ twice, so to find the real sum we divide the product (10×11) by 2. The same reasoning can be used for any natural number n, and not only for the case of 10.

Figure 2: Example of a generic example proof given by Reid and Vallejo Vargas (2018, p. 248-249)

The example given by Reid and Vallejo Vargas (2018) above includes a narrative reasoning as described by Kempen and Biehler (2015). Reid and Vallejo Vargas proclaim that there are other ways besides the use of narratives for students to communicate generality. They suggest that students who struggle with linguistic formulations may be able to express generality in alternative ways such as symbols, arrows, and other indications to the general structure of the example with some statement saying the procedure or transformation completed can be done with all objects in the domain. Figure 3 shows such an argument. They claim that the argument is proof because the symbols express the same generality that was captured in the written words in the prior argument.

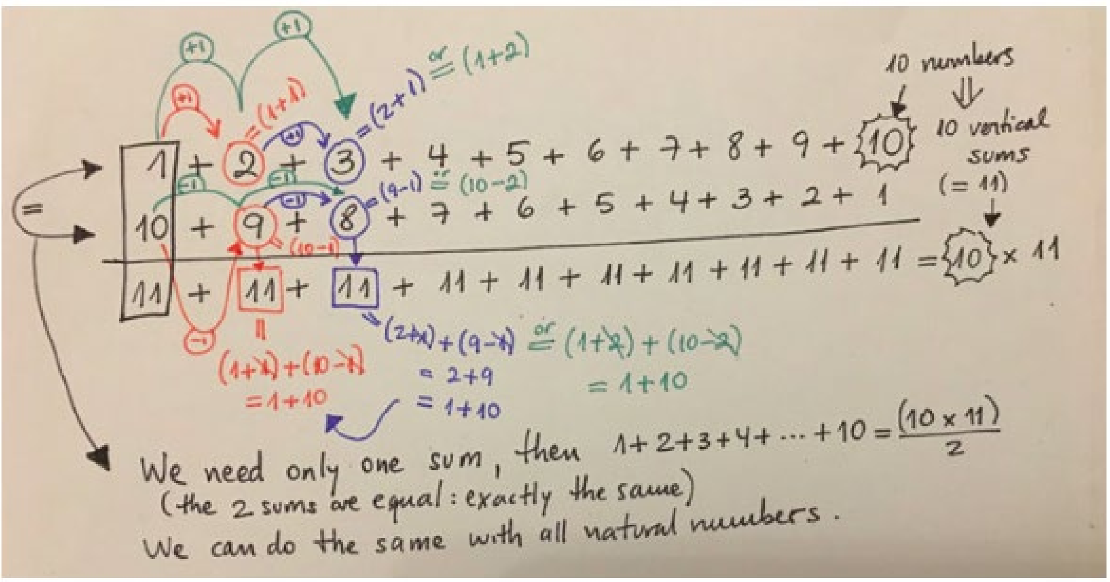


Figure 3: Example of a generic example proof given by Reid and Vallejo Vargas that does not include a narrative (2018, p. 249)

The definition given by Reid and Vallejo Vargas has since been used and adapted by researchers to classify generic examples in studies exploring characteristics of arguments written using generic example proofs (Rø & Arnesen, 2020) and the perceptions of generic example proofs by in-service teachers (Dogan & Williams-Pierce, 2021). Rø and Areneson used the definition given by Reid and Vallejo Vargas and explored the arguments developed by student teachers when asked to write a generic example proof. They found that none of the teachers wrote arguments that fully satisfied their criteria for generic example proof and thus suggest that generic example proof has an opaque nature requiring attending to criteria and emphasizing the structural nature of generic examples. Dogan and Williams-Pierce also utilized the definition by Reid and Vallejo Vargas in tandem with Stylianides sociocultural definition of proof:

> …a mathematical argument, a connected sequence of assertions for or against a mathematical claim, with the following characteristic:
>
> 1) It uses statements accepted by the classroom community that are true and available without further justification.
>
> 2) It employs forms of reasoning that are valid and known to, or within the conceptual reach of, the classroom community.
>
> 3) It is communicated with forms of expression that are appropriate and known to, or within the conceptual reach of, the classroom community (Stylianides, 2007, p. 291).

In their analysis, they found three categories of proof using generic examples: 1) empirical arguments enhanced with generic language, 2) incomplete generic examples, and 3) complete generic arguments (p. 133). The first category is characterized by examples accompanied by what Rowland (1998) defined as empirical generalization—a generalization "derived from the form of results (usually numerical) and observed relationships" (p. 67). Rowland elaborates, stating "empirical generalizations may possess predictive potential but lack explanatory power" (p.67). This can be seen in the examples given by Dogan and Williams-Pierce of empirical arguments enhanced with generic language. In these examples, the student teachers consider several examples and then notice patterns in the results which they describe generically. Incomplete generic examples are characterized as satisfying the awareness of generality criteria but not including sufficient mathematical reasoning. In the example given by Dogan and Williams-Pierce, the in-service teacher generalizes their example but relies on example-based reasoning to test her generalization. The in-service teacher does not provide valid justification for why the generalization will work for all objects in the domain. In their study they found that the Reid and Vallejo Vargas criteria for generic example proof was well suited for analyzing classroom proving activity in combination with the Stylianides (2007) framework.

The definitions and interpretations of generic example have evolved since the conception of the idea in the seventies from an example that a reader can perceive as representative of generality to an example that includes work showing awareness of generality and mathematical reasoning. Within the recent body of research, the criteria given by Reid and Vallejo Vargas have allowed for researchers to identify generic examples in their analysis and classify generic arguments as proof when they meet the outlined criteria and even find nuanced categories of arguments when one of the criteria is met and the other is not. What has not been explored is whether the definitions or representations students have access to influences their decision to use generic examples as opposed to empirical examples, or more formally accepted algebraic notation.

2.2.2 Empiricism

To understand the motivation for my study and my focus on general representations, it is important to discuss the prevalence of empiricism in both proof construction and proof validation. Research exploring the work produced by students in response to proof prompts has found that students often respond with empirical work (Chazan, 1993; Harel & Sowder, 1998; Healy & Hoyles, 2000; Knuth et al., 2002; Martin & Harel, 1989; Simon & Blume, 1996; Stylianides & Stylianides, 2009). This means that when given a statement that is general, or referring to a set of objects, students demonstrate the truth of the claim for some finite subset of the domain. For example, when given the claim "the sum of three consecutive natural numbers is divisible by three," a student may develop the following argument:

2+3+4=9, 9/3=3. I showed it worked so the claim is true.

This argument uses one example in the domain to demonstrate that the claim is true for all sums of three consecutive natural numbers.

Empiricism has been of interest to mathematics education researchers because it has been found to be incredibly prevalent in student work and thinking. Students across grade levels and ages have been found to develop empirical responses (Chazan, 1993; Harel & Sowder, 1998; Healy & Hoyles, 2000; Knuth et al., 2002; Martin & Harel, 1989; Simon &

Blume, 1996; A. J. Stylianides & G. J. Stylianides, 2009). Knuth et al. found in their study that 70% of the middle school students participating used examples to justify the truth for infinite domains. Goetting (1995) found 80% of participating preservice teachers accepted empirical arguments and more recently Morris (2007) found 41% of preservice teachers accepted empirical arguments.

These findings are troublesome to mathematics education researchers because empirical proof is not in alignment with what mathematicians view as proof and is perceived by many in the mathematical community as evidence of unsophisticated or immature mathematical thinking (Balacheff, 1988; Harel & Sowder, 1998). Thus, mathematics education research has been interested in understanding why students develop empirical arguments and exploring how to shift student proofs from empirical proof to analytical proof.

To make sense of student empiricism many researchers ascribe to the theoretical assumption that student proof is indicative of the student's scheme for what is convincing (Weber et al., 2020). This theoretical assumption, referred to by Weber et al. as the "proof as convincing" paradigm, means that an empirical argument is interpreted as evidence that the writer believes that their empirical work is convincing and sufficient to prove the claim at hand (2020). The writer of the proof is then described as having an empirical proof scheme (Harel and Sowder, 1998). Conversely, a deductive proof is then considered evidence that a student is convinced by deductive reasoning. Weber et al. (2020) summarizes this body of research categorizing the studies into two different categories: 1) justification studies where researchers ask a group to justify mathematical statements and examine what is produced and 2) evaluation studies where researchers present a claim and justification for the claim to a participant and ask the participant whether they are convinced that the claim is true by the justification. In both types of studies, when participants either develop an empirical proof or choose an empirical proof as convincing, the participants are classified as having an empirical proof scheme. The "proof as convincing" paradigm dictates how researchers propose supporting students to produce deductive proofs as their theoretical framing assumes that if students learn that empiricism is not convincing in mathematics and that deductive reasoning is convincing, then they will develop deductive arguments.

In recent research the "proof as convincing" paradigm has been challenged by findings that suggest that empirical proofs may be developed independent of an empirical proof scheme. Findings that challenge the paradigm include that students have different criteria for viable proofs than for convincing proofs (Knuth, 2002), the percentage of empirical responses is dependent on the difficulty of the claim (Knuth et al., 2009), participants that develop empirical arguments do not always consider these arguments to be convincing nor do they necessarily believe they are valid (Healy & Hoyles, 2000; Stylianides & Stylianides, 2009; Weber, 2010), participants reported not gaining certainty from empirical proofs (Bieda & Lepak, 2014), and participants apply different criteria or use different conceptions of proof depending on the context in which the proof is presented or developed (Healy & Hoyles, 2000; Stylianides & Al-Murani, 2010). These findings were detailed by Weber et. al (2020) and motivated their exploration of an expectancy value model to explain student empiricism.

Weber et al. (2020) used expectancy value theories to account for the types of proofs students develop. Expectancy value theories are theories of motivation that study the relationship of beliefs, values, and goals with actions (Eccles & Wigfield, 2002). Using their study, Weber et al. investigated three alternative factors that may dictate whether participants develop empirical justifications:

> (a) Participants might not be interested in being certain of the conjecture in question and thus settle for the first empirical justification that they produce (an issue of values); (b) they might find searching for a proof to be an unpleasant endeavor that is not worth the effort and thus prefer an empirical justification that takes less time and effort to generate (an issue of cost); or (c) they might settle for the empirical justification because they believe that they lack the capacity to find a proof (an issue of likelihood of success) (Weber et al., 2020, p. 32).

These factors demonstrated to be relevant to the participants in their study and furthermore reflect the practices of mathematicians. The work by Weber et al. and the prior research that challenges the "proof as convincing" paradigm demonstrates that there is more to discover about why students generate empirical justifications for the purpose of supporting their development of deductive arguments. In particular, how does a participant's familiarity with the domain and ability to represent it in a general manner contribute to their perceived cost of developing an argument or their perceived likeliness of success?

2.3 Components of Deductive Proofs

In section 2.2 I summarize how the spectrum of arguments produced by students has been classified. The classification of proofs affords ways of sanctioning student general direct arguments that are "proof-like"—including general and valid representations and reasoning which are not necessarily presented in a manner that is accepted as formal mathematical proof. These arguments are called "proofs" by some researchers (Balacheff, 1988a; Russell et al., 2011; Stylianides, 2007), but are sometimes called "deductive arguments" (Chazan, 2009; Knuth, Choppin & Bieda, 2009; Morris, 2009) or "viable arguments," (Yopp & Ely, 2016). In this section I summarize a model for arguments and features of arguments that influence the construction of viable arguments to general claims.

2.3.1 Models for Argumentation

Toulmin's 1958 model for argumentation was developed to examine moral reasoning. The model was adopted and used by mathematics education researchers to analyze components of student proof and argument. The model has three base structures: 1) a claim, assertion, or opinion, 2) data produced to support eh claim, and 3) a warrant that serves as justification for why the data supports the claim. In mathematics the warrant is often an axiom, theorem, or definition. The model has 3 additional elements. *Qualifiers* which acknowledge limitations of the claim, *rebuttals* which indicate circumstances where the warrant will not hold, and *backing* which is the theoretical framework that supports the warrant.

The model for argumentation has been used to analyze and document proof in various contexts. It has been used to analyze the process of students learning proof in classrooms (Forman, Larreamendy-Joerns, Stein, & Brown, 1998; Moore-Russo, Conner, & Rugg, 2011; Krummheuer, 1995; Yackel, 2001; Yackel & Rasmussen, 2002; Stephan & Rasmussen, 2002), student and teacher interview data (Hollebrands, Conner, & Smith, 2010; Inglis, Mejia-Ramos, & Simpson, 2007; Nardi, Biza, & Zachariades, 2012; Steele, 2005), the nature of student arguments under certain conditions (Hollebrands et al., 2010), how student and teachers connect parts of their arguments (Gonzales & Herbst, 2013), and to consider the relationship between argumentation and proof (Knipping, 2008; Lavy, 2006; Pedemonte, 2007, 2008; Weber & Alcock, 2005).

The model is not specific to mathematical argumentation and such researchers have found some limitations of using the model for understanding student proving activity. In particular, warrants as defined by Toulmin seem to have distinct meanings. Stranieri and Zeleznikow (1999) found that warrants can be a reason for a fact or a rule that leads to an inference. This was then extended by Pedemonte and Balcheff (2016) into three distinct meanings of a warrant: 1) a method of inference, 2) reasons that explain how the data demonstrates the truth of the claim, and 3) reasons that explain why the method of inference is appropriate.

To better account for warrants, Pedemonte and Balacheff (2016) used Toulmin's model in conjunction with the ck¢ Model, a model of a learner conception. The ck¢ Model was developed by Pedemonte and Balacheff and has four elements. The first is the *sphere of practice* which is the "set of problem-situations within which [a conception] proves to be an efficient tool for building a solution" (p. 109). The second and third elements are the *operators* and *representations* used within the sphere of practice. The fourth element is the *control structure*, "the set of means learners have" to make decisions and to assess to take decisions, and to assess their production (p. 105). The ck¢ Model affords more nuance when considering the warrants and backings in mathematical arguments. Pedemonte and Balacheff described how the model appears in proof construction. They stated,

> a given conception a mathematical problem can be represented by a set of statements expressed using the representation system. Hence, the application of a rule transforms an initial set of statements data into a new claim. The series of transformation ends when it reaches an ultimate statement claimed "true" based on the control structure (p. 108).

This characterization of an argument includes a reoccurring theme that the original claim, or domain of the claim, is transformed, operated on, and then reinterpreted. This characterization appears in the work of Weiss and Herbst (2007) and Duval (2006) in their discussions of registers and representations as well as the model for definition use outlined by Esty and Esty (2009).

Pedemonte (2007, 2008, 2016) utilized Toulmin's model and the ck¢ Model of a learner conception to explore the challenges students experience in proof writing by analyzing the transition from informal explorations to formal proof. Among the findings of her studies were the following: the validity of the operator influenced whether students successfully developed proof, abductive and inductive reasoning with empirical data at times led to larger structural distance between the informal argument and the proof, and generalization about the process observed in empirical data was useful to students to transition to proof.

When students used operators that were valid in their informal arguments, they were able to directly replace the operator with a theorem (Pedemonte, 2005). When the operator was not mathematically valid, two possible outcomes resulted: 1) the proof was not constructed because the student could not replace the operator with a theorem or 2) an incorrect proof was constructed based in the conceptions of the informal argument. Pedemonte and Balacheff found instances where the operator was not valid, correlating with the use of empirical data in tandem with abductive and inductive reasoning. What was useful for students in their transition was generalizing the process. This finding fits with the findings of Harel's (2001) teaching intervention where he found that generalizing the result of a pattern resulted in empirical proofs and generalizing the process of a pattern resulted in more deductive proofs.

An additional finding from Pedemonte and Balacheff's study was that interpreting the role that a representation is playing within a student proof is difficult; "it can be read as the actual mathematical object or as its representation" (p. 121). This supports the need to examine how students are intending to use the representations they develop or choose in a given argument.

2.3.2 Registers and Representations

The notion of representation has various meaning attributed to it in literature on teaching and learning mathematics (Zazkis & Liljedahl, 2004). Researchers and educators have struggled to define representation as it is considered both a dynamic process tied to an individual's mathematical thoughts (Vergnaud, 1998), a product (NCTM, 2000), and is further complicated by researchers viewing representations as existing internally and externally (Goldin & Shteingold, 2001; Goldin, 2003; Zhang, 1997). Internal representations describe mental processes (Goldin, 2001) or arrangements of ideas in an individual's mind (Janvier, 1987). Internal representations are not observable and difficult to study (Haciomeroglu, Aspinwall & Persmeg, 2010). External representations are forms of communication of mathematical ideas such as diagrams, signs, figures, characters, symbols, etc. (Mainali, 2021). Stylianou defined them as "configuration[s] that stand for something else" (2011, p. 266). This study is exploring how the inclusion of definitions which include symbols and characters to stand for classes of objects influence the construction of student argument. Hence, I will focus on external representations and refer to them as representations.

Representations are a fundamental part of communicating proof as seen by their inclusion in the model for learner conception. Pedemonte and Balacheff stated, "for researchers as well as for teachers making sense of students' understanding and activity always starts from the evidence provided by representations" (p.121). Frameworks leveraging registers and representations have been used to understand what it takes to prove. Mathematical objects can only be accessed through their representations and transformations that are made on those representations (Duval, 2006). Duval defined representational register as a semiotic system in which transformations can be completed on the representations. He outlined two types of transformations critical to mathematical activity.

Treatments are transformations that remain within the same register. For example, simplifying an expression or visually reconfiguring a diagram of a parallelogram into a rectangle of equivalent area. Conversions are transformations that "change the register without changing the objects" (p. 112). For example, converting a written description of a relationship into algebraic notation, or converting algebraic notation into a graphical representation. Duval stated that these types of transformations are more complex than treatments because "any change of register first requires recognition of the same represented object between two representations whose contents have very often nothing in common" (p.

112). Interpreting this within the notion of structure,[2] this seems to mean that the forms have little in common and recognition of the same represented object means identifying the structure of the object within the different form.

Duval (2006) used his framework to identify conversions as a source of difficulty for students. He found that students not successfully converting a relationship from a Cartesian graph representation to an equation representation did not correspond with the students not understanding the concept of function. Conversions are challenging to students and from Pedemonte and Balcheff's (2016) model they appear fundamental to argument and proof.

Weiss and Herbst (2007) used Duval's notion of conversions to examine proving activity in geometry classrooms. They defined three distinct registers: conceptual, generic, and diagrammatic. The conceptual register describes when objects are referred to by the name of their abstract class. The generic register describes when "particular instances of the class are taken as generic representations" (p. 2). And the diagrammatic register describes when a diagram is used to represent a class of objects. Weiss and Herbst found that students had very few opportunities to complete conversions between the registers. Most conversions were completed by the teacher or by the textbook. In particular, they noted that when students were presented theorems, the conceptual register was used and when students were asked to prove, they were presented representations in either the generic or diagrammatic register.

Duval highlights the importance of representations in mathematics. General, direct proofs are about infinite sets of abstract mathematical objects. Can we expect students to be successful at communicating their proof without access to a general representation of the domain? Furthermore, does the presence of a definition support students converting into that representation system?

2.3.3 Mathematical Structure

In observing how a student use representations and the warrants they develop I found there was a need to identify the properties of the mathematical objects and the ways in which

---

[2] Definitions of structure vary across the literature. In this study I use structure to describe the characterization of properties of a mathematical object (see the theoretical framework for more detail).

those properties are seen and characterized by student. The term “structure” has been used in various contexts across mathematics education research. Yet there is a certain degree of vagueness about its meaning as it has been used to describe many different phenomena. Kieran noted in 2018 that:

> Structure is often treated within the mathematics education community as if it were tantamount to an undefined term; it is further assumed that there is universal agreement on its meaning (p. 80).

In existing literature there are several distinct ways that structure is defined and the connections between structure and generalization, abstraction, representation, and properties are neither consistent nor clear (Venkat, Askew, Watson, & Mason, 2019). In the existing literature there are four distinct ways that structure has been defined: 1) the class or syntax of a problem, 2) the relationship between features within an object, 3) the external interrelationship between objects, and 4) as one of the two ways to conceive of mathematical objects.

Class or Syntax of a Problem

Hoch and Dreyfus (2004) used structure to refer to the class of problems that can be solved similarly. The use of the word structure to identify syntactic components of math tasks is not uncommon and also appears in Vergnaud (1992) along with other studies. This view of structure has been used to examine the relationship between the syntax of a problem and student response. This definition of structure does not address the structure of the mathematical objects themselves.

Features within an object

Malle (1993) considered recognizing structure as identifying partial terms. For instance, Malle identified three different partial structures within the equation $4 * x + 3 = 11$.

(1) $\boxed{4 \cdot x \quad + \quad 3} = 11$

(2) $\boxed{4 \cdot x} + \boxed{3} = 11$

(3) $\boxed{4} \cdot \boxed{x} + \boxed{3} = 11$

Figure 4: Depiction of structure from Malle (1993, p. 189)

This definition of structure was used to investigate unpacking features within algebraic equations and applying order of operations.

Within the specific content of arithmetic and algebra, Banerjee and Subramaniam (2012) defined structure as “identifying the components of expressions which contribute to its value and which remain invariant through valid transformations " (p. 356). The instruction with a focus on structure involved stating the information contained in the expression, the units that compose the expression, the values of each of the units and discussing the transformations that will not alter the value of the expression. They stated that perceiving expressions using a structural way allows students to develop a deeper understanding of the equality of expressions (Banerjee & Subramaniam, 2012). Similarly, Kieran (2018) saw structure in activities that decompose and recompose arithmetical or algebraic representations. These definitions focus on relationships between internal elements of mathematical objects or representations.

Kieran (1989) defined two types of structure within the context of algebra—surface (internal) and systemic (external) structure. Surface structure is similar to the definition used by Banerjee and Subramaniam (2012). As defined by Kieran (1989), surface structure “describes the relationship between the partial terms of the term in accordance with the hierarchy of operations, as well as the equality of two terms on the left and right side of an equal sign” (Kieran, 1989, p. 387). System structure “comprises all equivalent forms into which a term can be transformed in accordance with mathematical laws, or rather, all equations that are equivalent to an equation” (Kieran, 1989, p. 387). With these definitions, Kieran distinguishes between the structures found within an object and external structure found by making a comparison to equivalent objects.

Relationships Between Objects

The third category of definitions are those that view structure as the relationships between objects, i.e., external relationships. Morris (1999) defined mathematical structure to be knowledge about mathematical objects themselves and the knowledge of the relationships between the objects and the properties of the object. Warren (2003) expanded this definition into four parts: 1) relationship between quantities, 2) group properties of operations, 3) relationships between the operations, and 4) relationships across the quantities (Warren 2003, p. 123). Simpson and Stehlíková (2006) define apprehending structure as "the shift of attention from the familiarity and specificity of objects and operations to the sense of interrelationships between the objects caused by the operations" (p. 532).

Mason, Stephens and Watson (2009) define mathematical structure as "the identification of general properties which are instantiated in particular situations as relationships between elements or subsets of elements of a set" (Mason 2009, et. al, p.10). This definition makes it unclear if transformations of an object due to the interpretation of a definition, for instance $2^2 = 2 * 2$, are considered to reveal structure. If the object is the expression, then it can be considered the instantiation of the exponent definition but if the object is 2^2 then the object is unchanged, and the expression is merely an equivalence relationship.

Kieran's systemic definition and the definitions from Morris (1999), Simpson and Stehlíková (2006), and Mason, Stephens, and Watson (2009) all share a focus on interrelationships between objects. The definitions given by Morris (1999) and Warren (2003) are less specific. Kieran's (1989) definition is distinct in that she categorizes the equivalent forms as structure; in contrast Mason, Stephens, and Watson (2009) name the "identification of general properties" as structure (p.10).

Rüede (2013) expands on the idea of external structure and defines personal structure within algebra to be a person's individual interpretation of an expression. The personal structure identified will impact which systemic structures, as defined by Kieran (1989), are used as a person interacts with an expression or an equation.

Structure as A Way of Conception

Sfard (1992) discusses what she refers to as the “dual nature of mathematical conceptions”. The dual nature includes a coexisting operational and structural conception of mathematical objects. She defines operational conception to be viewing a notion as a process and structural conceptions as viewing a notion as an object (Sfard, 1992, p.60). Sfard reports a “deep ontological gap between operational and structural conceptions” (Sfard, 1991, p.4). This gap is partly attributed to the nature of the two conceptions, mainly that structural conceptions develop after operational conceptions through reifications (Sfard & Linchevski, 1994, p. 191). She characterizes reification as “an instantaneous quantum leap: a process solidifies into an object, into a static structure.” She clarifies that multiple representations are unified, the object is “detached from the process which produced it” and it becomes defined by the abstract category to which it belongs (Sfard, 1991, p. 20). To better support the development of structural thinking Sfard lists four factors to consider:

1. Supporting students to gain understanding of the processes that underlie mathematical concepts
2. Proficiency with executing algorithms
3. An adequate representation and exposure to many kinds of representations
4. Open discussion on the nature of mathematical entities and the differences between processes and objects (Sfard, 1991, p. 78)

Sfard’s duality framework differs from the other definitions of structure in that she is describing, in general, the two ways of conceiving of all mathematical objects. Sfard’s work also differs as it suggests a progression for development of structural thinking.

Path of Structure to Reading Proof

The model developed by Ahmadpour et al. (2019) describing how students read proofs identified structure as a defining component of one of three pathways to understanding proof when reading. Ahmadpour et al.’s model contains three pathways to characterize formally accepted proof for students reading a proof: path of structure, path of procedure, and path of form. They define formally accepted proof as a proof presented in a form that is acceptable to a given community.

In their study they examined how students made sense of different arguments to the claim “all sums of three consecutive natural numbers are divisible by three.” The path of form is characterized by students focusing on the symbols and surface-level characteristics of a proof without an understanding of the argument. Ahmadpour et al. examined how students read arguments, thus the path of form was evidenced by the reader considering a proof to be viable because it had familiar surface-level features that they recognize from class. However, they do not attribute any deeper understanding to the notation. Ahmadpour et al. include as an example a student who selected an argument that used algebra to demonstrate the truth of the claim, i.e., $n+n+1+n+2 = 3n+3=3(n+1)$. When asked what n is, the student responded it was an unknown. When asked the meaning of adding n, n+1, n+2, the student responded “now we should calculate them.” This demonstrated that the student did not have an understanding of how the algebraic representation captures the structure of sums of three consecutive numbers. When writing arguments, the phenomenon of the path of form would be captured by the Formal Mimicry Scheme detailed by Harel and Sowder (1998).

The remaining two pathways, the path of structure and the path of procedure, share two states: Naïve Experience and General Procedure. Naïve Experience is defined as students considering examples sufficient to justify a general claim. Students in the General Procedure state identify a procedure that can be applied to all cases to show the result, but they do not identify why the procedure can be generalized. As an example of this state, Ahmadpour et al. use the following interview excerpt:

> If we separate them into threes, for example if 13 is separated into threes, it makes 12; If 14 is separated into threes, it makes 12 again. Here is 1 [left over], and here is 2 [left over], we add 3 to it. 15 itself is divisible by 3 (p. 8).

In the excerpt the student uses an example to demonstrate how to divide each addend in a sum of three consecutive numbers by three and sum the remainders to show that the claim is true for that example. At this state there is no evidence of the student understanding why the procedure can be applied to all cases in the domain or any case in the domain. The two paths split from the General Procedure state to form the two distinct paths.

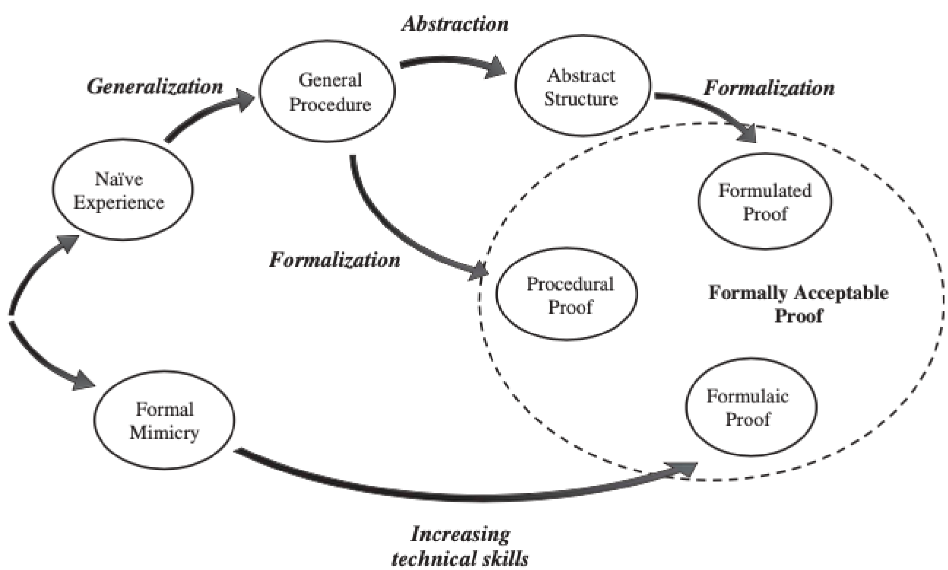


Figure 5: Model for how students understand proof while reading (Ahmadpour et al., 2019, p. 3)

For students who follow the path of procedure to read a proof as a Procedural Proof, the transition from the *General Procedural* state to the *Procedural Proof* state involves connecting surface-level symbols with a procedure, which the authors described as a formalization. Students on the path of procedure read notation that is designed to represent a procedure for transforming the domain as a recipe for inputting cases in the domain, following steps, and then confirming that the conclusion holds. Ahmadpour et al. (2019) state that in a procedural proof a representation “records a general procedure rather than representing an abstract structure” (p. 11). Their examples include a student reading an algebraic proof but conceptualizing it as a recipe for confirming individual proofs.

Students are classified as being on the path of structure if abstraction has taken place. Ahmadpour et. al (2019) define abstraction as “the construction of an idea that stands for a class” (p. 9). Student understanding is classified as being in the Abstract Structure state if students abstracted the general procedure. They take as evidence of abstraction instances where students identify how the algebraic notation $n + n + 1 + n + 2$ describes all sums of three consecutive natural numbers. Specifically, instances where students connect components of the algebraic expression with sums of three consecutive natural numbers.

In their model, abstract structure, abstraction, and thus generality are necessarily connected. They use as evidence of abstraction students identifying the abstract structure of the domain which is their indication that the student perceives the representation as standing for all objects in the domain, i.e., the representation is general.

2.3.4 Conceptual Insight and Technical Handle

Sandefur et al. (2013) developed a theoretical framework that ties together the representation used by a student in their proof construction with the idea the student is seeking to communicate with the representation using mathematical structure. Their framework builds on the ideas presented by Raman (2003). Raman identified three important components in proof evaluation and production: heuristic ideas, procedural ideas, and key ideas. Heuristic ideas give a sense of understanding that a claim is true from empirical data. Procedural ideas are based on logic and formal manipulation and lead to a sense of conviction that a claim is true without understanding of why. Finally, key ideas are "heuristic ideas that can map to a formal proof" (p. 323). They give both a sense of conviction that a claim is true and an understanding of why this is so.

Key ideas are critical in the construction of proofs. Raman et al. (2009) explored how novices and experts develop proofs and identified three significant moments: 1) getting a key idea, 2) discovering a technical handle, and 3) culminating the argument into a standard form. They further clarify from Raman's (2003) definition that key ideas are properties of a proof that give a sense of understanding but which do not always indicate a way to develop the formal proof. They used plural to describe key ideas because some proofs have multiple key ideas. The technical handle is used to "communicate a particular idea" (p.155). It can be based on a key idea, or it can be based on "informal thoughts or intuition" (p. 155). The final moment, culminating the argument into a standard form, involves "logically connecting given information to the conclusion" in a form that is the correct level of rigor for the given audience (p.155). In their study Raman et al (2009) found instances where students found a key idea and did not have access to a technical handle or had access to a technical handle but did not have a key idea.

Raman et al.'s (2009) key ideas and technical handle were then used and adapted by Sandefur et al. (2013) to explore how students use examples while developing proofs. Sandefur et al. refined key ideas into conceptual insights which they defined as "a sense of a structural relationship pertinent to the phenomenon of interest that indicates why the statement is likely to be true" (p. 328). They used the term technical handle and defined them to be "ways of manipulating or making use of the structural relations that support the conversion of conceptual insights into acceptable proofs" (p. 328). Sandefur et al. hypothesized that this framework would afford them access to the connections the representations used and the properties they are being used to represent.

Sandefur et al. (2013) found that example generation supported students in clarifying concepts and producing proofs. They witnessed that through generating examples, students developed conceptual insight and technical handles they could use to develop proofs. The discovery of conceptual insight and technical handle does not occur in a set order. Some students developed one before the other, as Raman et al found, and at times students will have access to one and not the other. Sandefur et al. also observed that some students stayed within the given representation system in hope of producing a proof quickly.

In the last few years conceptual insight has been used to examine the process of student proof production (Reed, 2021) and to examine the viability of student arguments (Yopp, 2020; Yopp, Ely, Adams, & Nielsen, 2022). When classifying student arguments, Yopp (2020) searched for a viable conceptual insight to categorize student responses that included evidence of argumentation but were too vague to classify as viable arguments. He defined viable conceptual insights as a conceptual insight, that could be developed into a viable argument[3]. Conceptual insight and technical handle in tandem with a focus on structure provide useful tools for examining student argument.

---

[3] Yopp defined a viable argument as "an argument that can be taken as proof but may have features that are less than formal, such as implicit inferences or intuitive argumentation/proof approaches that may not align with or attend to canonical methods or modes of proof" (2020, p. 5).

2.3.5 Definitions

Mathematical definitions relate registers and technical handles through conversions. Esty and Esty (2009) describe mathematical definitions as connecting a sentence with a new term to equivalent sentences expressed in more primitive terms. They give the example that a function f is increasing if and only if for all x and z in the domain if $x < z$ then $f(x) < f(z)$. In this example, we see that the mathematical definition outlines a conversion between the conceptual register and symbolic notation. The symbolic notation can serve as a technical handle giving access to the structure of increasing functions that could be leveraged to develop an accepted argument about increasing functions. In their model for "a way to work with a new term," Esty and Esty show how a definition can be used to translate or convert a sentence to "primitive terms." Esty and Esty show how the primitive terms can be used to "do work" and then a translation can occur back to interpret the result of the work (p. 121)

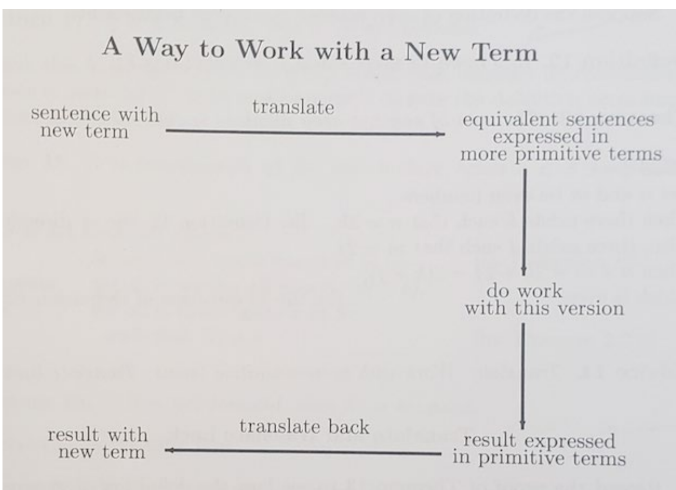


Figure 6: Esty and Esty model for working with a new term (2009, p. 121)

The model Esty and Esty developed for working with a new term (see Figure 6) can be natural extended to model completing a general direct argument with access to definitions that can serve as technical handles. The claim refers to some general domain. The definitions then describe a possible conversion from the language of the claim to a different register. Within the new register, treatments can be applied to the representation of the domain to demonstrate the desired property. The product of the treatments can then be translated back to the original register of the claim using definitions. This mirrors the model described by Pedemonte and Balacheff (2016).

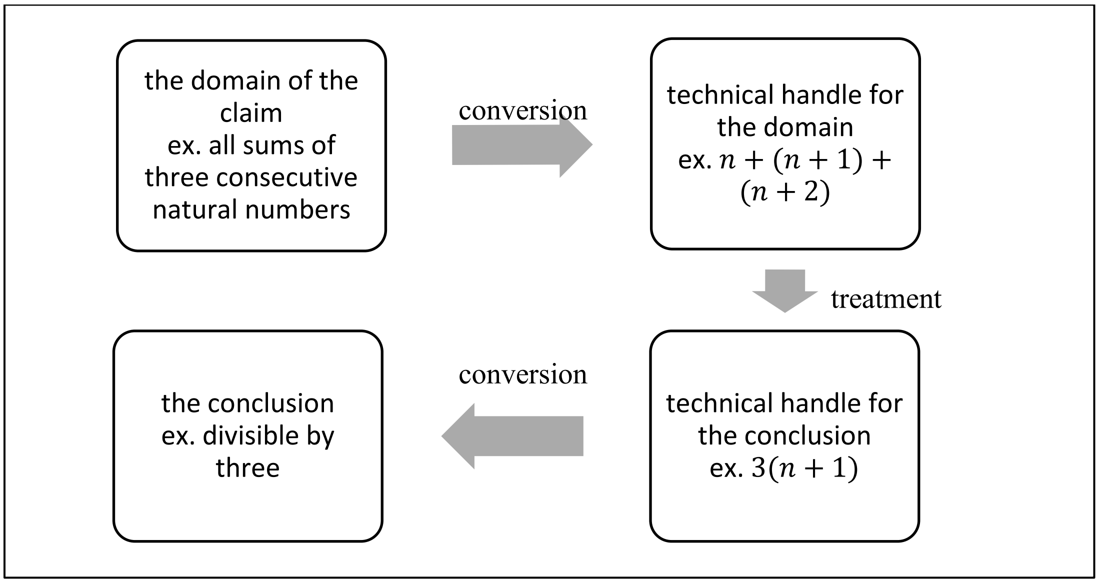


Figure 7: Extension of Esty and Esty's model to writing a general direct argument.

The model in Figure 7shows one possible path to developing a general direct argument. Key to this path is the conversion from the register in which the claim is presented to some register that the student sees as useful to communicating the structure of the domain they see as pertinent.

2.4 Relation of Previous Research to my Research Questions

Studies demonstrate that students of all ages find developing proofs to be challenging. Studies consistently find that students develop empirical proofs in response to general claims (Coe & Ruthven, 1994; Goetting, 1995; Harel & Sowder, 1998; Healy & Hoyles, 2000; Sowder & Harel, 2003). Historically, this phenomenon has been explained by asserting that students find examples are sufficient for demonstrating the truth of an infinite set because they have immature mathematical thinking (Balacheff, 1988; Harel & Sowder, 1998). However, there is a growing body of research finding that empirical arguments are not evidence of students being empiricists. My study is situated within this growing body of research that is seeking to understand when and why empirical arguments happen as I explored whether the presence or absence of definitions and general representations influenced whether students developed arguments that were empirical or deductive.

In my study I decided to examine student responses to tasks when varying definitions are provided because of the role definitions may play in supporting students with discovering pertinent structure of the domain and discovering a technical handle they can use to access the structure. Duval (2006) highlighted that the conversion from one register to another is significant mathematical work that is challenging to students. Combining the ideas of Duval and Sandefur et al., there is likely a conversion that must take place from the statement of the claim within the conceptual register to a register that supports the technical handle.

Definitions as described by Esty and Esty (2009) provide a means of converting from one register to another. I hypothesized that deductive arguments would happen more frequently when students were given definitions. The mechanism that I hypothesized would be at work is that by utilizing the given definitions, students would be more likely to perform a conversion to a register that would support them in discovering a conceptual insight and technical handle.

Chapter 3: Theoretical Framework

3.1 Proof and Argument

Mathematicians and researchers disagree the criteria for proof (Balacheff, 2008; Cirillo, Kosko, Newton, Staples, & Weber, 2015; Reid & Knipping, 2010; Stylianides, 2007). Yet to teach proof and study student proving it is essential to know what constitutes a classroom proof (Czocher & Weber, 2020). Stylianides (2007) naviagated the challenge of defining proof by developing a definition specific to a mathematical community. The criteria he presents are all situated in what is appropriate and acceible for the community in which the proof is developed or presented. Proof, as defined by Stylianides (2007), is a "connected sequence of assertions for or against a mathematical claim, with the following characteristics: 1) It uses statements accepted by the classroom community that are true and available without further justification; 2) It employs forms of reasoning that are valid and known to, or within the conceptual reach of, the classroom community; and 3) It is communicated with forms of expression that are appropriate and known to, our within the conceptual reach of, the classroom community" (p. 291). In this manner referring to a response to a claim as a proof includes an assertion of validity of the statements and logic contained in the response. Mathematical argument as defined by Reid and Knipping (2010) includes three components: a claim, data, and a warrant. The claim is a mathematical conjecture, the data is information that supports the claim, and the warrant provides the link between the data and the claim. Mathematical argument provides a less restrictive definition than proof as it includes work that is not based on deductive reasoning or work that has invalid steps or conclusions.

For the purposes of this study, "argument" will be used to describe the work produced by students in response to a claim. Argument will be used as it has no assumption of validity, allowing the study to explore what students produce and how they approach producing a proof. The term "viable argument" will be used to describe arguments that share the properties of proof as described by Stylianides (2007).

A general direct claim asserts that all objects in one set belong to a second set. With this framing we can consider the first set to be the set defined by the conditions of the claim. This set will be referred to as the domain of the claim, and the second set to be defined by the conclusions of the claim. Thus, for the claim "all sums of three consecutive natural numbers are divisible by three", the domain is the set of all sums of three consecutive natural numbers and the set defined by the conclusion is the set of numbers divisible by three. A viable argument to this claim is tasked with demonstrating that every object or all objects in the domain are in the set of numbers divisible by three.

A general direct argument is one way to form a proof to a general claim. A general direct argument starts with objects in the domain and demonstrates that they must be in the set defined by the conclusion. An argument can be formed by considering all objects in the domain simultaneously and demonstrating that they all satisfy the conclusion. One way to accomplish this is to represent all possible objects in the domain of the claim as a single entity and to perform a transformation on that representation to demonstrate that the properties of the conclusion are satisfied. Alternatively, the argument can be formed by representing an arbitrary object in the domain of the claim and demonstrating via transformations that the object satisfies the conclusion. These approaches sound similar, yet they are distinct in how the representation is viewed and used. The first approach uses a representation as a stand-in for all objects in the domain of the claim; this will be referred to as a class representation as it stands for the class of objects. The second approach uses a representation that is a placeholder where any of its kind can go; this will be referred to as a placeholder representation. For each of these approaches there is a set, the perceived domain, that describes the set of objects the arguer intends the representation to capture. In the case of a class representation, the perceived domain is the set of objects the arguer intends the class representation to stand for. In the case of the placeholder representation, this is the set of objects from which the arguer draws to then place into the placeholder representation. For the representation to be of the domain, the perceived domain and the domain must be the same or equivalent sets.

3.2 Mathematical Structure

In existing literature, structure has been used to describe the syntax of problems (Hoch & Dreyfus, 2004), the relationships between internal parts of an object (Banerjee & Subramaniam; Kieran, 2018; Malle, 1993), the relationship between an object and other objects (Morris, 1999; Simpson &Stehlíková, 2006; Warren, 2003), and one of two ways to conceive of mathematical objects (Sfard, 1992). In this study I will use structure to describe the way a participant characterizes the properties of the mathematical objects.

Using only properties and representations to describe arguments is not sufficient because there is variation in how properties and representations are used that is identified in the argument's conceptual insight. To illustrate this, I will outline two example arguments that are attending to the property consecutiveness using the conceptual register, yet which have fundamental differences that structure as described above allows me to describe.

When students develop a viable argument for the claim "any sum of three consecutive natural numbers is divisible by three", they attend to the property of consecutive numbers and sums of three such numbers. The way they attend to this property can differ.

> Argument 1: Since the numbers are consecutive the first addend will be one less than the second addend and the third addend will be one greater than the second addend. This means I can always take one from the greatest addend in the sum and add it to the smallest addend to get three copies of the middle number.

> Argument 2: Since the numbers are consecutive the second addend is one more than the first and the third addend is two more than the first addend making a total of three. This means I can think of each sum as three copies of the first number plus an extra three.

Each of these arguments represents the domain using the conceptual register. So there is no notable difference in representation usage. Each argument attends to the properties of the domain (i.e., sums of three numbers where each number is one greater than the previous). However, the manner in which "consecutive" is characterized is different. Argument one is characterizing "consecutive" as $(x-1)+x+(x+1)$ while the second is characterizing "consecutive" as $x+(x+1)+(x+2)$. Thus while the representation used and the property described is the same the manner in which the student is characterizing the objects is quite different.

The structure the student is attending to may be inferred by a representation, for example if one student constructs the representation $A + B + C$ and another constructs the representation $(x - 1) + x + (x + 1)$ to represent the set of all sums of three consecutive natural numbers a researcher could infer that the first student is attending to the objects being sums of three numbers and the other is attending to consecutiveness centered around the middle addend. By defining structure to be the manner in which students characterize the properties of the domain I can examine the possible conceptual insights available to the student

Table 1 shows four different structures of the domain which a student could identify. The structures are depicted using the conceptual register and the symbolic register. Some structures seem to lend themselves more easily to particular representations. It is, for example, easier to imagine a student identifying the structure that all sums of three consecutive natural numbers will have one remainder equal to 0, 1, and 2 respectively using the conceptual register than when using the symbolic register. It is possible to imagine how a student might build a generic example to depict the first two characterizations of consecutiveness. However, a generic example may be less intuitive in some of these cases.

	Register	
	Conceptual Register	Symbolic Register
Structure	The second addend is one greater than the first and the third addend is one greater than the second or two greater than the first.	$x+(x+1)+(x+2)$ for $x \in N$
	The first addend is one less than the second and the third addend is one greater than the second.	$(x-1)+x+(x+1)$ for $x \in N$
	One addend is divisible by three. Another will have a remainder of 1 when divided by three because it is either one greater than or two less than the number divisible by three. The final addend will have a remainder of 2 when divided by three because it is either one less than or it is two greater than the number divisible by three.	Every sum is one of three forms: $3k+(3k+1)+(3k+2)$ Or $(3k-1)+3k+(3k+1)$ Or $(3k-2)+(3k-1)+3k$ for k$\in N$
	The sum 1+2+3 is divisible by three. Every sum of three consecutive natural numbers can be built by adding one to each addend repeatedly.	$1+2+3$ $2+3+4$ $3+4+5$ ... $(1+k)+(2+k)+(3+k)$ for k$\in N$

Table 1: Structure communicated in the conceptual and symbolic registers.

3.3 Generic Example Argument

Within the wide breadth of types of arguments produced by students participating in proving activities is the generic example proof. Generic example proofs are formed using a generic example—an example that is seen by the arguer as representative of a typical object in the domain. Rø and Arnesen (2020) define a generic example argument as an argument that "takes as its basis an example of the claim to be proved, continues with a mathematical reasoning on why the claim holds for the example, and winds up with a lifting of this mathematical reasoning to the general claim" (p.13). To determine whether the reasoning is "lifted" to the general claim in my coding I will check two criteria: 1) Do the actions performed demonstrate either a transformation that can be applied to all objects in the domain or a procedure that can be applied to any object drawn from the domain? And 2) Does the arguer reference the domain in a general manner in connection to their example and the actions they perform? If an arguer produces a generic example satisfying these conditions, the work will be interpreted as general and can be used to form a viable argument.

3.4 Representations and Registers

Representations are attributed different meanings depending on whether they are taken to describe the "act of capturing a mathematical concept or relationship in some form" or the form itself (NCTM, 2000, p.67). Some researchers consider three modes of representation: enactive, iconic, and symbolic (Bruner, 1964). Where enactive representations are characterized by a sensory-motor actions, iconic representations are characterized by bearing a selective organization of the perceived event or object, and symbolic representations words or symbols are used to communicate particular referents. Other frameworks for representations include classifying the representation as either internal or external (Goldin & Shteingold, 2001; Goldin, 2003; Zhang, 1997). In this study as I was studying proofs constructed by students, I was interested in external, symbolic representations, in particular the form in which students communicated the mathematical concepts.

Stylianou (2011) defines representations as "configuration[s] that stand for something else" (p. 266). She identifies "symbolic expressions, drawings, written words, graphical displays, numerals, and diagrams" as examples of representations of mathematical concepts (p. 266). Stylianou's definition of representation is useful for this study as it attends to written and drawn notation students create to communicate their reasoning. This study is exploring the written arguments students construct and the examples detailed by Stylianou are the tools students use to communicate written arguments. This definition has been used in similar studies to explore student proving activity in classrooms (Weiss & Herbst, 2007). Representations are important in the study of mathematics as mathematical objects can only be accessed through representations and transforming representations is at the heart of mathematical activity (Duval, 2006).

Registers have been defined to discuss and distinguish between systems of representations. Duval (2006) defines register as a representation system that allows for transformations on the representation. Transformations can take one of two forms: 1) treatments, which remain within the same register, and 2) conversions, which translate a representation from one register to another without altering the object that is represented. The registers defined by Weiss and Herbst (2007) allow them to consider three distinct categories of representations within geometry:

1. the conceptual register: objects are referred to by the abstract classes they belong to i.e., isosceles triangle,
2. the generic register: a particular instance is used as a representative of the class of objects i.e., ΔABC with the property $\overline{AB} = \overline{BC}$. And,
3. the diagrammatic register: a diagram is used to represent a class of objects i.e., a diagram of a triangle with vertices labeled A, B, and C with tick marks to indicate that sides AB and BC are congruent.

In their definitions, Weiss and Herbst (2007) focus on the conversion from one register to another. They do not emphasize the ideas of treatments within registers.

Extending the generic and diagrammatic register outside of the context of geometry introduces ambiguity. This is because whether students intend their representation to objectify the domain as an abstract class or as a particular instance that stands for the class involves assuming how the student conceives of their representation. For any given representation the way it is conceived may not be universal in a manner that fits with the three categories provided by Weiss and Herbst.

In this study, I use the categories for the conceptual register as defined by Weiss and Herbst (2007). The generic register will not be used as it assumes that the representation stands for a particular case. Instead, additional categories will be used that can be applied from examining written work alone: 1) symbolic notation, 2) generic examples, 3) partial examples, and 4) conforming examples. Symbolic notation includes representations where symbols are used to objectify the domain. This includes classic mathematical notation, i.e., n+n+1+n+2, and student generated symbolic notation, i.e., dots drawn on a diagonal with the vertical distance labeled as "+1." Generic examples are as defined above. Conforming examples have no indication of generality; they are examples in the domain that demonstrate that the example satisfies the conclusion of the claim, i.e., 1+2+3=6, 6/3=2. Partial examples are sets of three consecutive natural numbers such as 3, 4, 5. If partial examples are the only representations present, it is not clear that the participant knows what the domain of the claim is or what properties of the domain the claim is asserting.

3.5 Conceptual Insight

Conceptual insight and technical handle have become tools for researchers to consider features of student proof with more nuance than considering only whether the argument counts as proof (Reed, 2021; Yopp, 2020; Yopp, Ely, Adams, & Nielsen, 2022). It is widely accepted that determining what is proof is context dependent. Consider Stylianides' (2007) commonly used definition of proof,

> connected sequence of assertions for or against a mathematical claim, with the following characteristics: 1) It uses statements accepted by the classroom community that are true and available without further justification; 2) It employs forms of reasoning that are valid and known to, or within the conceptual reach of, the classroom community; and 3) It is communicated with forms of expression that are appropriate and known to, our within the conceptual reach of, the classroom community (p. 291).

Utilizing this definition to evaluate a proof requires immense knowledge about the community in which the proof was made. Each statement, form of reasoning, and form of expression must be assessed within the context to determine whether it is appropriate for the given community.

The notion of conceptual insight can be used to classify the variety of arguments produced that are general without making judgements about the viability of the argument or proof. Sandefur et al. (2013) defined conceptual insight to be "a sense of a structural relationship pertinent to the phenomenon of interest that indicates why the statement is likely to be true" (p. 328). Given a general claim there may be multiple possible ways the pertinent structure can be related to the phenomenon, resulting in multiple possible conceptual insights for a student to leverage. Consider the claim that the sum of three consecutive natural numbers is divisible by three. One possible conceptual insight a student could leverage is that all sums of three consecutive natural numbers have the form $n + n + 1 + n + 2 = 3(n + 1)$ where n is a natural number. Alternatively, a student could note that $1 + 2 + 3$ is divisible by three and that they can construct any sum of three consecutive natural numbers by adding one to each addend recursively, i.e., $2 + 3 + 4 = (1 + 1) + (2 + 1) + (3 + 1) = (1 + 2 + 3) + 3$. So, any sum of three consecutive natural numbers is equal to $1 + 2 + 3$ plus some multiple of three. These two conceptual insights lead to different arguments that require different uses of representations as well as different uses of logic.

Above, conceptual insights are given that can be used to form a viable argument. They will be referred to as viable conceptual insights (Yopp, 2020). However, a student can identify structural relationships that are not pertinent. For example, students can identify the alternating even and odd structure of consecutive numbers and seek to use that structure to demonstrate why the claim is true. This is a non-viable conceptual insight because this structure cannot be used to demonstrate that the sum is divisible by three without relying on an additional structure, property, or prior result. The distinction between viable and non-viable conceptual insight is not always clear and will involve some degree of reader bias because whether the reader perceives the conceptual insight to be viable depends on whether the reader sees a path to proof using that conceptual insight.

Chapter 4: Methods

An embedded mixed method approach was used in this study utilizing both quantitative and qualitative data. This is a popular design for investigating an intervention within a school setting (Creswell & Creswell, 2017). Quantitative data was collected from participants to address research questions one and two. Within the quantitative study, assumptions are made about the arguments students make. For example, the coding scheme for the quantitative study classifies written arguments as including general representations. However, non-conventional representations may be interpreted as not general when the student developing them intended them to be general or, as Ahmadpour et al. (2019) found, when students used variable expressions in ways that were not general that would be assumed by the mathematical community as being general.

The embedded qualitative study was used to explore the arguments as intended by the participants. In particular, how participants perceived their representations in relationship to the domain of the claim. The embedded mixed method approach is appropriate for this study as it allows for the quantitative results to be contextualized by qualitative data (Creswell & Creswell, 2017). Table 2 includes a summary of the methods including the research questions, the data collected, and the data analysis.

Research Question	Data Collected	Data Analysis
(1) How does access to a definition or general representation of the mathematical objects in the claim influence the generality of a student's argument and the way the student represents the domain?	170 written arguments from 15 different math classes. Coded using coding scheme (Table 3).	Randomized Block Design to consider the relationship between task version and generality of participants' written work (Kuehl, 2000).
(2) How does access to a definition or general representation of the mathematical objects in the claim influence the conceptual insights that are used in the argument?	170 written arguments from 15 different math classes. Coded using coding scheme (Table 4).	Randomized Block Design to consider the relationship between task version and viability of the conceptual insight used in the written response (Kuehl, 2000).
(3) How do students describe the representations they develop or choose to utilize in their arguments?	9 filmed interviews Thematic analysis (Table 14) with a combination of open and closed coding (Table 13).	Thematic Analysis was used with both inductive themes and theoretical themes that are drawn from extending the work of Ahmadpour et al. (2019) (Braun & Clarke, 2006).

Table 2: Summary of Methods

4.1 Quantitative Study

4.1.2 Participants

This study took place at universities in Washington, Idaho, Montana, and Oregon in courses designed to support mathematics content knowledge for elementary school teachers and prospective elementary school teachers. These students have met an algebra requirement. The math courses for elementary school teachers and prospective elementary school teachers do not focus algebraic solving methods. I anticipated that these students would have seen and worked with algebraic notation in previously and would not be primed by their current coursework to immediately produce and algebraic representation to solve a problem.

4.1.3 Recruitment

Participating instructors were contacted first with a summary of the project and a detailed explanation of what participation in the study would involve. Follow-up conversations by phone were used to connect with instructors or clarify any questions that arose. To recruit student participants, a mixture of methods were used depending on the instructors' preferences and feasibility. For the courses that were located within accessible driving range, I supported recruitment in person by offering to attend in person to introduce my project and describe to the students what participation would entail. For courses that were located at a distance, the instructors were given a script introducing the project that they read to their students. It is possible that the different methods of recruiting students impacted how participants interacted with the tasks. However, the randomized block design controls for variations between classes.

4.1.4 Instruments

Participating instructors were sent a collated stack of the tasks and asked to distribute them in order by walking a systemic path through their class. This distribution technique ensures that each participating student is randomly given one of the three versions of the same task to complete and that roughly the same number of students will complete each version of the task. Each task includes the following prompt and claim:

> Prove that the sum of any three consecutive natural numbers is divisible by three. In your work, show and explain why this is true.
>
> **Claim:** The sum of any three consecutive natural numbers is divisible by three.

The claim was chosen because the content is approachable, participants can develop their own general representations, there is a symbolic representation that will be of a familiar form to participants, and the same claim has been used in previous studies exploring how students interpret written arguments (Ahmadpour et al. , 2019).

I felt the content of the claim would be approachable because I anticipated that students enrolled in university math courses for prospective elementary school teachers would have encountered consecutiveness, natural numbers, sums, and divisibility by three at some point in their mathematics educations. The structure of the domain is also accessible in non-symbolic representation systems. Students can access the domain using generic examples and informal representations. The participating students have also met an algebra requirement which led me to hypothesize that the algebraic representation for three consecutive natural numbers ($n, n + 1, n + 2$) would be familiar to the students. I further hypothesized that participants would see how this representation connects to consecutive numbers and that they would utilize the definition to represent the domain of the claim. Finally, Ahmadpour et al. (2019) examined students reading proofs to this claim to develop their model. Using the same claim allowed me to directly compare my findings to research question three to their model for how students understand proofs they read.

The three versions of the task vary in the definitions that are provided or not provided. Task A includes no definitions. Task B includes definitions in the conceptual register for both consecutive and divisible by three as well as examples of those definitions (see Figure 8).

Task B

Definitions:

Definition: Two natural numbers are consecutive if one of the numbers is one greater than the other number.

Example: 4,5,6 are consecutive natural numbers because 5 is one greater than 4 and 6 is one greater than 5.

Definition: A number is divisible by three if there is no remainder when the number is divided by 3.

Examples: 6 is divisible by three because 6 divided by 3 is 2 with no remainder.

Prove that the sum of any three consecutive natural numbers is divisible by three. In your work, show and explain why this is true.

Claim: The sum of any three consecutive natural numbers is divisible by three

Figure 8: Task B

Task C includes definitions in the symbolic register for both "three consecutive numbers" and "divisibility by three" as well as examples of the definition (see Figure 9).

Task C

Definitions:

Definition: Three consecutive natural numbers are numbers that can be represented as n, n+1, n+2 for some natural number n.

Example: 4, 5, 6 are three consecutive integers because they can be written as 4, 4+1, 4+2

Definition: A number m is divisible by 3 if it can be written as $3 * k$ for some integer k

Example: 6 is divisible by three because $6 = 3 * 2$. This means $6 = 3 * k$ when $k = 2$.

Prove that the sum of any three consecutive natural numbers is divisible by three. In your work, show and explain why this is true.

Claim: The sum of any three consecutive natural numbers is divisible by three

Figure 9: Task C

4.1.6 Measures

To address the first and second research question, the written arguments were coded along two variables: 1) type of representation and 2) conceptual insight. The coding schemes were developed using prior literature and were fine-tuned during the pilot study.

The closed coding scheme for the type of representation (Table 3) includes six codes that are drawn from previous literature as well as the pilot study. The conceptual register as defined by Weiss and Herbst (2007) is included. The generic register from their work has been further subdivided by the coding scheme into three subcategories: algebraic notation, informal symbolic notation, and generic example. It is worth noting that Weiss and Herbst assume that within the generic register, the representation is of a particular instance in the domain that is representative of the set. While these algebraic notations, informal symbolic notations, and generic examples can be interpreted in this manner by a reader, the follow-up interviews were used to explore whether the writer intended the representations to stand for a particular instance or had some other intention. In the qualitative study I explored whether these types of representations belong to the generic register as defined by Weiss and Herbst or whether the student perceives their representation as defining an abstract class.

Arguments that were coded as including the conceptual register, algebraic notation, informal symbolic notation, and/or generic examples will be coded as general as these ways of representing the domain are accepted in mathematics as being general. The other codes will be coded as not general. These codes include partial examples, conforming examples, and blank or unrelated work.

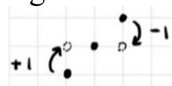
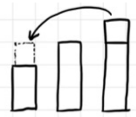
The closed coding scheme for conceptual insight (Table 4) was initially developed by anticipating arguments that participating students may develop. Additional codes were added to describe the data collected in the pilot study. Arguments that include conceptual insights that are not included in the existing coding scheme were coded as other (viable) or other (non-viable). After initial coding I considered the arguments which were coded as other using inductive thematic analysis to determine what themes exist in the conceptual insights of these arguments (Braun & Clarke, 2006). This led to the formation of a new code to describe the viable conceptual insight that one of the addends will be divisible by three and the sum of the other two addends will also be divisible by three. The details of this revision of the coding scheme can be found in the results section.

Conceptual insight is the structural relationship that is used to indicate the claim is true. Each conceptual insight present in the arguments will be coded. Viable conceptual insights will be defined as conceptual insights that are sufficient to form an argument without needing to reference another structure of the domain. Non-viable conceptual insights are conceptual insights that cannot be used to form a viable argument without referencing a different structure of the domain. For example, we can imagine a participant developing a viable argument by first starting with the structure “odd and even”. Perhaps they note that there are two cases; the sum’s addends are of the form odd-even-odd or even-odd-even. This conceptual insight cannot on its own be leveraged to form a viable argument to the claim. However, this is not to say that this line of reasoning would not support a student to develop a viable argument by discovering some other structure. Non-viable, as defined here, includes no judgement on whether the insight is a useful tool for participants in the proof development or part of a successful learning trajectory. Non-viable in this study is used only to categorize conceptual insights that would need to pair with some other fundamental structure of the domain to form a viable argument to the claim.

Code	Sub Code	Description
General	Conceptual Register	The domain is described as an abstract class. Properties (correct or incorrect) are attributed to that set as a class. Ex. All sums of three consecutive numbers have the property that the next number is one greater than the previous… Ex. All sums of three consecutive numbers are divisible by three because they can be broken into three groups because the total is a sum of three numbers. Ex. It is true because they are consecutive and there are three numbers. Ex. The average of the three consecutive natural numbers will always be the middle number.
	Structural Algebraic Notation	The domain is objectified using algebraic notation that demonstrates a structure shared by all objects in the domain. The representation may include a transformation that may or may not have an algebraic error (e.g., 3n+3=3(n+3)) Ex. n+n+1+n+2 Ex. n+n+1+n+2=3(n+1) Ex. n+n+1+n+2=3n+3=3(n+3)
	Non-Structural Algebraic Notation	The domain is objectified using algebraic notation that does not objectify a pertinent structure shared by all objects in the domain. Ex. A+B+C

Table 3: Closed coding scheme for type of representation and generality

Table 3 Continued

General	Informal Symbolic Notation	The domain is objectified using symbols or diagrams that are not algebraic. The representation may depict a transformation. Ex.  Ex. 
	Generic Example	A generic example is used. This includes an example in the domain, actions performed on the example that demonstrate a structure shared by all objects in the domain or a procedure that can be applied to any such object from the domain, and a statement or markings referencing the domain in a general manner in connection to the arguer's example and the actions they performed. Ex. 4+5+6 +1 -1 5+5+5 I can always even the sum into a sum of three copies of the middle number
	Other	A representation that is general that does not fit the above categories.
Not General	Blank or non-examples	The argument includes calculations that are not examples in the domain, is blank, or has statements such as "I don't know."
	Partial Examples	The argument includes examples that share some property of the domain. It is unclear from the work if the writer understood the full definition of the domain. Ex. 1, 2, 3 Ex, 6, 6+1, 7+1
	Examples in the domain	The argument includes an example in the domain. The example is not transformed to demonstrate the truth of the conclusion. Ex. 4+5+6
	Conforming examples	The argument includes an example in the domain transformed to demonstrate the truth of the claim. Ex. 4+5+6=15, 15/3=5
	Other	A representation that is not general that does not fit the above categories.

Code	Sub Code	Description of the Conceptual Insight
Viable	Redistribution	The sum of three consecutive natural numbers can be redistributed as the sum of three copies of the middle addend.
	Symbolic Manipulation	The sum of three consecutive natural numbers is represented symbolically and manipulated into a form that is recognized as demonstrating divisibility by 3. If the argument includes an interpretation of the algebraic work using another conceptual insight (ex. Redistributing or summing remainder), then code the argument as both symbolic manipulation and the other conceptual insight. Ex. The sum of three consecutive natural numbers can be written as x-1+x+x+1=3x so we know the sum is always divisible by 3. Ex. n+n+1+n+2=3n+3 Ex. n+n+1+n+2=3n+3=3(n+1)
	Three copies plus three	The sum of three consecutive numbers is equivalent to three copies of one number plus 3. Since three copies of a number is divisible by three and three is divisible by three the sum is divisible by three.
	Recursive	The sums for a base case or a set of base cases satisfy the conclusion, and the remaining cases can be constructed by either adding sums from the set of base cases or adding some multiple of three. Ex. The sum 1+2+3 is divisible by three and every sum of three consecutive natural numbers can be constructed by adding one to each addend of 1+2+3 recursively. Thus, every sum of three consecutive numbers is of the form 1+2+3 + 3k for some natural number k. Ex. The claim is true for sums of three consecutive integers between 0 and 10 and all other sums of three consecutive integers can be constructed by adding combinations of those sums.
	Summing Remainders	In any sum of three consecutive natural numbers one of the addends will be divisible by three. When divided by three, the remainders of the other two addends will be 1 and 2 respectively. This means the sum of the remainders will be 3 making the sum of the other two numbers divisible by 3.
	Average	Adding three numbers and dividing by three finds the average of the three numbers. Since the three numbers we are adding are consecutive, the middle number is the average.
	Other	A conceptual insight that is viable that does not fit the above categories.

Table 4: Closed coding scheme for conceptual insight

Table 4 Continued

Not Viable	Even and Odd	The sum of three consecutive natural numbers will always have either one even and two odd addends or two even and one odd addend.
	Groups of Three	The sum of three consecutive natural numbers is divisible by three because we are adding three numbers. May describe three numbers as three groups.
	Division Mix up	All numbers are divisible by three but there may or may not be a remainder.
	Observed Property	A property or pattern observed from examples is used to justify the truth of the claim. The pattern may or may not hold for all cases in the domain. Ex. The claim is true because the quotient is the second number in the sum. Ex. For any sum of three consecutive natural numbers the sum of the first two numbers will be divisible by 3 and the last number will be divisible by three making the sum divisible by 3. Like, 4+5+6 where 4+5 = 9 is divisible by 3 and 6 is divisible by 3.
	Empiricism	The conforming examples show that the sum is divisible by three OR when you add the three numbers and divide by three you get a whole number so the sums are divisible by three. Ex. 4+5+6= 15, 15/3=5 The sum of any 3 consecutive natural number is divisible by 3 because as shown if you take any of the numbers and add them all together you get a number that will be divisible by 3. Ex. 4+5+6=15, 15/3=5 and 7+8+9= 24, 24/3=8. Using these two examples I can infer that the claim is true.
	Other	A conceptual insight that is non-viable that does not fit the above categories. This may include arguments that do not appear to address the claim. Ex. An argument for why a multiple of three is divisible by three.

Table 4 Continued

No Concep tual Insight		The argument includes no reason or justification for why the objects in the perceived domain satisfy the conclusion. This may manifest as summarizing the conditions as the reason for the conclusion holding. Or student may state that they do not know why the claim is true. Ex. It's true because if they are all divisible by 3 then they are all equal to the sum. ➔ Perceived domain= factors of three ➔ Conclusion = sums ➔ No justification connecting the two sets. Ex. 4+5+6= 15, 15/3=5 Ex. 3, 4, 5 are consecutive because they can be written as 3, 3+1, 4+1. 12 is divisible by 3 because 3x4=1. This means 12=3xk when k=4. No justification connecting the two sets sums of three consecutive natural numbers and numbers divisible by three.
	Unclear Conceptual Insight	The conceptual insight is unclear.

4.1.7 Interrater Reliability Study

To determine the validity of coding, an interrater reliability study was completed. The study was used to provide objective evidence that the categories described in the coding scheme exist through identifying whether independent researchers could reliably code consistently (Stemler & Tsai, 2008).

An additional mathematics education researcher was given a subset of the data and the coding schemes. We coded the subset of the data independently and then percent agreement and Cohen's kappa were calculated. The random sample of the data was created by first collating all data into a single PDF. Each of the one hundred and seventy student work samples were numbered based on its sequential order in the document. A random number generator was used to generate 30 integers in the range 1-170. The tasks corresponding to these 30 integers formed the subsample. We coded the subsample independently using the schemes. Once the subsample was coded, the percent agreement was calculated. Each argument is assigned as zero or one depending on the properties of the argument. This means the data is nominal in nature as it is describing the presence of features and will not be normally distributed. For that reason, using precent agreement for consensus estimates of interrater reliability is appropriate (Stemler & Tsai, 2008).

In addition, Cohen's kappa was calculated for the codes' generality and viability of conceptual insight. Cohen's kappa is interpreted as the level in which scores agree after taking into account the agreement that would occur simply by chance (Landis & Koch, 1977). Percent agreement can become inflated if codes are of low or high incidence of occurrence (Hayes & Hatch, 1999). Thus, the value of Cohen's kappa informs whether the percent agreement is due to chance.

Cohen's kappa is given by $\kappa = \frac{P_A - P_C}{1 - P_C}$, where P_A is the proportion of arguments on which the scorers agree on the score and P_C is the proportion of arguments for which agreement is expected by chance. The value P_C is given by the below formula:

$$P_C = \frac{total\ number\ of\ 1\ scores\ by\ 1st\ scorer}{total\ number\ of\ responses} * \frac{total\ number\ of\ 1\ scores\ by\ 2nd\ scorer}{total\ number\ of\ responses} + \frac{total\ number\ of\ 0\ scores\ by\ 1st\ scorer}{total\ number\ of\ responses} * \frac{total\ number\ of\ 0\ scores\ by\ 2nd\ scorer}{total\ number\ of\ responses}$$

(McHugh, 2012). Then value κ can be interpreted as described in Table 5 (Cohen, 1960). A sample size of 11 to 30 is recommended for initial calculations of Cohen's kappa (Bujang & Baharum, 2017). With this consideration, the 30 participant responses used for percent agreement are sufficient for the calculation.

Range	Interpretation
$\kappa \leq 0$	No agreement
$0.01 \leq \kappa \leq 0.2$	Slight agreement
$0.21 \leq \kappa \leq 0.4$	Fair agreement
$0.41 \leq \kappa \leq 0.6$	Moderate agreement
$0.61 \leq \kappa \leq 0.8$	Substantial agreement
$0.81 \leq \kappa \leq 1$	Almost Perfect agreement

Table 5: Interpretation guidelines for Cohen's kappa (Cohen, 1960)

Reliability of generality code.

There was agreement on 27 of the 30 codes resulting in 90% agreement. This is comfortably over the 70% minimum agreement required (Stemler & Tsai, 2008). Cohen's kappa for this code is approximately 0.8, suggesting there was substantial agreement. The calculations of Cohen's kappa can be found in Appendix G.

Reliability of viability of conceptual insight code.

There was agreement on 27 of the 30 codes resulting in 90% agreement. This is comfortably over the 70% minimum agreement required (Stemler & Tsai, 2008). Cohen's kappa for this code is approximately 0.7, suggesting there was substantial agreement. The calculations of Cohen's kappa can be found in Appendix G.

4.1.8 Power and Sample Size Analysis

The below calculations were performed to gain insight into what sample size will be needed to demonstrate a difference between the student responses to the three task conditions with the Randomized Block Design (RBD). Pilot data was gathered from two classes in the Fall of 2021. These student arguments were coded using the coding scheme. The full data set can be found in Appendix E.

I first considered what number of classes (blocks) needed to see a difference between the proportion of responses that are general and the proportion of those that contain viable conceptual insight for each of the three tasks. To gain a heuristic answer to this question, I grouped the two tasks (Task B and C) that include definitions and compared their proportion of responses that are general or contain a viable conceptual insight respective to the responses to task A which did not include a definition. This calculation has some limitations as it is comparing samples that are not equal in size. However, it indicated a reasonable number of classes to begin my investigation.

The formula $n = \sigma^2 \frac{2(z_{\propto/2}+z_{\beta})^2}{\Delta^2}$ can be used for choosing a sample size when testing the difference between two means (Ott & Longnecker, 2015). Table 6 gives the meanings of each of the symbols in the formula for sample size. For this calculation, instead of means I considered the proportion of participant responses that are general or the proportion of participant responses that contain a viable conceptual insight (see Table 7 and Table 8). For example, in block one of the pilot data, 2 of the 5 responses to Task A were coded as general. So, the proportion for Task A in block one is 2/5=0.4. The proportion of general responses for Tasks B and C, the tasks that include definitions, is 4/8=0.5. Thus, for block one the difference in proportions is 0.4 - 0.5 = -0.1.

Symbol	Meaning
n	Sample Size (number of classes or blocks needed for the RBD)
σ^2	Variance in the differences in proportions
Δ	Difference in proportions
z_α, z_β	The z-values associated with the chosen α and β values
α	The probability of incorrectly rejecting the null hypothesis (type I error)
β	The probability of incorrectly failing to reject the null hypothesis (type II error)

Table 6: Meaning of symbols in the formula for sufficient sample size.

Class	Proportion of General Responses to Task A	Proportion of General Responses to Tasks B and C	Difference in Proportions
1	2/5 = 0.4	4/8 = 0.5	-0.1
2	3/3 = 1	3/8 = 0.375	0.4096

Table 7: Summary of Pilot Data with Respect to Generality

Class	Proportion of Viable CI Responses to Task A	Proportion of Viable CI Responses to Tasks B and C	Difference in Proportions
1	0/5 = 0	3/8 = 0.375	-0.375
2	1/3 = 0.33	3/8 = 0.375	-0.045

Table 8: Summary of Pilot Data with Respect to Viability of Conceptual Insight

To calculate the sample size, i.e., the number of classes needed, I first found the variance in the difference in proportions for both Generality (${\sigma_G}^2$) and Viable Conceptual Insight (${\sigma_{CI}}^2$) using the pilot data ${\sigma_G}^2 \approx 0.13$ and ${\sigma_{CI}}^2 \approx 0.027$.and the standard accepted values for alpha and beta, $\alpha = 0.05$ and $\beta = 0.2$. This corresponds to 5% probability of falsely rejecting a true null and a 20% probability of failing to reject a false null hypothesis.

The alpha and beta values have the corresponding z values $z_{0.05/2} = 1.96$ and $z_{.01} = 0.842$. With the above values fixed we can now vary the desired difference in proportion to examine what sample sizes would be necessary to find that difference with confidence.

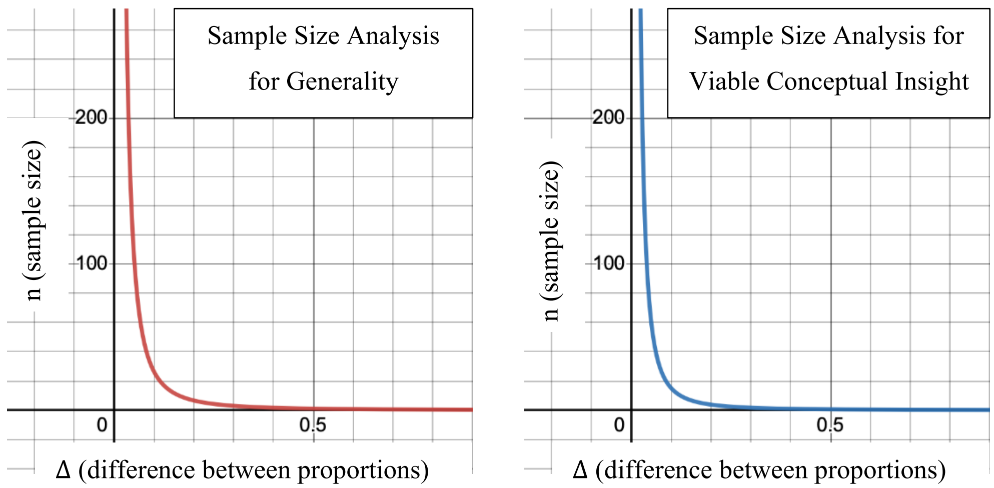


Figure 10: The relationship between difference in proportion and the sample size needed to find that difference with confidence.

Generality	
Difference in proportions (Δ)	Sample Size needed (n)
0.1	204
0.15	11
0.2	7
0.25	4
0.3	3

Table 9: Relationship between difference in proportion of general responses and needed sample size.

Viable Conceptual Insight	
Difference in proportions (Δ)	Sample Size needed (n)
0.1	15
0.15	7
0.2	4
0.25	0

Table 10: Relationship between difference in proportion of responses with viable conceptual insight and needed sample size.

Considering these calculations, it seems that to find a small difference in the generality of participant responses a larger sample size of approximately 73 classes is needed. However, to determine a larger difference, say of 0.2, 18 classes should be sufficient. The power and sample size analysis for viable conceptual insight resulted in much smaller sample sizes needed to find differences with confidence. Thus, a sample size of approximately 18 should be sufficient to find differences in viability of conceptual insight in student responses and differences in generality that are 0.2 or larger.

4.1.9 Randomized Block Design

Research questions one and two were addressed using statistical methods. A randomized block design was used to explore the two dependent variables: generality and conceptual insight, and their relationship to the task version respectively. This method is appropriate as the participants are naturally grouped by their setting (class). The randomized block design allowed for the differences from the class setting to be accounted for and removed from the error component (Kuehl, 2000). To apply the randomized block design the data must be continuous in nature. Each argument was assigned a 1 if general and a 0 if not general and a 1 if a viable conceptual insight is used and a 0 if a non-viable insight is used respectively. The model then examined the proportions of each task response that was general or viable respectively for each block making the data continuous in nature, and thus appropriate for the randomized block design.

The model for the Randomized Block Design is given by

$$Y_{ij} = \mu + T_i + B_j + random\ error,$$

where each of the variables is as described in Table 11 (Kuehl, 2000).

Symbol	Meaning
Y_{ij}	The proportion of students in classroom j with treatment i which include a general representation. OR The proportion of students in classroom j with treatment i which include a viable conceptual insight. * [Treatment 1 is Task A, treatment 2 is Task B, and treatment 3 is Task C.]
μ	The mean.
T_i	The effect for being in the treatment i.
B_j	The effect for being in block j.

Table 11: The meaning of the symbols in the Randomized Block Design Model
*This model is used twice: Once for the variable generality of representation and once for the variable viability of conceptual insight.

The model was applied twice resulting in two rounds of calculations. In the first round the dependent variable was generality of the argument and in the second round the dependent variable will be the viability of the conceptual insight. With the randomized block design both the block (B_i) and treatment (T_i) effects were tested. The block effect is included because I anticipated there would be an effect on the outcome based on the class in which the participant is enrolled. The null hypotheses and alternative hypotheses are given in the Table 12.

Hypotheses	
H_0: $B_i = 0$ for all i	H_0: $T_i = 0$ for all i
H_1: $B_i \neq 0$ for some i	H_1: $T_i \neq 0$ for some i

Table 12: The null and alternative hypothesis of the randomized block design

The statistics program SAS was be used to perform the calculations. For each dependent variable (generality and viability of conceptual insight) an overall ANOVA was calculated. First the F-statistic for the model was considered.

After determining whether the null hypothesis was accepted or rejected for each independent variable, I computed the effect size of the task version and blocking variable on generality and viability of conceptual insight using eta squared (η^2). Finding eta squared is done using values from the ANOVA table and the formula $\eta^2 = \frac{SS_{SOURCE}}{SS_{TOT}}$ where SS_{SOURCE} is the sum of squares for the variable in question and SS_{TOT} is the total sum of squares. The value eta squared is interpreted as the percent of the variance of the dependent variable for which the independent variable accounts.

4.2 Qualitative Study

4.2.1 The Setting

This portion of the study took place virtually with a subset of the participants involved in the quantitative study. Selected participants who indicated their willingness to be interviewed were contacted using the contact information provided by the student and follow-up interviews were scheduled within a week of the participant completing the written interview. Interviews were completed and recorded using Zoom.

4.2.2 Recruitment

On the bottom of the task, participants were asked to indicate whether they were willing to discuss their work in a follow-up interview (see Appendices A-C). They were then asked to give their preferred contact information and their preferred way of being addressed. All willing participants were contacted for follow up interviews.

4.2.3 Instruments

A semi-structured interview protocol was used to guide the interviews (see Appendix E). The protocol included two parts. The first part included questions meant to facilitate the researcher's understanding of the written argument as the participant intended it to be understood. This method of exploring the intending meaning of student representations builds on the research by Weiss and Herbst (2007) and Ahmadpour et al (2019) who, through discussing arguments with students, determined students were interpreting representations taken by mathematicians as general to be particular instances of objects from the domain of the claims.

The second part of the interview sought to establish the set of objects the student intended to represent with their objectifications and how the objectifications represented those objects. In the second part of the interview the interviewer chose one of two lines of questioning depending on whether the domain is described using words or if there is some representation, either symbols, generic example, or some other student generated configuration, present in the work. This portion of the protocol is very exploratory because existing literature has not attempted to determine the set students intended to describe by a representation and whether they see themselves as characterizing the set as a class representation, a placeholder representation, or some other characterization.

4.2.4 Thematic Analysis

Thematic analysis was used to analyze the transcribed interviews. When using thematic analysis, Braun and Clarke (2006) outline a number of decisions the researcher must make and communicate in their work including: what counts as a theme, whether inductive or theoretical analysis is used, and whether the themes are semantic or latent. Below I will expand on each of these decisions and then I will detail how my research is situated within the framework for thematic analysis.

Themes as defined by Braun and Clarke (2006) capture something important in the data in relation to the research question. They represent a pattern in the data. There is no one method for determining what level of pattern counts as a theme in thematic analysis. It is instead suggested that the researcher must use their judgement and, ultimately, key to the validity of the research is that the researcher is consistent in how they determine what counts as a theme throughout their analysis.

When conducting thematic analysis, the themes can be predetermined by existing theoretical frameworks or the themes can arise directly from the data using inductive thematic analysis. From the prior literature I theorized that students could develop representations that are empirical, representations that serve as placeholder representations, or they could develop class representations (see Table 13). Thus, during the thematic analysis

I used these theoretical themes to explore how these themes arise as students describe their representations.

Theorized Themes	Description
Empirical Representation	The representation is described as standing for one specific case.
Placeholder Representation	The representation is described as standing in place of any object drawn from the perceived domain.
Class Representation	The representation is described as standing for a class of objects.

Table 13: Theorized themes for theoretical thematic analysis.

Furthermore, the themes were based on semantic rather than latent features of the data. I examined directly what the participants said and or wrote as they discussed their argument. Instances where students discussed their representation and or the domain of the claim were coded by characteristics of their descriptions. These codes were then used to first formulate the inductive themes.

I followed the six phases of thematic analysis outlined by Braun and Clarke (2006) for the two rounds of thematic analysis. Each of the phases, with a description of my actions, is included in Table 14. For the second round of thematic analysis, the first two phases were completed identically to the first round. However, the following phases proceeded differently since the themes were drawn from theory rather than arising inductively from the data.

	Phase	Description of the process
Round 1: Inductive Thematic Analysis	1. Familiarized myself with the data.	I transcribed and read the data while reflecting on initial ideas.
	2. Generated initial codes.	I identified instances where participants described their representation or the domain and assigned initial codes to characterize how the participant described their representation and or the domain of the claim.
	3. Searched for themes.	I collated the codes into potential themes and gathered the data for each evolving theme.
	4. Reviewed themes.	I reviewed the themes by first examining the data extracts for each theme and then reflecting on the entire data set with respect to the theme.
	5. Defined and named themes.	I adjusted themes based on the review in Phase 4. I generated clear definitions and names for each theme.
	6. Produce the report.	I selected extracts for each theme that concisely capture the theme and used them to address my research question.
Round 2: Theoretical Thematic Analysis	Phases 1 and 2 were repeated as described above.	
	3. Searched for themes.	In this round, I used theorized themes. Thus, I examined the coded data to see how the data fit within the existing themes.
	4. Reviewed themes.	I collated the data extracts for each theme and reflected on whether the theme appropriately characterized the extracts with respect to the data set as a whole.
	5. Defined and named themes.	I adjusted and expanded upon the definitions for the theorized themes as necessary.
	6. Produce the report.	I selected extracts for each theme that concisely capture the theme and used them to address my research question.

Table 14: Phases of thematic analysis.

Initial codes were generated (step 2 in Table 14) by highlighting sections of the student interview where students referred to representations they used in their argument or classified the domain of the claim and characterizing how the student was representing or classifying the domain. During this initial coding, 32 codes were generated and assigned to excerpts from the interviews. In step 2 of the data these 32 codes were collated into the following five themes:

1. Examples as evidence
2. Algebraic representation
3. Patterns as a tool for seeing structure
4. Shifting representation or procedure
5. What it takes to prove.

A table depicting the initial codes organized into themes with example excerpts can be found in Appendix E.

The initial codes and themes were then reviewed by first examining the excerpts and reflecting on the entire data set with respect to the theme. During this process the theme "what it takes to prove" was absorbed by the other themes as these codes fell into two categories. The first category are comments that solidified that the participant saw their representation as either general or not general and the second category of responses were ones where participants described what they thought would be needed to show the claim was true for all sums of three consecutive natural numbers. These excerpts did not pertain to the participant's description of the domain or their representations. A summary of the final themes is given in the results section.

Chapter 5: Results

The purpose of this study was to investigate whether access to definitions key to the domain of the claim and the conclusion influenced the manner in which participants objectified the domain and the conceptual insight they used to connect the conditions of the claim to the conclusion. The study was guided by three research questions:

1. How does access to a definition or general representation of the mathematical objects in the claim influence the generality of a student's argument and the way the student represents the domain?
2. How does access to a definition or general representation of the mathematical objects in the claim influence the conceptual insights that are used in the argument?
3. How do students describe the representations they develop or choose to utilize in their arguments?

To address the first two research questions, the written arguments were first analyzed for their use of representation and conceptual insight using the coding scheme in chapter 4 (see Table 3 and Table 4). A randomized block design was used to test the two independent hypotheses that a definition would influence 1) the generality of the representations used as well as 2) the viability of the conceptual insight. Both the models were found to not account for statistically significant variation in their respective dependent variables—generality of representation and viability of conceptual insight. To further contextualize the results and answer research questions one and two, the types of representations and conceptual insights used across the task versions were examined.

To address the third research question, the participant interviews were analyzed using two rounds of thematic analysis. The first round used inductive thematic analysis and the following themes emerged: examples used as evidence, algebraic representations, shift in representations, and patterns as tools to see structure. In the second round of analysis, theoretical themes drawn from Ahmadpour et al. (2019) model for how students understand proof while reading were used to examine the manner in which students utilized the representations they chose and developed.

5.1 Research Question One

How does access to a general representation or general description of the mathematical objects in the domain of the claim influence the generality of a student's argument and the way the student represents the domain?

5.1.2 Examining Generality of Representations with a Randomized Block Design (RBD)

The ANOVA table for the RBD model with the dependent variable generality of representation shows that the model does not account for a significant portion of the variation in generality of participant representations (F=1.33, p=0.1849 > 0.05). Furthermore, with each of the Type I and Type III Sum of Squares, neither the task variety nor the block had a statistically significant effect on the generality of the representation (see Table 16 and 17).

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	5.1759	0.3234	1.33	0.1849
Error	153	37.1769	0.2429		
Corrected Total	169	42.3529			

Table 15: ANOVA table for the dependent variable Generality of Representation

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Block	14	4.7440	0.3388	1.39	0.1618
Variety	2	0.4319	0.2159	0.89	0.4133

Table 16: Type I Sum of Squares

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Block	14	4.6707	0.3336	1.37	0.1726
Variety	2	0.4319	0.2159	0.89	0.4133

Table 17: Type III Sum of Squares

Figure 11 and Table 18 below, show the relationship between the blocks and the generality of the representation used. Along the horizontal axis of the graph in Figure 11. are the 15 blocks that represent the 15 classes in which data was collected. Along the vertical axis is the mean or proportion of the arguments that included a general representation for each task version. The proportions for each task version are calculated by summing the total number of arguments in response to each task that include general representations and dividing that value by the total number of responses collected for that task. Table 18 shows the proportion of arguments including a general representation for each task version varied across the classes where data was collected. Furthermore, the number of responses for each task collected varied from one to nine. In some classes, as few as one response to each task version was collected. Also, in some classes significantly fewer responses to Task C were collected, compared to the other tasks.

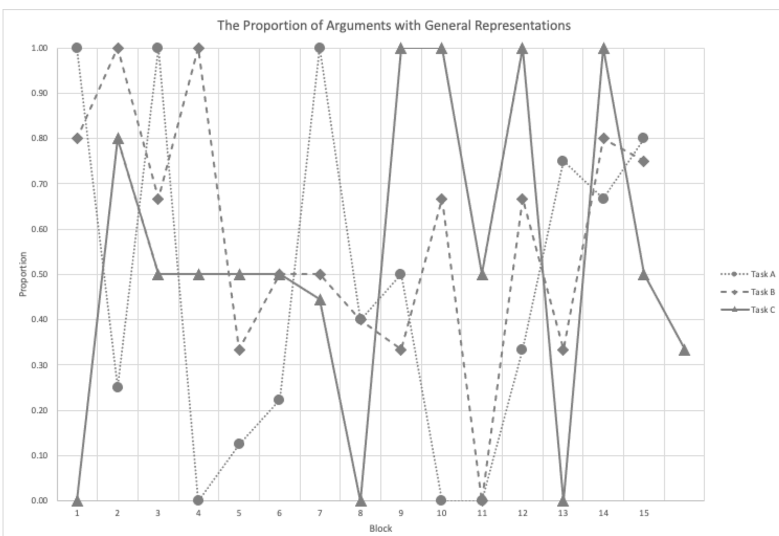


Figure 11: Interaction Plot for Blocks and Generality of Representations

Proportion of Arguments with General Representations by Block						
Block	Task A		Task B		Task C	
	Number of Arguments	Mean	Number of Arguments	Mean	Number of Arguments	Mean
1	5	1.00	5	0.80	5	0.80
2	4	0.25	4	1.00	4	0.50
3	3	1.00	6	0.67	2	0.50
4	2	0.00	3	1.00	2	0.50
5	8	0.13	9	0.33	8	0.50
6	9	0.22	8	0.50	9	0.44
7	1	1.00	2	0.50	1	0.00
8	5	0.40	5	0.40	1	1.00
9	4	0.50	3	0.33	1	1.00
10	1	0.00	3	0.67	2	0.50
11	1	0.00	1	0.00	1	1.00
12	3	0.33	3	0.67	2	0.00
13	4	0.75	3	0.33	2	1.00
14	3	0.67	5	0.80	2	0.50
15	5	0.80	4	0.75	6	0.33
Mean		0.47		0.58		0.57
Standard Deviation		0.36		0.27		0.32

Table 18: Proportion of Arguments with General Representations by Block

While the inferential statistics do not support the generalizability of these results, within the data collected the mean for Task B, the task with definitions in the conceptual register, was the highest, followed by Task C (symbolic definition), and followed by Task A (no definition) (see Table 19).

To examine the effect size, I calculated eta squared for both the variety in task and the blocking factor. The effect size for the version of the task is 0.0119, indicating that the task version was responsible for approximately 1% of the variation in generality of representations. The effect size for the blocking factor was 0.1102, corresponding to the class the data was collected in accounting for 11% of the variation in generality.

Task Version	Number of students	Generality	
		Mean	Standard Deviation
A	58	0.4655	0.4988
B	64	0.5938	0.4911
C	48	0.5208	0.4996

Table 19: Mean and standard deviation of generality.

Table 19 shows the mean and standard deviation for each version of the task. The coding scheme resulted in codes of 0 for non-general responses and 1 for general responses. Consequently, with means very close to 0.5 it is logical that the standard deviation is close to 0.5 as the standard deviation indicates the spread of the data.

5.1.3 Power and Sample Size Analysis

The power and sample size analysis suggests that more data is needed to determine whether the small difference in proportion of general responses is statistically significant. In the data collected, the difference between the proportion of general responses to any two of Task A, Task B, and Task C is less than 0.15. The power and sample size analysis suggests that to determine whether this difference is statistically significant, data from 204 classes would be needed. This means 189 additional classes of data are needed to determine whether the small difference in proportions is statistically significant.

Task Version	Proportion of General Responses
Task A	0.47
Task B	0.59
Task C	0.52

Table 20: Proportion of general responses.

5.1.4 Examining Representation Use with Descriptive Statistics

This section will examine the registers used within the arguments collected. Across the task versions the occurrence of each type of representation were often within a few percentages of each other. The precents are the number of occurrences of that representation out of the total number of tasks of that version. Since the arguments were coded to include every type of representation present in the argument, the total number representations is not necessarily equal to the number of arguments. For example, conforming examples occurred frequently in tandem with other types of representations. The use of conceptual register, structural algebraic notation, and blank or non-examples were the only categories that had variation greater than a few percentages.

Task Version	General Representations				
	Conceptual Register	Structural Algebraic Notation	Non-Structural Algebraic Notation	Informal Symbolic Notation	Generic Example
Task A (58)	21 (36%)	2 (3%)	1 (2%)	0 (0%)	4 (7%)
Task B (64)	28 (44%)	5 (8%)	0 (0%)	1 (2%)	3 (5%)
Task C (48)	19 (40%)	7 (15%)	2 (4%)	1 (2%)	3 (6%)

Table 21: Frequency and percent table for general representations in participant arguments.

Task Version	Non-General Representations		
	Blank or Non-Examples	Partial Examples	Conforming Examples
Task A (58)	12 (21%)	0 (0%)	46 (79%)
Task B (64)	6 (7%)	5 (8%)	49 (77%)
Task C (48)	5 (10%)	3 (6%)	37 (77%)

Table 22: Frequency and percent table for non-general representations in participant arguments.

Task B had by percent the most responses, 44% of the responses (28 of 64), that included the conceptual register. Task C included 40% of responses in the conceptual register (19 of 48). Task A had 36% of the responses including the conceptual register (21 of 58). The arguments were coded to include all the representations present. This means arguments identified as having the conceptual register may also have included another general representation. The percentage of arguments that include the conceptual register as the only general representation of the domain is reported in Table 23. Participants who were given a definition in the conceptual register (Task B) developed more arguments using the conceptual register as the only general representation with 37.5% of arguments using only the conceptual register. In comparison, approximately 29% of responses to both Task A and Task C used only the conceptual register. Of the responses in the conceptual register to Task A, 6.9% included some other general representation. In response to Task B, 10.4% included some other general representation.

Task Version	Conceptual Register	Only Conceptual Register
Task A	36.2%	29.3%
Task B	43.8%	37.5%
Task C	39.6%	29.2%

Table 23: Responses with the conceptual register and only the conceptual

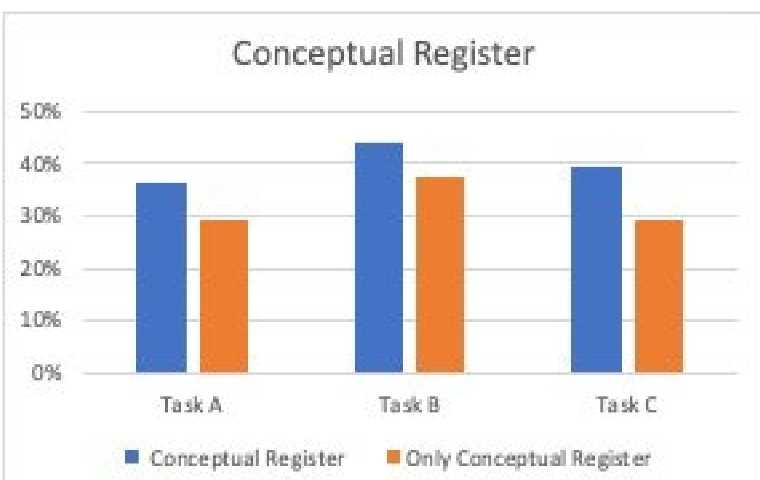


Figure 12: Comparing conceptual register to only conceptual register.

The use of structural algebra occurred with the highest proportion among responses to Task C, where approximately 15% of the arguments used structural algebra. In comparison approximately 8% of the responses to Task B and 3% of the responses to Task A included structural algebra.

By proportion, there were more responses that were blank or contained non-examples among responses to Task A than among the responses to the tasks that included definitions, Task B and Task C. In response to Task A, 21% of the participants included no work or non-examples. For Task B and Task C, 9% and 10% of the responses, respectively, were blank or had non-examples.

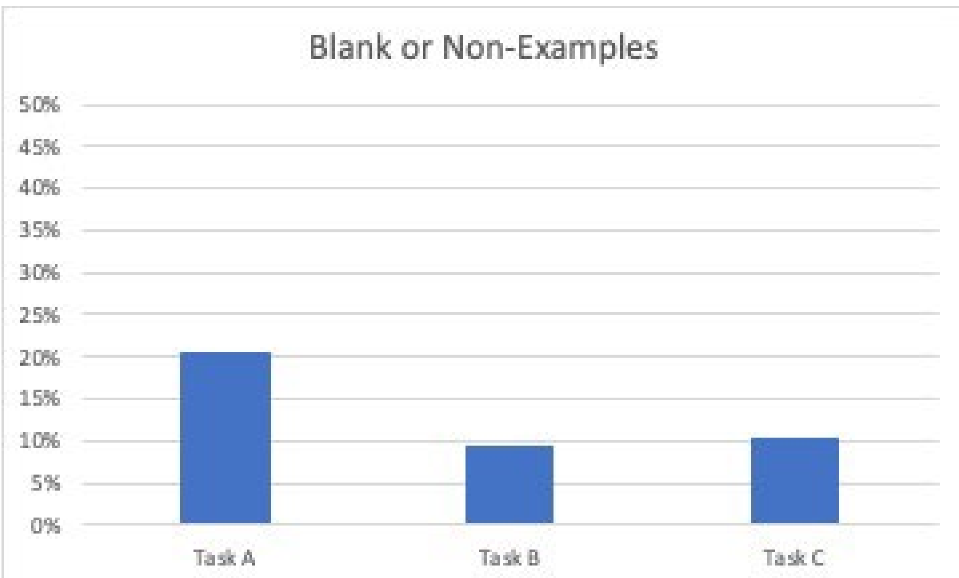


Figure 13: Arguments that were blank or included non-examples.

5.2 Research Question Two

How does access to a general representation or general description of the mathematical objects in the domain of the claim influence the conceptual insights that are used in the argument?

5.2.1 Changes to the Coding Scheme

While coding the data there were seven occurrences of *other viable conceptual insight* and 22 instances of *other non-viable conceptual insight*. Upon examination of the seven occurrences of the viable conceptual insight that did not fit the coding scheme, each occurrence included two components: 1) recognizing that one of the three addends would be divisible by three and 2) a statement that the sum of the other two addends would be divisible by three as well. This conceptual insight is similar to summing remainders; however, these arguments do not reference remainders. They did not have any structural reason for why the sum of the two other addends would be divisible by three. I classified this conceptual insight as viable because it is possible to envision how an argument using this conceptual insight can be adapted to form a viable argument using remainders.

No additional categories were added to the coding scheme to classify the occurrences of other non-viable conceptual insights. The instances of other non-viable conceptual insights did not form clear categories. The 22 instances of non-viable conceptual insights included 9 non-examples demonstrating that the argument was not addressing the domain. These conceptual insights addressed domains such as products of three consecutive numbers and sums of the form $n + n + n$ for some natural number n. Of the 13 remaining instances of other non-viable conceptual insight there was a variety of responses including examining the place value, stating that the first and third addend are three apart, one number in the sum is divisible by three making the sum divisible by three. These responses also included justifications for why the claim is false. None of these conceptual insights reoccurred and so I did not add additional codes.

5.2.2 Examining Viability of the Conceptual Insight with a Randomized Block Design (RBD)

The ANOVA table for the RBD model with the dependent variable viability of the conceptual insight shows that the model does not account for a significant portion of the variation in viability of conceptual insights (F=1.30, p=0.2060>0.05). Furthermore, with both the type I and type III sum of squares, neither the task variety nor the block had a statistically significant effect on the viability of conceptual insight (see Table 24).

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	3.9503	0.2468	1.30	0.2060
Error	153	29.1379	0.1904		
Corrected Total	169	33.0882			

Table 24: ANOVA table for the dependent variable viability of conceptual insight

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Block	14	3.5596	0.2542	1.34	0.1929
Variety	2	0.3907	0.1953	1.03	0.3610

Table 25: Type I Sum of Squares

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Block	14	3.6338	0.2595	1.36	0.1778
Variety	2	0.3907	0.1953	1.03	0.3610

Table 26: Type III Sum of Squares

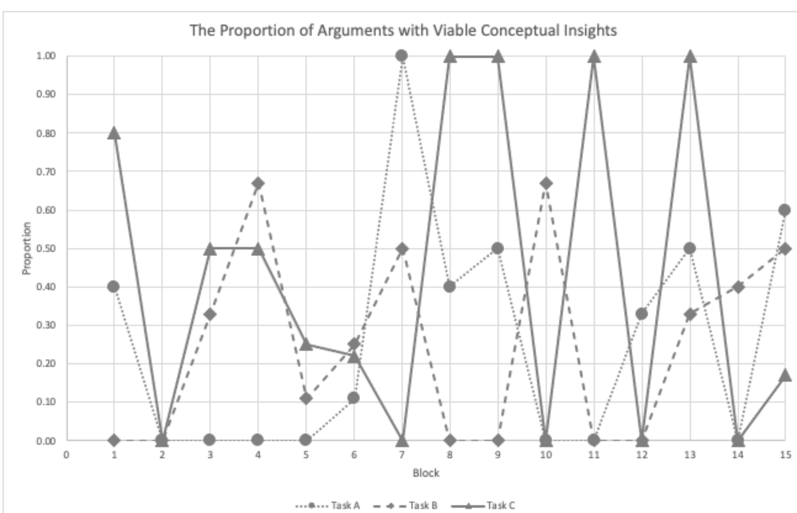


Figure 14: Interaction Plot for Blocks and Viability of Conceptual Insight

Figure 14 and Table 27 show the relationship between the blocks and the viability of the conceptual insight used. The proportions for each task version are calculated by summing the total number of arguments in response to each task that include a viable conceptual insight and dividing that value by the total number of responses collected for that task.

Proportion of Arguments with Viable Conceptual Insights by Block						
Block	Task A		Task B		Task C	
	Number of Arguments	Mean	Number of Arguments	Mean	Number of Arguments	Mean
1	5	0.40	5	0.00	5	0.80
2	4	0.00	4	0.00	4	0.00
3	3	0.00	6	0.33	2	0.50
4	2	0.00	3	0.67	2	0.50
5	8	0.00	9	0.11	8	0.25
6	9	0.11	8	0.25	9	0.22
7	1	1.00	2	0.50	1	0.00
8	5	0.40	5	0.00	1	1.00
9	4	0.50	3	0.00	1	1.00
10	1	0.00	3	0.67	2	0.00
11	1	0.00	1	0.00	1	1.00
12	3	0.33	3	0.00	2	0.00
13	4	0.50	3	0.33	2	1.00
14	3	0.00	5	0.40	2	0.00
15	5	0.60	4	0.50	6	0.17
Mean		0.26		0.25		0.43
Standard Deviation		0.30		0.25		0.41

Table 27: Proportion of arguments with viable conceptual insight by block

Examining the means for each task version, within the data there was a greater proportion of arguments with viable conceptual insights on Task C where participants were given the symbolic definition (see the means in Table 28). The proportion of arguments with viable conceptual insights for participants responding to Task A and Task B were very similar. While this occurred within the data collected, the lack of statistical significance in the difference of means indicates that this variation may be due to chance instead of the variation between task versions.

To examine the effect size, I calculated eta squared for both the variety in task and the blocking factor. The effect size for the version of the task is 0.0095, indicating that the task version was responsible for approximately 1% of the variation in viability of conceptual insight. The effect size for the blocking factor was 0.1098corresponding to the class the data was collected in accounting for 11% of the variation in viability of conceptual insight.

Task Version	Number of Students	Viability of Conceptual Insight	
		Mean	Standard Deviation
A	58	0.2414	0.4279
B	64	0.2344	0.4236
C	48	0.3333	0.4714

Table 28: Mean and standard deviation of viability of conceptual insight

Table 13 shows the means and standard deviation for each version of the task. The coding scheme resulted in codes of 0 for responses with non-viable conceptual insights and 1 for responses with viable conceptual insights. With means very close to 0.5, it is logical that the standard deviation is close to 0.5 as the standard deviation indicates the spread of the data.

5.2.3 Power and Sample Size Analysis

The power and sample size analysis suggests that more data is needed to determine whether the small difference in proportion of responses with viable conceptual insights is statistically significant. In the data collected, the difference between the proportion of general responses to any two of Task A, Task B, and Task C is less than 0.1. The power and sample size analysis suggests that to determine whether a difference of 0.1 is statistically significant, data from 15 classes is needed. However, to determine whether a difference in proportions less than 0.1 is statistically significant data is needed from at least 61 classes.

Task Version	Proportion of Responses with Viable Conceptual Insights
Task A	0.24
Task B	0.23
Task C	0.33

Table 29: Proportion of responses with viable conceptual insights

5.2.4 Examining Conceptual Insight with Descriptive Statistics

Table 30 and Table 31 report the frequency of each type of conceptual insight as well as the percent of the responses to each respective task version. Since the arguments were coded to include every type of conceptual insight present in the argument, the total number of conceptual insights is not necessarily equal to the number of arguments. For example, one student included both symbolic manipulation and redistribution as conceptual insights in their argument. Of the viable conceptual insights, the difference in the percentages across task versions were within a few percentages of each other for all except symbolic manipulation, three copies plus three, and recursive.

For the conceptual insight symbolic manipulation, Task C had the most occurrences by five percentages. On Task C, 8% (4 of 48) of response included the conceptual insight symbolic manipulation. Participants made this algebraic argument on the tasks where they were not given the symbolic definition; on both Task A and Task B, approximately 3% (2 of 58 and 2 of 64) included the symbolic manipulation conceptual insight.

The three copies of three plus three conceptual insight occurred the most with Task C with 8% (4 of 48) of the tasks including this conceptual insight. For Task A and Task B this conceptual insight occurred less frequently with a total of 3 occurrences, 3% (2 of 58) of the Task A responses and 2% (1 of 64) of the Task B responses.

The recursive conceptual insight also occurred most frequently with Task C with 6% (3 of 48) of the response including this conceptual insight. The conceptual insight occurred in 5% (3 of 64) of Task B responses and 2% (1 of 58) of Task A responses (see Figure 15).

Task Version	Viable Conceptual Insights						
	Re-distribution	Symbolic Manipula-tion	Three copies plus three	Recursive	Summing Remainder s	Average	One addend divides and sum of others does too
Task A (58)	2 (3%)	2 (3%)	2 (3%)	1 (2%)	2 (3%)	1 (2%)	2 (3%)
Task B (64)	4 (6%)	2 (3%)	1 (2%)	3 (5%)	1 (2%)	1 (2%)	3 (5%)
Task C (48)	2 (4%)	4 (8%)	4 (8%)	3 (6%)	0 (0%)	1 (2%)	2 (4%)

Table 30: Frequency and percent table for viable conceptual insight.

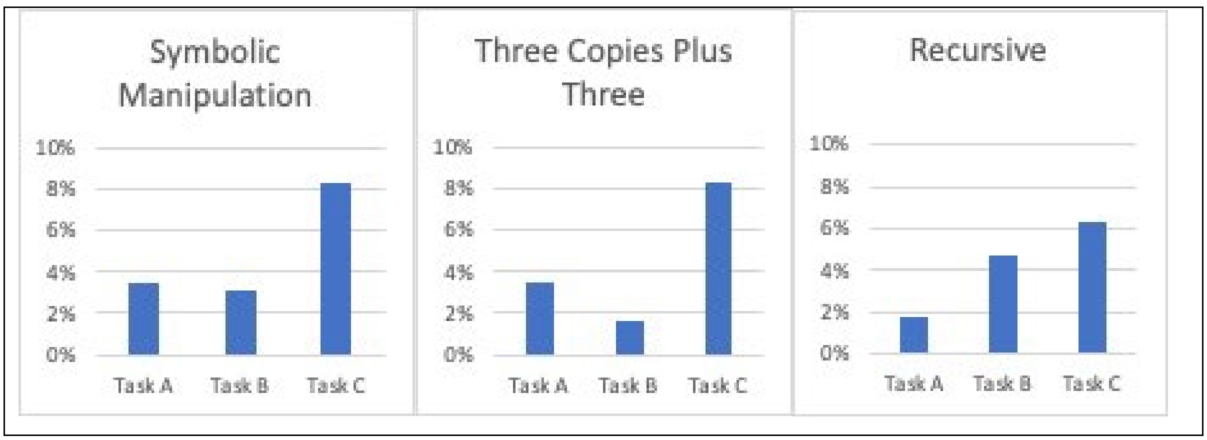


Figure 15: Conceptual insights with the greatest variation across task versions.

Table 31 shows the frequency of each of the non-viable conceptual insights as well as the percent of the total responses to each task these occurrences account for. Comparing the percents the occurrence of even and odd and unclear conceptual insight are within a few percentages of each other across tasks. The occurrence of observed property, empiricism, and other non-viable have a difference of less than 10%. The conceptual insight groups of three was never used in response to Task C but showed up on 5 % of Task A responses and 15% of Task B responses.

Task Version	Non-Viable Conceptual Insights						
	Even and Odd	Groups of Three	Observed Property	Empiricism	Other Non-viable	No Conceptual Insight	Unclear CI
Task A (58)	2 (3%)	3 (5%)	5 (9%)	10 (17%)	9 (16%)	16 (28%)	2 (3%)
Task B (64)	3 (5%)	9 (15%)	4 (8%)	9 (15%)	8 (16%)	17 (27%)	1 (2%)
Task C (48)	1 (2%)	0 (0%)	1 (2%)	5 (10%)	5 (10%)	19 (40%)	1 (2%)

Table 31: Frequency and percent table for non-viable conceptual insights

The largest difference occurred in the category no conceptual insight. For this code, 40% (19 of 48) of the responses to Task C included no conceptual insight compared to only 27% (17 of 64) for Task B and 28% (16 of 58) to Task A. To further examine this category, Table 32 shows the instances where no conceptual insight occurred simultaneously with the code blank or non-examples, partial example, and conforming example. This table shows that most of the responses of this type were responses that were conforming examples with no statements justifying the truth of the claim.

Task Version	No Conceptual Insight	No Conceptual Insight AND Blank and non-examples	No Conceptual Insight AND partial examples	No Conceptual Insight AND conforming examples
Task A (58)	16 (28%)	5 (9%)	0 (0%)	13 (22%)
Task B (64)	17 (27%)	2 (3%)	1 (2%)	14 (22%)
Task C (48)	19 (40%)	4 (8%)	2 (4%)	13 (27%)

Table 32: No conceptual insight and blank and non-examples

5.3 Research Question Three

How do students describe the representations they develop or choose to utilize in their arguments?

5.3.1 Thematic Analysis

The two rounds of thematic analysis resulted in themes that interact with each other. The inductive thematic analysis resulted in four themes:

1. examples as evidence,
2. algebraic representations,
3. shifts in representations, and
4. patterns as a tool to see structure.

The first two themes were broad themes describing instances where participants described their representation as examples or algebraic. Examining the codes and the original excerpts, the purpose of the examples and the way the algebra was described had nuance that was not captured by the overarching categories titled examples as evidence and algebraic representations. The theoretical thematic analysis brought clarity to these inductive themes.

During the theoretical thematic analysis the theme placeholder representation developed two distinct sub themes. There were instances where a representation was used to stand for any object form the domain and the conceptual insight leveraged by the student applied to all of the domain. In these instances the placeholder stands of any object drawn from the domain. In other instances representations were used to describe a general procedure that they have applied to some subset of the domain but the conceptual insight does not extend to all instances of the domain. In these instances the representation is intended to stand for any object from the domain however, either structure of the domain is not used to justify the steps of the procedure or the structure referenced only applies to a subset. I call these placeholder representations for some subset of the domain.

After completing this round of thematic analysis, I found that the themes from the two rounds of thematic analysis interact with one another. The theoretical themes and the additional theme that arose from this framework characterized the purposes of example use as well as the characterizations of the algebraic notation described by participants.

The remaining inductive themes are strongly tied to each other. I observed three shifts in representations during the interviews. Two of these shifts occurred when the participant was asked to explain a pattern. Explaining patterns brought the participant's attention to the structure of the domain when the prompt to prove the claim had previously not done so.

Table 33 details the relationship between each of the inductive themes and the theoretical themes. The four uses of examples that arose within the category of examples fit within the theoretical themes. The two ways in which algebraic representations were described also fit within the theoretical themes. The table shows that patterns were a mechanism facilitating shifts in representations toward more general representations.

		Theoretical Themes			
		Empirical Representation	Placeholder representation for some subset of the domain	Placeholder Representation for any object from the domain	Class Representations
Inductive Themes	Examples as Evidence	Examples used to show truth of claim	Examples used to demonstrate a procedure without justification OR Examples used to demonstrate a procedure that cannot be generalized to the whole domain.	Examples used to illustrate algebra with a particular case	Examples used to demonstrate a transformation
	Algebra Representations			The expression stands for “any” object in the domain OR The variable stands or “any” natural number	The expression represents the domain as a set
	Shifts in Representations				
	Patterns as a tool to see structure				

Table 33: Interaction between inductive and theoretical themes

Examples as Evidence

Examples were used by every participant interviewed. The purpose of the example and what the example was used to represent spanned the theoretical themes. Examples of each will be outlined below.

Examples as Empirical Representations

Jess, Marge, Kendra, and Lucy (pseudonyms) made arguments relying on examples. The participants chose their examples either because it was the example given, they perceived it as random, or it had some structure that they saw as useful or important. Kendra used the example that was given in the definition. She stated,

> I was like looking at the definition provided at the top and it was saying like I didn't really know what natural numbers really meant for some reason, but so I was saying that if they're consecutive, so like four, and they gave the example 4, 5, 6. So that's actually what I use for my first time up there… I just wanted to check with other numbers so I just did like kind of the basic numbers.

Jess, on the other hand, chose what she perceived as three random numbers. She explained,

> I picked three random numbers. My dad one time told me when I was like in first grade that 47 is… the least common number to come up… and so I picked 48, 49, 50. Super easy to add up.

After finding that 48+49+50 was divisible by three, Jess identified this example as having "friendly numbers." In her process of adding the sum she used a commonly taught addition technique of making tens by taking one from 47 to add to 49 so that she could complete the easier sum 47+50+50. She saw this structure as important to the example and chose additional examples she characterized as "not having friendly numbers," meaning no numbers that end in five or ten. Audrey also chose examples to include a particular structure.

She chose two examples "to show that single digit versus multi digit numbers would both work."

None of the participants were confident that their examples showed that the claim was true for all sums of three consecutive natural numbers. When asked, "does your work show that all sums of three consecutive natural numbers are divisible by three?" and following up with the question, "How do you know?" Jess, Marge, Lucy, and Audrey gave clear indication that the examples did not demonstrate truth for the whole domain.

> "Yes, I think so…um. I guess technically not."—Jess

> "I don't know, I feel like you would have to find it some sort of like proof that shows and I don't know what that is… I feel like you have to come up with some sort of proof to prove that any three numbers can do it."—Marge

> "It doesn't show that all numbers are, no, but it does prove that it is a pattern that can be seen at least with numbers extending up to 12. …[because] I did a few examples within that range, and kind of tested it out." —Lucy

> "So the hard thing with that is it would be impossible to show every single example of that because there's infinitely natural numbers."—Audrey

Kendra indicated that the examples are not representative of the domain. However, she did find that confirming three examples verifies the truth of the claim. She stated, "So I felt like it was pretty accurate if I could verify it three times. So, I wrote that I said it's true because there's three natural consecutive numbers as defined and then I added them together, I divided them by three, and then each example was divisible by three. So, this claim is correct." Verifying the truth of the claim and showing that the claim is true for all objects in the domain are not the same for her. This is apparent when she responds to the question about whether her work shows that all sums of three consecutive numbers are divisible by three with, "No, because that would. Um, I would have to do so many numbers. I kind of just did the basic ones where I didn't need a calculator because I could do like 20, 21, 22 but I would have to like pull out a calculator and stuff like that. I probably should have done that just to verify."

Examples to Demonstrate a General Procedure

Nessa used three examples to illustrate a procedure that she views as applying to all objects in the domain. Her procedure has two steps: 1) find the number in the sum of three consecutive numbers that is divisible by three and 2) sum the other two numbers. She claimed that this procedure will result in a sum of two numbers divisible by three which shows that the original sum is divisible by three. She described the procedure using three examples below:

> First, I find the like the number who can directly be like divided by three like 3, 9, 21 here. So, like 3 divided by 3, 9 divided by 3, 21 divided by 3 and 1 + 2 = 3 and here is 7 + 8 = 15 and 22 + 23 = 45 which is also to be able to divide it by three so. Yeah, that's just how I thought.

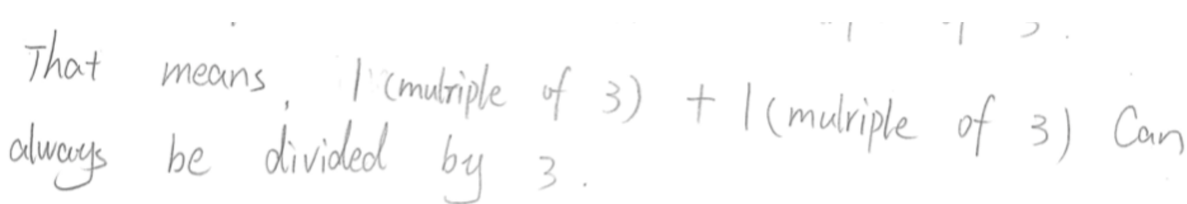


Figure 16: Nessa's representation of the transformed sum

Nessa does not use the defining structures of the domain to justify why her procedure will always work. When asked why the sum of the two remaining numbers is always divisible by three Nessa uses a prior result about divisibility she learned in elementary school. The prior result is that if the sum of the digits of the number is divisible by three than the original number is divisible by three. This prior result is not used to connect structure belonging to all sums of the two remaining numbers. Rather, she uses it to identify that in the examples she works with the sum of the two remaining numbers is indeed divisible by three.

To Nessa, the examples are placeholders for any object from the domain. However, the structure she uses to justify the second step in her procedure is occurring at a case-by-case level. For each example, she sums the digits to determine that the sum of the two addends will be divisible by three. She has not identified a structure of the domain shared by all for why the sum of the digits will always be three. Akin to a generic example, the purpose of her example is to show a procedure that can be applied to all. However, unlike a generic example, the examples do not illustrate structure shared by all objects in the domain. The examples illustrate steps she believes can be applied to any object in the domain to confirm that the particular sum of three consecutive natural numbers is divisible by three.

Examples as Placeholder Representations

Cleo and Winston both chose to supplement their algebraic argument with examples. Cleo plugged seven in for n, simplified the expression, and confirmed that the result was divisible by three. She then stated, “obviously, that works with literally any number.” Winston did not discuss his example. He did however write under his algebra a conforming example.

Examples as Class Representations

In the interview Freya used an example to describe a “leveling” procedure that she discovered while exploring examples and looking for a pattern. She stated,

> I realized that when you would add three consecutive numbers, you could take the number, the larger numbers, like in 5, 6 or 7, you can take seven and you can move one of the numbers from 7 to the 5 to make them all equal. And then that would kind of, which is why we need to divide by three you're gonna always get the middle number.

At this point in the interview, it was unclear if Freya was viewing this as a procedure that can be applied to any sum of three consecutive natural numbers or if she was leveraging the structure of the domain to view this as a transformation that applies to all objects in the domain. When asked if this procedure will always work, she replied,

> I think it'll always work with three consecutive numbers because. They're obviously all going to be like 9, 10 and 11 or one after the other, and so you can always take the top one which is 2 away from the bottom one. And take that extra 1 and put it over to 9 so they're all equal.

In her initial description of her pattern, she gave no evidence of why the procedure applies to all sums of three consecutive numbers. Using Ahmadpour et al.'s (2019) model, this description would indicate that she is in the general procedure state. When asked if her pattern will always work, Freya gave a justification that referenced the structure of consecutive numbers. She identified that they are "one after another" resulting in the "top" number being 2 more than the "bottom" number. This justification gives evidence of abstract structure and aligns with Ahmadpour et al's classification of the representation standing for all objects in the domain. It is evidence Freya sees the structure of consecutive and thus her leveling is a transformation that applies to all objects in the domain. By the coding scheme this is now a class representation.

Interestingly, Freya does not see her transformation as being limited to only sums of three consecutive natural numbers. When asked if her leveling could be applied to other sums to show divisibility by three she responded, "I do think it's possible, but they would have to be like separated the same amount. So if you did like it would have to be like 1, 3, and 5." The purpose of this question was to determine using set equivalence whether Freya intended her generic example to stand for exactly the domain. Instead, the interview question prompted her to imagine a larger domain for which the conclusion holds.

Algebraic Representations

Of the participants interviewed, two students had written algebraic arguments in their initial response. These participants moved fluidly between thinking of the representation as standing *for all* sum of three consecutive natural numbers and standing for *any* sum of three consecutive natural numbers.

Algebra as placeholder representation for any

There were two participants who described their representation as standing for “any”—Winston stated his representation $n + (n + 1) + (n + 2)$ stood for any sum of three consecutive natural numbers and Cleo stated that x in her representation $x + x + 1 + x + 2$ stood for any real number which she later corrected to be any natural number.

Algebra as a class representation

Winston and Cleo both wrote arguments using algebra and indicated that their representation stood for all objects in the domain as a single class of objects. I used two forms of evidence to determine whether they considered their algebraic representation as standing for all. First, they described how the representation connects to the structure of the domain. This satisfies the criteria used by Ahmadpour et al. (2019) to distinguish whether a representation stands for a class of objects. Secondly, they indicated that the algebraic representation stood for the domain as a class. They also indicated that the set of objects that the representation stands for is equivalent to the domain by describing that their representation shows that all objects of in the domain are divisible by three and their notation does not include other objects that are not in the domain.

Winston was given the task version with the algebraic definition, Task C. He used the given definition to write $n + (n + 1) + (n + 2)$. He simplified the expression to $3(n + 1)$ and concluded that the claim was true. In the interview, Winston indicated that the algebraic expression stands for the domain of the claim, standing for all sums of three consecutive natural numbers. He stated,

> [the work shows that all sums of three consecutive natural numbers are divisible by three] because we're using n as the variable, right? So, so long as we follow what's

> given to us. Like if we can say whatever n as starting point and we can say then. Is that is the lowest number of the consecutive, right? Because that's what n stands for in this very specific case. That three of those plus 3 over 3 will always do that, but that's also the same as saying that number, it's next consecutive number and the one directly after that as well. And we can see that because that was the initial line is.

In this excerpt Winston identifies how the algebraic notation connects to the mathematical objects sums of three consecutive natural numbers. In the model by Ahmadpour et al. (2019) they use this as evidence that the student is interpreting the representation as standing for the class of objects, the class of all sums of three consecutive natural numbers. However, Winston is utilizing the "specific cases" for particular values of n to describe how the notation stands for all. When asked if it is possible for his representation to stand for something else, he responded,

> I don't think so…because no matter what, you substitute the, the problem with that is that it there's no other way to interpret it specifically because of the parenthesis. It might have been different if I hadn't parenthesized it, but the fact that each parenthesis is there indicates that. It itself is that n + 1 is an entire unit number.

He clarifies that the "algebraic translation could definitely represent something else…and become any number" referring to the expression once the parenthesis are removed and like terms are potentially combined or partially combined. This indicates that the original expression stands for only sums of three consecutive natural numbers. Thus, Winston is using the algebra as a class representation.

Cleo was given the task version with no definition, Task A, and started her work by writing $x + (x + 1) + (x + 2)$. She combined like terms, simplified to $3(x + 1)$, and identified that the expression is always divisible by three. She gave evidence that she views her representation as standing for all sums of three consecutive natural numbers when asked if her work shows that all sums of three consecutive natural numbers are divisible by three. She responded, “Hopefully, uh, as I did use a variable which you can plug any number in and still get the same result that it's divisible by three.” She went on to clarify that x represents any natural number and “$x + 1$ would be the next number.” This also satisfies the criteria of Ahmadpour et al as evidence of abstract structure indicating, in their model, that her representation stands for the class of consecutive natural numbers.

Cleo also indicated that the sum only represented sums in the domain by stating “it should represent only sums that are three consecutive natural numbers if we’re using the same number for x repeatedly.” Thus, the set that she views the representation standing for is exactly the set made up by the domain.

Shifts in Representations

During their interviews, Freya, Audrey, and Nessa described their written work in a way that did not align with the researcher’s interpretation of their work.

Freya’s written work includes two conforming examples and an objectification of the domain as a set where the sum of first and third addend will always be double the middle number in the sum. From this written work it is unclear if Freya had observed this property as a pattern in her two examples or if she saw the structure of three consecutive natural numbers. In her interview, Freya developed a generic example to explain her thinking:

> I realized that when you would add three consecutive numbers, you could take the number, the larger numbers, like in 5, 6 or 7, you can take seven and you can move one of the numbers from 7 to the 5 to make them all equal. And then that would kind of, which is why we need to divide by three you're gonna always get the middle number. So, I thought that was really interesting, but yeah, my thought process was just kind of like grouping it to make sure it was all level.

In her interview, her first example 5+6+7 is used to demonstrate the “leveling” process that can be applied to all sums of three consecutive natural numbers. She clarified why the process will always work by stating,

> I think it'll always work with three consecutive numbers because they're obviously all going to be like 9, 10 and 11 or one after the other, and so you can always take the top one which is 2 away from the bottom one. And take that extra one and put it over to 9 so they're all equal.”

The property used in her written work, the first and third number are double the middle number, is a consequence of the leveling process as the leveling process results in three copies of the middle number.

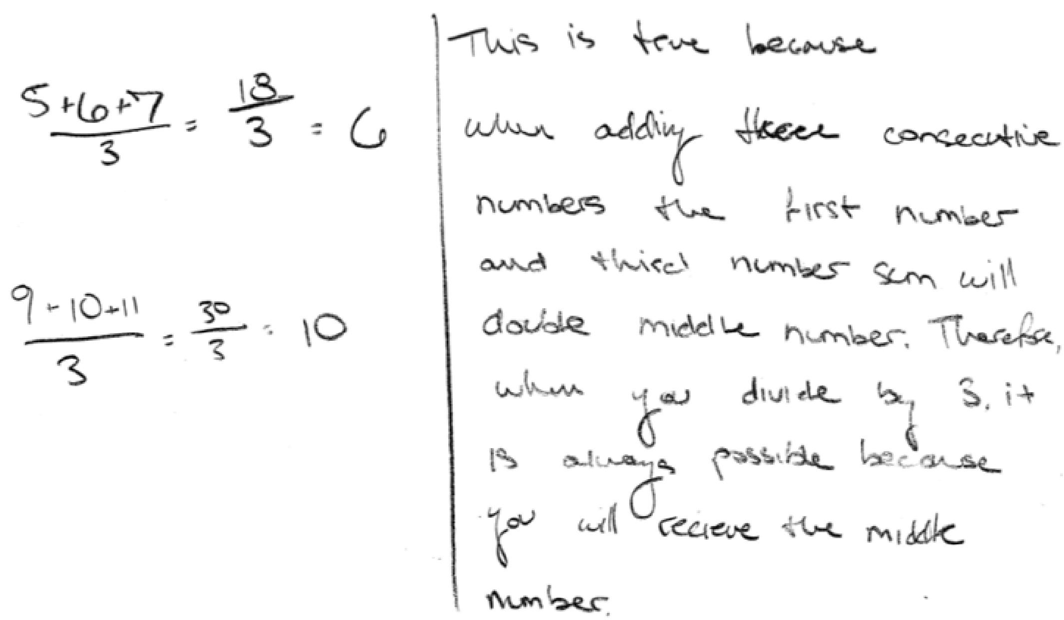


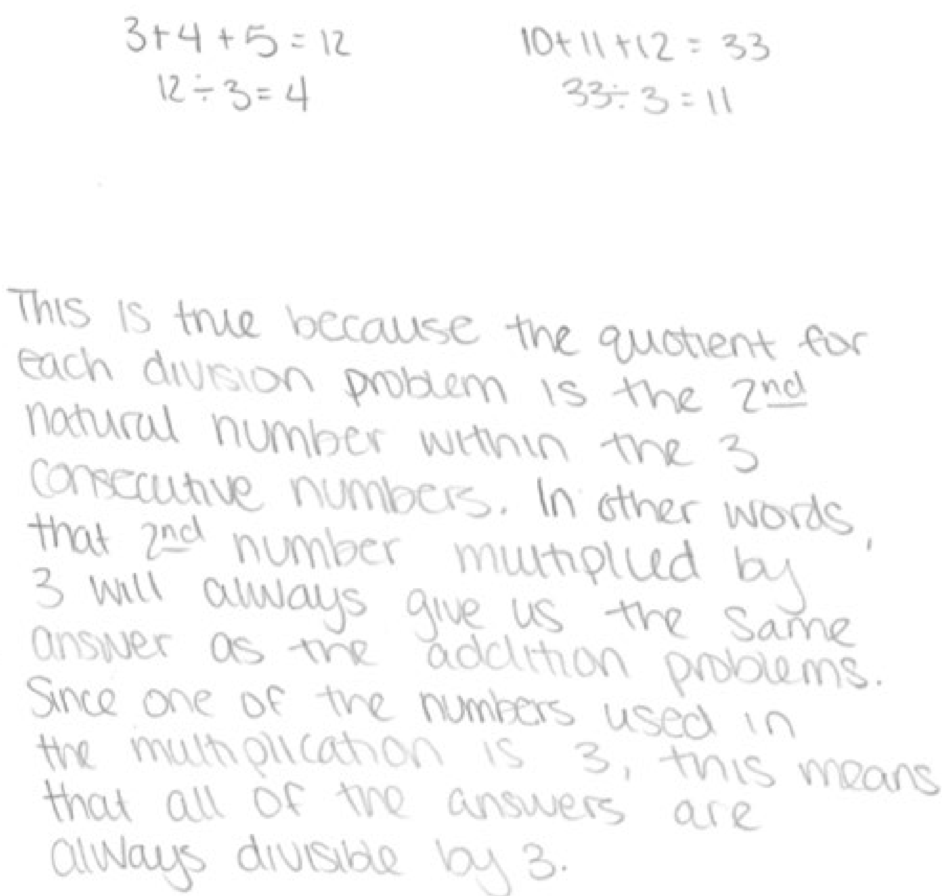
Figure 17: Freya's argument.

Audrey’s written work included conforming examples and the general statement “the quotient for each division problem is the 2nd natural number within the three consecutive numbers." The general statement can be interpreted as an objectification of the domain as a set where the sum divided by 3 is equal to the second addend. However, it can also be interpreted as an objectification of the set of quotients formed when dividing sums of three consecutive natural numbers by three. With either of these interpretations the claim is inherently assumed to be true because for the quotient to be the second natural number it must be an integer making the sum divisible by three. It seems likely then that Audrey observed the quotient was the second number in her examples and is assuming that this property will generalize.

In her interview, Audrey describes that she chose her examples "to show single digit versus multi digit numbers would both work." She described her work further by stating "I discovered when doing both of these equations that my answer in the division… it ended up being the second number in the sequence. [The second number in the sequence] was the same as the quotient from the division problem." When asked whether her work shows that all sums of three consecutive natural numbers are divisible by three she responded "so the hard thing with that is it would be impossible to show every single example of that because there's infinitely [many] natural numbers." She further clarified "I think that it's a pattern within numbers that's consistent throughout the numbers. So that's kind of where I made that educated assumption."

When asked to explain the pattern and how she knows it can be used to justify the argument she responded, "I guess for both of my equations that was the case, but I think it's because it's the average of the three numbers. So, when you take the average of the sum, you're always going to get the mean or the middle number, the median I guess. So, I think that's why the pattern is consistent the whole way through." She stated that the pattern does not work for any sum "it's only with the sum when the three numbers are consecutive and it's divisible by three…when you find the average of the consecutive numbers, it's the median of the consecutive numbers and so that wouldn't be the case if the numbers aren't consecutive."

Originally Audrey's written argument was coded as examples with the conceptual register used to generalize an observed pattern to the entirety of the domain however it is unclear if she is objectifying the domain or the transformation of the domain. In either case the warrant appears circular. In the interview Audrey identifies that the observed pattern is specific to sums of consecutive numbers and uses the notion of average to explain the pattern for all sums in the domain.



3+4+5=12
12÷3=4

10+11+12=33
33÷3=11

This is true because the quotient for each division problem is the 2nd natural number within the 3 consecutive numbers. In other words, that 2nd number multiplied by 3 will always give us the same answer as the addition problems. Since one of the numbers used in the multiplication is 3, this means that all of the answers are always divisible by 3.

Figure 18: Audrey's argument.

Nessa’s written argument uses the conceptual register to objectify the domain as having the property that the domain will always have one addend that is divisible by three and the sum of the other two addends will be divisible by three. Her page has additional markings including parenthesis containing three numbers with two numbers boxed and one number circled. When asked she explained that the circled number was the number divisible by three and the boxed numbers sum to a number divisible by three.

When asked to share her thinking as she completed the task, she described first an algebraic argument. She stated,

> So, when I saw like three natural like continuous national numbers. Um no three consecutive natural numbers so you just like 1,2, 3 or whatever. Like I use like n is the first number and the second is n + 1 and then it's n + 2. So those three that are like consecutive and if I just add them together and plus n + 1 + n + 2. So, it's 3n+3. And as we know that like any number which can like after multiply, not multiply, like adding up if they are like able to divide it by three, then that number is going to be able to divide it by three. So like 3n+3 no matter what n is, it will be able to divide it by three.

When asked if there is anything in her written work that represents the algebraic argument, she stated that, "I just got the examples here I didn't use like n. I used the examples like 1, 2, 3, or 7, 8, 9 or 21, 22, 23 or whatever numbers they are like consecutive." She then describes the process of adding 21+22+23 to get 69 and confirming that 69 is divisible by 3. Describing this she realizes that the process she used in her written argument is different than the algebraic argument. She stated,

> Wait hold up. So, for 21, 22, 23 [it] is actually different. So, the work that I give here is first I find the like the number who can directly be like divided by three like 3, 9, 21 here. So, like 3 divided by 3, 9 divided by 3, 21 divided by 3 and 1 + 2 = 3 and here is 7 + 8 = 15 and 22 + 23 = 45 which is also to be able to divide it by three so. Yeah, that's just how I thought.

She found the addend divisible by three in the sum of three consecutive numbers (3, 9, and 21, respectively) and then found that the sum of the remaining numbers in each sum is also divisible by three (1+2, 7+8, 22+23, respectively).When asked how she knows that the two remaining numbers will sum to a number divisible by three she cited a prior result from elementary school where if the sum of the digits of a number is divisible by three the number is divisible by three.

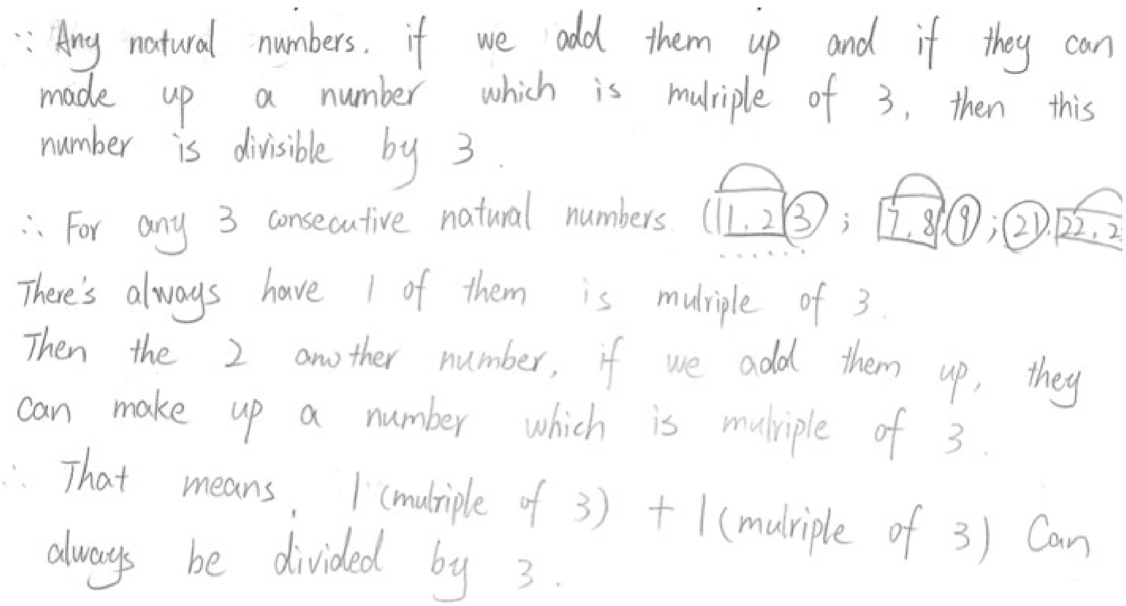


Figure 19: Nessa's argument.

Patterns as a Tool to See Structure

Three of the 9 participants mentioned looking for patterns and or discovering patterns in their examples. Freya, who described a general procedure to transform any sum of three consecutive numbers to get three copies of the middle number, started her interview by stating, “I like to kind of think about things in terms of patterns.” Audrey discovered the pattern that the middle number in the sum is equal to the quotient when the sum is divided by three. Her written argument did not address why this pattern occurs and instead worded it as a property that is true for all sums of three consecutive numbers. In the interview Audrey expressed that it was not possible to show the truth of the claim for all sums in the domain because there are infinitely many of them. However, when instead she was asked about her pattern, she began to describe general mathematical structures that could be applied to all objects in the domain.

Researcher: Does your work here show that all sums of three consecutive numbers are divisible by three? And how do you know?

Audrey: So, the hard thing with that is it would be impossible to show every single example of that because there's infinitely natural numbers. So, it's kind of a hard question to answer. I think that it's a pattern within numbers that's consistent throughout the numbers. So that's kind of where I made that educated assumption, I guess is what I would call it.

Researcher: So, tell me more about that pattern. So, when you stated here that the when for the division problem, the quotient, it's the second natural number and then you kind of use that to justify, tell me more about why that is. How do you know?

Audrey: I guess, um, well, for both of my equations that was the case, but I think it's because it's the average of the three numbers. So, when you take the average of the sum, you're always going to get the mean or the middle number, the median, I guess. So, I think that's why the pattern is consistent the whole way through.

Her attention to the general mathematical structure became more explicit when asked if the pattern extended to other sums. This question elicited her to identify two structures of the domain: 1) addends are consecutive and 2) the number of addends is equal to the number the sum is divided by to test the truth of the claim (in this case 3).

Researcher: OK, does that work for all sums or just special sums?

Audrey: I think it's only with the sum when the three numbers are consecutive and it's divisible by three.

Researcher: OK, why do you think that is?

Audrey: I think that's the case because when you find the average of a set of numbers, you add all the numbers up and then you divide by the number of numbers. So this I think would also work if. Actually, I don't. I don't want to say that because I haven't done the work for it. I was going to say it's possible that it could work for another set of consecutive numbers that's a different amount of consecutive numbers.

Researcher: OK,

Audrey: that I'm not certain that that is the case.

Researcher: OK, what about instead of? If we was instead of 3 + 4 + 5? If it was like 3 + 6, I'm sorry, 3 + 5 + 6. Would it still work?

Audrey: And. No.

Researcher: Why not?

Audrey: Because they're not consecutive. So, when you find that average of when you find the average of the consecutive numbers, it's the median of the consecutive numbers, and so that wouldn't be the case if the numbers aren't consecutive.

Researcher: gotcha. So, there's something about consecutive that makes it so the average and the median are the same.

Audrey: Yes.

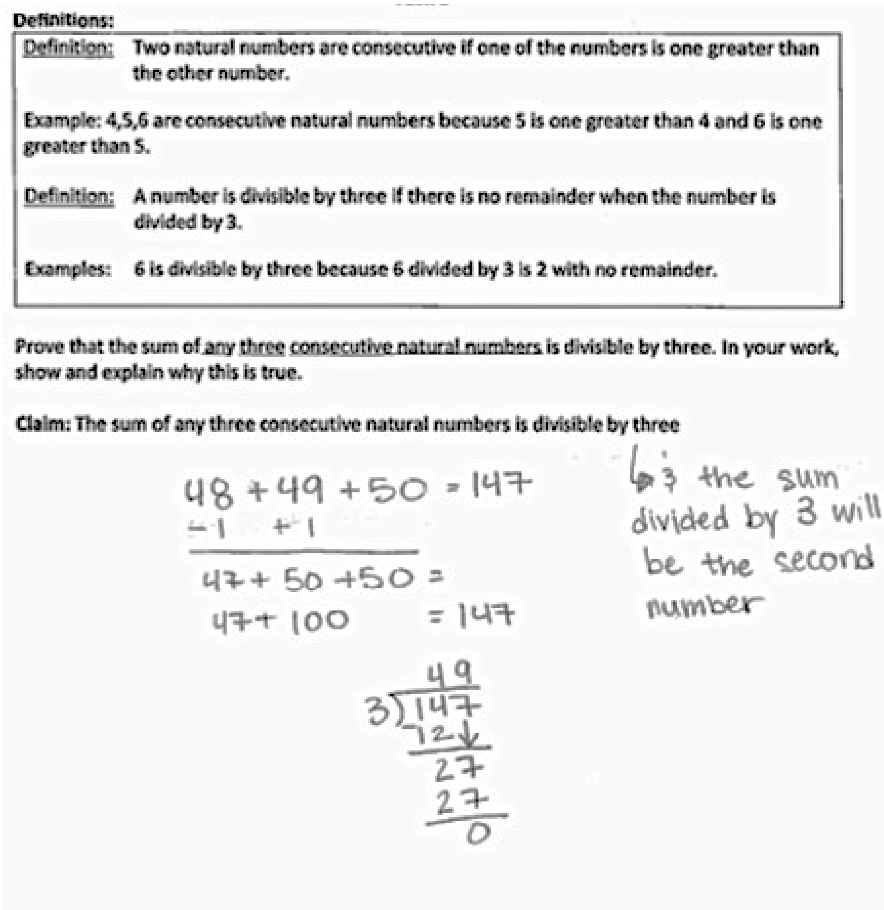
Researcher: Do you have any thoughts on why that might be?

Audrey: Um. Not really.

Jess's written work included one example and a statement that the "sum divided by 3 will be the second number." In her interview she described how she selected her example to be random and how she used adding strategies to redistribute the sum to make finding the sum "super easy." After working the one example she described how she did a few more in her head and then noticed a pattern. She stated,

> I just noted that because I like did a couple other like sets of three in my head like 4, 5, 6 and then like dividing that, I realized that, like the second number will always be the answer after it's divided by three, so I just kind of made a little note of that, but yeah, I don't know.

When asked “any ideas about why that might be?” She responded, “not really. I’m not like a big math whiz…I’m sure it has to do with the fact there are three consecutive numbers and like in doing the math, like adding them up, uh 49, 48, 49, 50 were pretty easy because I could make like 50 out of it. I could make like we called the friendly numbers in our classes like numbers that end in five or zero and so maybe this, like the second one coming up, has something to do with that, but like because they're all so close together.” When asked about the pattern she began to consider consecutiveness and the numbers being “close together.” She began to pay attention to structures shared by all objects in the domain.



Definitions:

Definition: Two natural numbers are consecutive if one of the numbers is one greater than the other number.

Example: 4,5,6 are consecutive natural numbers because 5 is one greater than 4 and 6 is one greater than 5.

Definition: A number is divisible by three if there is no remainder when the number is divided by 3.

Examples: 6 is divisible by three because 6 divided by 3 is 2 with no remainder.

Prove that the sum of any three consecutive natural numbers is divisible by three. In your work, show and explain why this is true.

Claim: The sum of any three consecutive natural numbers is divisible by three

Figure 20: Jess's argument.

5.4 Examining Why the Results of the Quantitative Study were Not Statistically Significant

The amount of variations in the generality of representations and viability of conceptual insight across the task versions had far less variability than expected. The power and sample size analysis revealed that the sample size needed to be bigger to determine whether the small variation that existed was statistically significant. However, I am also left to wonder why there was so little variation in the arguments produced by students given no definition and students given definitions.

Upon completing the statistical analysis and finding that the results were not statistically significant I returned to the data to make sense of why the results were not as anticipated. Below I will detail four factors that may have contributed to the limited variation in generality of responses and viability of conceptual insight. The factors are the span of experiences the participating students have had with mathematics, the prompt given on the tasks, the broadness of the code conceptual register, and the meaning of access to a representation.

5.4.1 Prospective Elementary School Teachers

I chose to use students enrolled in mathematics courses for prospective elementary school teachers because they had met an algebra requirement, but using algebra is not part of the day-to-day mathematical activity of mathematics courses for elementary school teachers. I anticipated that these two criteria would mean that the participants would be able to use algebra if given a symbolic definition for consecutive but that they would not be primed by their current course work to immediately approach the task algebraically. The lack of variation in student performance across the task versions can be partially explained by the variation in participating students' experiences with mathematics and their comfort with algebra.

In the interviews, the span of student prior knowledge and experiences with mathematics was demonstrated. The interviews represent a non-random subsample of the whole data set as the participants who were interviewed volunteered. Disproportionally many of the students who volunteered had completed Task A where they were given no definitions.

To highlight the span of student prior knowledge that exists within the participants, consider Cleo and Lucy. Both completed Task A. Cleo started to develop her argument by writing $x + (x + 1) + (x + 2)$. When asked to share about her thinking as she developed her argument she said, "honestly, when it said any three consecutive natural numbers is divisible by three, what I wrote out like $x, x + 1, x + 2$, that just makes sense to me." When asked if she had seen the claim or the algebraic representation for sums of three consecutive natural numbers before, she responded "that's just how I thought of it. I don't think I particularly, had somebody explain something like this to me before." Cleo is very fluent with algebra, and it was a tool that she reached for immediately when she read the task.

Lucy's experience with Task A is far different from Cleo's because her first step was to use examples to determine the domain of the claim. She started her explanation of her thinking by saying, "I wasn't exactly sure what it was asking, but then I realized like sum that's like the total of something, usually adding." Lucy realized that the claim was referring to adding three consecutive natural numbers by testing what happens when multiplication, subtraction, and addition are performed on three consecutive natural numbers. Through using examples, she determined that the claim is "true" for multiplication and addition.

Lucy may have been distracted from developing a proof by the work of determining what the domain is. Her final response does not include a conceptual insight and appears to be empirical. Her response includes the examples of adding, multiplying, and subtracting sets of three consecutive numbers with check marks indicating when the results is divisible by three and x's indicating when the result is not divisible by three. However, in her interview Lucy stated that "[her] mind went immediately to proofs" after reading the task and she saw her examples not as proof but demonstrating "a pattern that can be seen at least with the numbers extending up to 12." For Lucy a definition and general representation may have drastically changed the argument she produced.

Finally, the number of tasks returned by each class was not equally distributed across the task versions. This suggests that students chose not to participate depending on the task they received. Most classes included a variation of one or two tasks, but three of the fifteen classes had variations of three to four, which is with the small number of student participants resulted in the number of task version C collected being a fifth the size of the number of tasks of version B. On average more tasks of version B were returned than Task A and Task C. Perhaps if all students had returned their tasks, the proportion of arguments that included a general representation and viable conceptual insight for Tasks A and C would be lower.

5.4.2 Prove versus Explain a Pattern

The prompt "prove" has specific meaning within the mathematical community and it is unclear if the participants attribute the same meaning. During the interviews, two of the participants, Jess and Audrey, developed empirical proofs that did not attend to the structure of the domain. However, they both began to attend to the structure of the domain when asked to explain a pattern they had discovered. Both students identified that the result of dividing sums of three consecutive natural numbers by three is the middle addend of the sum. In their written arguments neither student referenced the structure of the domain. Audrey's written argument used the pattern to justify the truth of the claim and Jess's argument included a conforming example and no conceptual insight.

In the interviews, when asked to explain why there is a pattern, both students began to consider the structure of the domain. When asked Audrey responded,

> I think it's only with the sum when the three numbers are consecutive and it's divisible by three… I think that's the case because when you find the average of a set of numbers, you add all the numbers up and then you divide by the number of numbers. So, this I think would also work if. Actually, I don't. I don't want to say that because I haven't done the work for it. I was going to say it's possible that it could work for another set of consecutive numbers that's a different amount of consecutive numbers.

In her responses she identified that dividing the sum by three is finding the average. She initially states this will only work for sums of three consecutive natural numbers but then she wonders if it is possible for other "sets of consecutive numbers." When Audrey described another set of consecutive numbers being "a different amount of consecutive," she may have been attending to the structure that each number was 1 apart from the previous number. She may have observed that when each number is a set amount apart from the previous number, the resulting sum is divisible by 3. With that interpretation, the question about a pattern resulted in Audrey noting that the sum divided by three is the same as finding the average because the number of addends is three, the divisor is three, and the first and third addends are spaced an equal amount form the second addend. These structures could be leveraged into a viable argument in the future.

Jess gave the following response when asked about the pattern she found.

> Uh, not really. Not like a big math whiz, but um, yeah, I. Yeah, I'm sure it has to do with the fact that there are three consecutive numbers, and…so maybe this, like the second one coming up, has something to do with that, but like because they're all so close together.

In considering the pattern, Jess notes that there is something about the conditions "three numbers" and "consecutive" that cause the pattern and that the numbers are "all so close together." These observations are not explicit references to the structure of the domain and would perhaps take more support to become a viable argument, but nevertheless Jess originally did not attend to structure and after discussing her pattern she began to consider the structure of the domain.

These two interviews suggest that the prompt "why is there a pattern?" prompted different mathematical activity than the prompt "prove." When asked to explain a pattern, both students attended to the structure of the domain when they had not previously done so.

5.4.3 Unpacking the Conceptual Register

The code conceptual register in application was very broad. The code was described as follows: The domain is described as an abstract class. Properties (correct or incorrect) are attributed to that set as a class (see Table 3). In application this code arose almost any time a student wrote a sentence to explain their thinking. It captured descriptions of the domain that include a wide range of properties. In Table 34, I include some categories that arose when I went back and examined the conceptual register code.

The first category is when the structure identified to the sums of three consecutive natural numbers stems directly from consecutiveness and summing three numbers. Of the 68 arguments that used the conceptual register, 32 fell into this category (47%).

The next category is when the properties stem from consecutiveness and summing three numbers but there is some intermediate transformation or result applied that is not described. Of the 68 arguments using the conceptual register, 7 (10%) fell into this category. This, for example, came up in Freya's interview. Freya justified the truth of the claim by stating "…when adding three consecutive numbers the first number and the third number sum will be double the middle number." Originally when I read this, I assumed she had observed a pattern in her examples. However, in the interview she very clearly explained how the sum is three copies of the middle number and connected consecutiveness to the transformation that redistributes the original sum of any three consecutive natural numbers into the sum of three copies of the middle number. With her further clarification in the interview, the structure described in the written argument appears to not be an observed pattern. Rather, they are a consequence of the redistribution she was preforming but had not explicitly stated.

The observed structure does seem to be a relevant category for conceptual register. It arose when students stated the claim was true and then cited a property of the results of actions on the domain. For example, Audrey attributed the property that "… the quotient for each division problem is the 2^{nd} natural number within the 3 consecutive numbers." In her interview Audrey called this property a pattern. She observed the pattern after completing several examples and noting that when sums of three consecutive numbers are divided by three, the result is the middle number of the sum. This is a structure of the domain that already assumes the conclusion as in the quotient (the sum divided by 3) is the 2^{nd} number, which is by definition a natural number, meaning this structure already assumes that the sum is divisible by three. Of the 68 arguments coded as using the conceptual register, 6 characterized the domain as having the structure that the sum divided by three is the middle number.

Non-pertinent structures were used 20 times to describe the domain. This category was made of instances of the domain being characterized by either the number of addends in the sum, odd and even structure and/or explicit restatements of the conditions of the claim. These structures are not unique to the objects in the domain and end up characterizing a much larger superset of the domain. Finally, there were three instances where the structure was unclear.

Types of Structure	Example Excerpts	Number of Occurrences	Percent of total Conceptual Register Codes
(1) Structure of consecutiveness and sums of three numbers is described	"The sum of any three consecutive natural numbers is divisible by three because one number will be a multiple of three (or zero), one will be a multiple of three (or zero) plus one, and one will be a multiple of three (or zero) plus two." "Each consecutive natural number sequence adds 3 to the previous consecutive natural number sequence. Every sequence of 3 consecutive numbers increases by 3 total when each individual number is increased by 1."	32	47%
(2) Structure is related to consecutiveness and sums of three numbers	"…each consecutive number increases the total by 3, making it divisible by 3." "The average of the three consecutive natural numbers will always be the middle, or the second natural number." "…when adding three consecutive numbers the first number and the third number sum will be double the middle number."	7	10%
(3) Structure is an observed pattern in the result	"… the quotient for each division problem is the 2nd natural number within the 3 consecutive numbers." "The sum of any three consecutive natural numbers is divisible by three because of the middle number. Adding the first number and the middle number gives us a number divisible by 3. Then the final number keeps it like that."	6	9%

Table 34: Variation in conceptual register

Table 34 Continued

(4) Structure applies to a much larger superset of the domain	"The sum of any 3 consecutive natural numbers is divisible by 3 because there are 3 numbers to begin with." "…because they are consecutive and there's three numbers…" "…because adding a number 3 times is the same as multiplying it by 3."	20	29%
(5) Structure is unclear	"… because from the first number to the third number, they are 3 apart." "…because in the multiplies of 3 there is every number."	3	4%

The original purpose of the code for conceptual register was to indicate whether the participant was using words to describe the domain in general. In practice, when interpreting student language use it is less clear whether students are intending to describe a class of objects using shared properties or using language that is general to observe structures that they see in their explorations. This nuance of whether the structure described is viewed by the participant as existing in the domain or as a byproduct of actions taken on the domain can help distinguish whether the student is intending the describe the domain using shared structure or is observing structure in their empirical data. Using this lens property, Types (1) and (2) lead to general characterizations of the domain. Type (3) instances of the conceptual register are describing a set that is the domain with some action taken on it. The structure is no longer inherent to the domain but rather the domain after undergoing some action. Type (4) characterizes a large superset of the domain. Two supersets that were characterized include the set of all sums of numbers of the forms $odd + even + odd$ and $even + odd + even$ and the set of all sums of three numbers. These sets have too little in common with the domain to be productive for demonstrating the truth of the claim.

What can be seen in the data is that the participants who were given tasks with definitions used the structure of "consecutiveness" and "sums of three numbers" or structures related to consecutiveness and sums of three numbers more often in their arguments. For

Task A approximately 48% (10 out of the 21), of the instances use of the conceptual register was type 1 of type 2 where the structure of consecutiveness is explicit. For comparison, for task B approximately 54% (15 of 28) and for Task C approximately 74% (14 of 19) of the responses in the conceptual register included this level of detail to the structure of the domain.

Task Version	Type 5 Structure is unclear	Type 4 Structure applies to a much larger superset of the domain	Type 3 Structure is an observed pattern in the result	Type 2 Structure is related to consecutiveness and sums of three numbers	Type 1 Structure of consecutiveness and sums of three numbers is described	Total in Conceptual Register
Task A	0	8	3	2	8	21
Task B	1	9	3	3	12	28
Task C	2	3	0	2	12	19

Table 35: Types of descriptions in the conceptual register

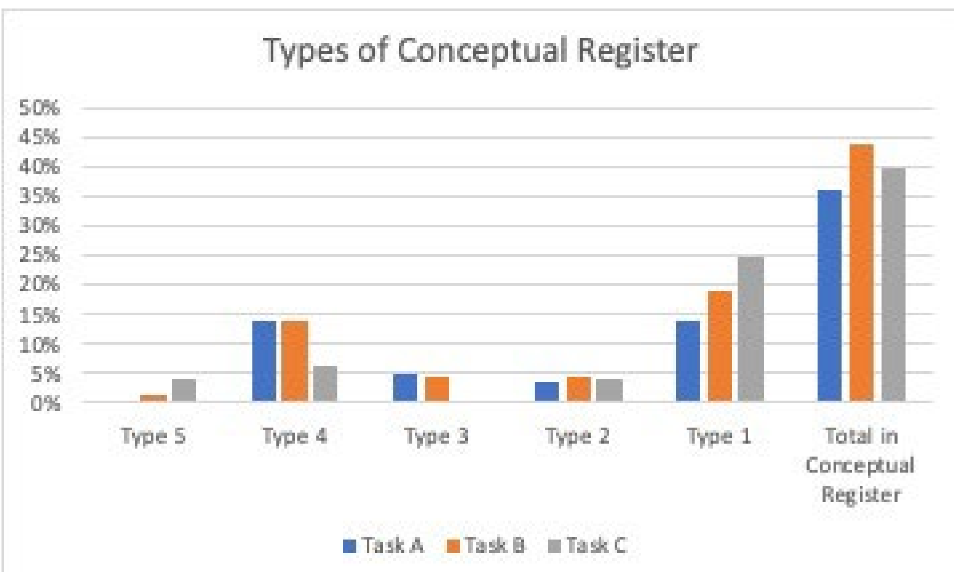


Figure 21: Types of descriptions in the conceptual register

5.4.4 Access to Representations

Part of the hypothesis was the assumption that giving a definition at the top of the task would give the participant "access" to the definition. I had assumed that a significant portion of the students given a task with a definition would read the definition, interpret it as a general representation of the domain, and use the representation to see and operate with the structure of the domain. Since the proportion of responses that contained general representations and viable conceptual insights was similar across task versions, there is more to participants using and identifying structure in a representation than having the definition given to them.

The participants given definitions did leave their responses blank or include non-examples less often compared to the participants who were given no definition. The percent of tasks that were blank or included non-examples was similar for the two version of the task that include definitions, with about 7% for Task B and 10% for Task C. This is lower than the number for Tasks A which was about 21%. This suggests that the definitions did support some students to correctly develop examples of objects in the domain.

Of the 48 participants that were given a symbolic definition only 7 used it to write the sum of three consecutive numbers algebraically and only 4 of these 7 simplified the expression to show that the sum is divisible by three. This means of the 48 students who were given the symbolic definition only 4 used it to develop an algebraic argument for the claim.

Examining responses to Task C suggests that for a student to use a representation as a technical handle representing the domain they must 1) see the representation as general, 2) identify structure of the domain within the representation that they can use to develop an argument, and 3) see the representation as useful in communicating the conceptual insight they have developed. To illustrate this, consider the following three responses to Task C. Figure 24 includes a structural algebraic representation of the domain that is set equal to 3k there is then an arrow drawn to the simplified equation 3n+3=3k followed by another arrow pointing to the equation n=k. The participant then wrote "This is false. Only certain groups of 3 consecutive numbers can be divided by 3." They also include 2 conforming examples. Instead of noticing 3n+3 is divisible by three, they simplified incorrectly, to n=k, which may have led them to determine that the claim is false except for when n is equal to k. This participant was able to use the symbolic definition to develop a general algebraic representation of the domain however it seems that the general representation they created did not make the structure of sums of consecutive natural numbers visible to them. In other words, they manipulated the algebraic expression, but it was never a technical handle that they saw as expressing a structure belong to the domain that they could use to show the claim was true.

In Figure 22 the participant includes two sets of three consecutive natural numbers generated by inputting natural numbers into $n, n+1, n+2$. The sets are then summed and the divisibility by three is confirmed. They then state that "this is true because three numbers that are the exact same and are added together will be divisible by 3 because there are 3 groups of the same number. When doing 3 consecutive numbers you are just adding another group of 3." This participant does not see the notation $n, n+1, n+2$ as operable. In their statement they identify the following structures as belonging to the domain: 1) "three numbers that are exactly the same" and 2) "another group of three." It seems these structures that can be seen in the notation $n, n+1, n+2$ did not become visible to the participant until they had used the notation to generate particular examples. It seems likely that for this participant, the symbolic definition eventually gave them access to the conceptual insight they used because this conceptual insight is far more visible when examining examples of the form 5, 5+1, 5+2 than when examining examples in the form 5, 6, 7. However, they did not see the symbolic definition as a technical handle that could represent the structure of the conceptual insight they discovered.

In Figure 23 the response includes examples that on the first line have the sum of three consecutive natural numbers. On the second line the sum is rewritten in the form of the symbolic definition. On the third line the sum is manipulated to show redistributing to form three copies of the middle number still shown with the same structure as the symbolic definition. Then on the final line the sum is simplified to be three copies of the middle number. The participant includes an explanation the uses the notation $n, n+1, n+2$ but never forms the algebraic expression $n+n+1+n+2$. Instead, they explain that 2 can be broken into 1+1 and so the 2 can be "distributed" to make any sum of the form $4+1, 4+1, 4+1$. This participant also did not operate on the symbolic definition. They chose not to show distributing the 2 by writing $n + 1, n + 1, n + 1$. The symbolic definition seems to have given them access to the structure of consecutive to discover a conceptual insight. They did not, however, view the symbolic definition as a usable technical handle and instead chose to use the conceptual register and generic examples to form a general argument.

From this small subsample we see three examples of the symbolic definition being used but not being a technical handle for the student to communicate a conceptual insight. In one case the algebraic representation created using the definition does not seem to communicate general structure to the student. In the other two cases the definition seems to support the student to identify general structure and yet they seek other ways to communicate the structure and do not use the definition to develop a technical handle.

This shows that access to the definition in the form of having the definition presented at the top of the task is not sufficient. There are three related ingredients there are necessary: 1) the student sees the representation as general, 2) the student identifies structure of the domain within the representation that they can use to develop an argument, and 3) the student sees the representation as useful in communicating the conceptual insight they have developed.

B05 T01

Task C

Definitions:

Definition:	Three consecutive natural numbers are numbers that can be represented as n, n+1, n+2 for some natural number n.
Example:	4, 5, 6 are three consecutive integers because they can be written as 4, 4+1, 4+2
Definition:	A number m is divisible by 3 if it can be written as $3 * k$ for some integer k
Example:	6 is divisible by three because $6 = 3 * 2$. This means $6 = 3 * k$ when $k = 2$.

Prove that the sum of any three consecutive natural numbers is divisible by three. In your work, show and explain why this is true.

Claim: The sum of any three consecutive natural numbers is divisible by three

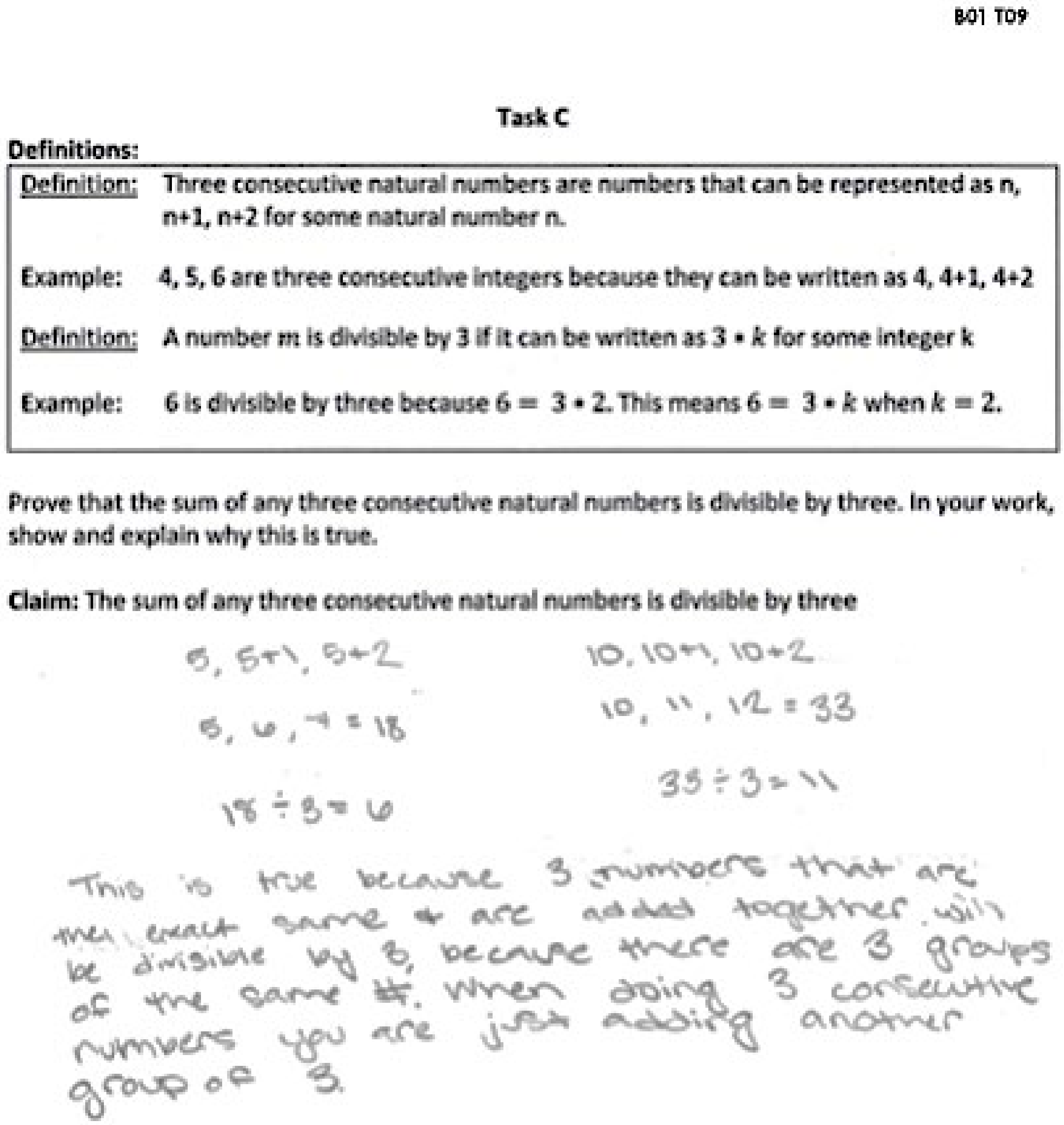
$n + (n+1) + (n+2) = 3 \cdot k \rightarrow 3n + 3 = 3 \cdot k \rightarrow n = k$

This is false. Only certain groups of 3 consecutive numbers can be divided by 3.

$4 + 5 + 6 = 15 \rightarrow 15 = 3 \cdot 5$ ✓

$8 + 9 + 10 = 27 \rightarrow 27 = 3 \cdot 9$ ✓

Figure 22: Response to Task C



B01 T09

Task C

Definitions:

Definition:	Three consecutive natural numbers are numbers that can be represented as n, n+1, n+2 for some natural number n.
Example:	4, 5, 6 are three consecutive integers because they can be written as 4, 4+1, 4+2
Definition:	A number m is divisible by 3 if it can be written as $3 \cdot k$ for some integer k
Example:	6 is divisible by three because $6 = 3 \cdot 2$. This means $6 = 3 \cdot k$ when $k = 2$.

Prove that the sum of any three consecutive natural numbers is divisible by three. In your work, show and explain why this is true.

Claim: The sum of any three consecutive natural numbers is divisible by three

5, 5+1, 5+2

5, 6, 7 = 18

18 ÷ 3 = 6

10, 10+1, 10+2

10, 11, 12 = 33

33 ÷ 3 = 11

This is true because 3 numbers that are the exact same & are added together will be divisible by 3, because there are 3 groups of the same #. When doing 3 consecutive numbers you are just adding another group of 3.

Figure 23: Response to Task C

B11 T02

Task C

Definitions:

Definition: Three consecutive natural numbers are numbers that can be represented as n, n+1, n+2 for some natural number n.

Example: 4, 5, 6 are three consecutive integers because they can be written as 4, 4+1, 4+2

Definition: A number m is divisible by 3 if it can be written as $3 * k$ for some integer k

Example: 6 is divisible by three because $6 = 3 * 2$. This means $6 = 3 * k$ when $k = 2$.

Prove that the sum of any three consecutive natural numbers is divisible by three. In your work, show and explain why this is true.

Claim: The sum of any three consecutive natural numbers is divisible by three

6+7+8 = 21
6+(6+1)+(6+2)
(6+1)+(6+1)+(6+1)
7+7+7=21

1+2+3 = 6
1+(1+1)+(1+2)
(1+1)+(1+1)+(1+1)
2+2+2=6

We know 3 consecutive numbers are written as n, n+1, n+2. Because 2 is an even number, it can be separated into 1+1. Because this is simple addition, it does not matter where the +1 is within the 3 natural numbers. Therefore, if I have; 4+5+6, I know 6 is 2 more than 4. This concept is true of any 3 consecutive natural numbers. Therefore 4, 4+1, 4+2 is equivalent to 4+1, 4+1, 4+1, since all I did was distribute the 2. This turns 4+5+6 into 5+5+5 which all equal 15.

Figure 24: Response to Task C

Chapter 6: Discussion

6.1 Motivation for the Study

In 2019 I hypothesized that students would be more likely to develop arguments that were structural or procedural, as defined by Ahmadpour et al. (2019), if there was or was not, respectively, a general representation for the domain readily available to them. I observed Preston (pseudonym), a grade 8 student, think aloud as he developed arguments to two claims. The first claim was “all sums of three consecutive natural numbers are divisible by three.” What I observed was Preston spent most of the interview working to develop a general representation to represent all sums. An algebraic representation was not readily available to him, so he spent the interview working to invent a way to represent consecutiveness using dots along a diagonal. He drew three dots on a diagonal with arrows with +1 and -1 to indicate moving the highest dot on the diagonal down vertically to be in an alignment horizontally with the middle dot and moving the lowest dot up vertically to be in alignment with the other two dots.

It became apparent to me that for a student well versed with algebra who has the representation $x + x + 1 + x + 2$ available to them, the process of developing a proof to this claim involves simplifying an algebraic expression and interpreting the simplified expression as a multiple of three. Once you have the algebraic expression it can be operated on and remain completely decontextualized. To write a proof that will be readily accepted, a student does not need to interpret the meaning of the representation along the way or keep track of any properties belonging to the original object. In many classes teachers would be satisfied to see the expression simplified to $3(x + 1)$ and a statement that says the claim is true. However, to a student without an algebraic expression, the task of developing this general argument involves much more. A student who instead must either develop their own general representation or use words to describe general properties has a much larger task. Without an algebraic definition, a student must develop a technical handle they can use to communicate the structure of the domain pertinent to the claim and discover a viable conceptual insight that uses that structure to connect the conditions to the conclusion of the claim.

I then became interested in how a student’s argument would be influenced by the student having access to a general representation. I was interested in 2 specific characteristics of the argument: 1) the representations and 2) the conceptual insight. The representations interested me because I was curious whether access to a definition with a description or an algebraic representation would influence the generality of the student's argument. Would students be more likely to represent or describe the domain in general if they were given a definition? I hypothesized that they would based off my interview with Preston and then a study by Weber et al. (2020). First, for Preston, finding a general representation that he confidently felt stood for all objects in the domain seemed to be the most challenging part of developing a general argument. Second, Weber et al. (2020) found that perceived cost, which includes cost in time, as well as perceived likelihood of finding a deductive proof were factors considered by students when deciding to develop a deductive proof or an empirical proof. Using this finding to contextualize what happened in my 2019 interview with Preston, I observed that while Preston did persevere to develop a deductive general argument, the brunt of his work was seeking a general representation, which he saw as pinnacle to his ability to develop the deductive argument. There was a high time cost for him to develop that argument because he did not have access to a general representation of the domain. This led me to hypothesize that students who do not have a readily available general representation may be more likely to develop empirical arguments. This hypothesis is situated within a growing body of research that is shifting away from judging students who produce empirical arguments as fundamentally believing that a few examples are sufficient to show the truth of a general claim. Instead, in recent years researchers have sought to consider the conditions that lead students to producing empirical versus deductive arguments.

The conceptual insight used in the argument was also of interest to me because I hypothesized that the representation available to a student would influence the structures of objects in the domain that the student was attuned to. While all viable conceptual insights rely in some fashion on the structure of consecutiveness and that the objects are sums of three numbers, the way these structures show up in the conceptual insight seems to reveal how these structures are seen by the students. For example, one student may make the argument by noting that every sum of three consecutive natural numbers can be redistributed into the sum of three copies of the middle number. This transformation on all objects of the domain involves the student identifying that the first addend is one less than the middle addend and the last addend is one more. In contrast, if another student notes that the claim is true because summing three consecutive natural numbers is the same as summing three copies of the first addend plus three, they are viewing consecutive as the first number, plus the first number plus one, plus the second number plus one which is equal to the first number plus two. I can imagine that if a student is given the symbolic definition $n, n+1, n+2$ for consecutive natural numbers they may be more likely to develop an argument noting that the sum is three copies of the first number plus 3. It is less intuitive perhaps for a student with this definition to redistribute to three copies of $n+1$.

Conceptual Insight	**Structure of the Domain attended to**
Redistribution	First number is one less than the middle number and the third number is one greater than the middle number.
Symbolic Manipulation	No explicit structure of domain needed once algebraic expression is generated.
Three copies plus three	Second number is one greater than the first number and the third number is 2 greater than the first number.
Recursive	$0+1+2=3$ which is divisible by three. Every sum of three consecutive numbers is of the form $(0+n)+(1+n)+(2+n)$ for some natural number n.
Summing remainders	One number in every set of three consecutive natural numbers is divisible by three. The remainders of the two remaining numbers divided by three will be 1 and 2 respectively.
One number divisible by three other two sum to number divisible by three	One number in every set of three consecutive natural numbers is divisible by three.
Average	The number of addends is equal to the number that is being used to divide.
Even and odd	Every sum of three consecutive natural numbers will be of the form $odd + even + odd$ or $even + odd + even$.
Groups of three	There are three addends, and the divisor is three.
Division mix-up	Unclear.
Observed property	The result of dividing the sum of three consecutive numbers by three is the middle addend of the sum. OR The first and last number sum to twice the middle number.
Empiricism	No general structure.

Table 36: Descriptions of the structure behind each conceptual insight.

The work of Ahmadpour et al. (2019) and their model for how students read proof also sits in the foreground of this work, as in extending their model to how students develop or write proof lies the question: Do procedural proofs as defined by Ahmadpour et al. occur when students are writing proofs? There are two key distinctions between procedural and structural proofs. The first distinction lies in the representation, i.e., does it stand for all? Or is it a placeholder for any? The second is whether there is evidence that the student has abstracted the structure of the domain. Ahmadpour et al. used abstract structure as evidence for the representation standing for all objects in the domain. They classified students as having abstract structure when they connected the algebraic representation to the structure of the domain. However, it is unclear when the model is extended to students developing arguments if using representations as placeholders and abstraction are mutually exclusive. So, I was interested in the interviews in understanding how students intend their representations. I looked for whether they saw their representation as standing for a class of objects, a placeholder, or an example.

In the following sections I discuss how my findings situate within existing literature and inform future research. First, I will discuss how my findings suggest another factor, student generalization about the domain, as contributing to student proof construction. Next, I will discuss how the consideration of how students intended their arguments underscored the distinction between student generalization of observed results and generalizations of procedures. In particular, how this distinction ties to research on transitioning from informal argument activity to viable arguments. Thirdly, I will discuss how an exploration of the role of examples and algebraic representations supported a need to focus instead on the structure students are attending to. I will conclude this chapter by discussing limitations and avenues for future research.

6.2 Definitions Without Generalization

My finding suggest that the inclusion of definitions does not have a high impact on student development of empirical arguments with the population of prospective elementary school teachers and the mathematical content of basic number theory. This builds on the work by Weber et al. (2020) by indicating that presence of definitions does not significantly influence the factors they identified: students' perceived cost and their perceived likelihood of success in developing a deductive proof. There are other factors at play that determine whether the student utilizes the definition and the register it is presented in to construct an argument.

One of these factors may be whether students have generalized about the domain. Ellis (2007) adapts Kaput's (1999) definition to define generalization as "one of three activities: 1) identifying commonality across cases, 2) extending one's reasoning beyond the range in which it originated, or 3) deriving broader results from particular cases" (p. 197). Utilizing the symbolic definition involves a conversion from the conceptual register in which the claim is presented to the symbolic register. Duval (2006) identified that performing conversions is a challenging activity for students. Key to student success in converting from one register to another is identifying the common structure of the mathematical objects that stays consistent across the differing forms the object takes in different registers (Duval, 2006). This description of what conversion entails aligns directly with the first of the three generalization activities described by Ellis (2007). This indicates that identifying structure of the domain within an algebraic representation, viewing an algebraic representation as general, and having performed a generalization about the domain of the claim are one and the same.

Generalizing about the domain of the claim is a prerequisite to developing a general direct proof for a claim with an infinite domain that is not given enough attention. Existing research on generalization may provide insight into how to support student proof construction. For example, upon examining actions in a seventh-grade classroom, Ellis (2011) found encouraging students to justify and clarify promoted students to generalization. She defined these actions to include asking students to "clarify a generalization, describe its origins, or explain why it makes sense" (p. 316). This aligns with the findings from my qualitative study where I found asking students to explain patterns they had discovered supported them to identify structure of the domain, i.e., generalize about the domain, when they had not previously attended to shared structure of the domain. In fact, in the interviews students who developed general arguments with viable conceptual insights had one of two experiences: the student had already generalized about the domain, or they strategically used examples to identify the pertinent structure belonging to the domain that allowed them to generalize and develop their argument.

If the mechanism for converting from the conceptual register to the symbolic register is as Duval (2006) suggests, then generalization is a key mechanism in the choice to adopt an algebraic representation from an available definition as a technical handle. Sandefur et al. (2013) and Raman (2003) found that the discovery of a conceptual insight and technical handle does not occur in a particular order and that a student can discover one without finding the other. In my study there were arguments collected where students used the algebraic representation to generate examples and ultimately identify a conceptual insight but chose not to utilize the algebraic representation as a technical handle. They instead used examples that followed the form of the algebraic representation (ex. $4 + (4 + 1) + (4 + 2)$) to develop generic examples. These students did not volunteer to be interviewed so I can only hypothesize their approach to developing their arguments. Their written argument indicates that the algebraic representation likely is serving as a placeholder representation standing for a procedure that can be used to generate objects in the domain. Thus, they utilized the algebraic representation to generate examples and were able to discover a conceptual insight. However, the structure was not identified in the algebraic representation itself, or if it was the representation was not viewed as a means to communicate the conceptual insight.

A student utilizing an algebraic representation to represent the domain of a claim as a class of mathematical objects necessitates that they have generalized. This means that a student's ability to construct a proof and intend it as general is dependent on the mathematical content of the claim and whether the student has generalized the domain or is able to do so in the time given to develop a proof. When Ellis (2007) explored the relationship between justification and generalization, she found that when students explained their generalizations, they developed increasingly deductive justifications. She recommended proof writing be used as a tool to support students in generalizing more effectively. Proof writing with opportunities for feedback and revisions would allow for students to develop increasingly deductive justifications, allowing for students to attend to the pertinent structure of the domain by performing generalizations.

6.2 Examining the Arguments as Intended

One of the common concerns that has arisen for previous researchers studying student proof construction is whether these assumptions align with the student's intentions for the description or representation of the domain (Ahmadpour et al., 2019; Pedemonte, 2007, 2008; Yopp & Ely, 2016). To evaluate the generality and viability of a student argument a reader makes assumptions based on the context of the proof and the reader's own ability to see the generality in the student argument (Yopp & Ely, 2016). In my quantitative study, to examine the influence of the inclusion of definitions, it was necessary for me to make assumptions as to whether students were intending their descriptions, examples, and algebraic representations of the domain to be general. The interviews completed for this study served two purposes: 1) explore whether students are intending the descriptions, examples, and algebraic representations they develop and use as general or not and 2) see whether the interpretations of algebraic representations that define the states in the model for how students read proof are present when students construct proof.

Through examining how students intended their arguments I found that: 1) the roles of examples and algebraic representation aligned with the states of student understanding while reading proof detailed by Ahmadpour et al. (2019), 2) the structure of the domain students attend to influences their ability to transition to a viable argument, and 3) whether students were generalizing a pattern in the results or a pattern in the process was critical to whether the student developed an argument that described a general procedure or used structure.

6.2.1 The Role of Examples and Algebraic Representations

From the model by Ahmadpour et al. (2019) I developed three classifications to describe the role of examples and algebraic representations within student argument—empirical, placeholder, and class. I hypothesized these classifications based on the states of understanding found by Ahmadpour et al. (2019). Each state of understanding was characterized by how the student was interpreting the algebraic representation. Within the model of how students understand proof while reading was the procedural pathway where representations are used as recipes to check individual cases but not as representative of all objects of that class. This became my working definition of a placeholder representation. In the interviews, students did use their representations in this manner falling into two subcategories: one in which the students simultaneously described the representation as standing for all, and the other where the student did not attend to structure of the domain belonging to all.

I found that students who intended the algebraic representation as a class representation described it in a manner that aligned with the placeholder representation when asked how they knew the representation stood for all objects in the domain. Here they viewed the algebraic representation as standing for “any” object in the domain or x as standing for any natural number to create a representation for any object from the domain. This language of the variable representing “any” is also common within formal proofs in the mathematical community. The other instances were when students used examples to communicate a procedure. Examples were used to explain what Ahmadpour et al. (2019) referred to as general procedures. In these cases, the examples were used to communicate a procedure that they student believed could be applied to any object from the domain to show the truth of the claim. These instances did not include justifications that connected to the structure of the domain. The steps of the procedure were described in a manner that made them specific to each individual example.

6.2.2 Transitioning to Viable Arguments

Research studying the transition from informal argumentation activity to viable argument have found two primary factors at play: 1) the viability of the operators (Pedemonte, 2005) and 2) the type of generalizations (Harel, 2001; Pedemonte, 2005). The interviews with Nessa and Freya demonstrated how the structure of the domain the student is attending to is another factor to consider that is related to these existing factors. While Freya and Nessa both originally described general procedures without any justification, the conceptual insights they were using relied on different structures of the domain and ultimately Freya developed a viable argument while Nessa did not.

To transition to a viable argument Freya needed to only identify the structure of the domain that the first addend is one less than the middle and the third is one greater than the middle and connect that structure to the “leveling” procedure she had described. When asked to explain her general procedure she was able to use the structure to justify why the procedure applies to all objects in the domain to transition to a viable argument. In her interview it becomes apparent that the two factors identified by Balacheff and Pedemonte are present in her argument. Her operator, the "leveling" process, is viable and she is generalizing the process of a pattern. In her explorations of the examples, Freya identified the structure of the domain by generalizing the about the process of the pattern she noticed and the operators available to her are determined by the structure she sees belonging to the domain. The structure Freya identified as belonging to the domain influenced how she was able to transition to a viable argument.

In contrast, the conceptual insight used by Nessa requires attending to the remainders of the individual addends when divided by three. Nessa likely observed that every sum of three consecutive natural numbers has an addend divisible by three and was able to justify that observation however she was unable to justify why the sum of the other two addends was also divisible by three because she was not attending to the remainders of all the addends. The operator Nessa used in her original argument to justify the second part of her argument was dependent on a generalization she had made about the result of a pattern and the operator utilized was not viable. The structure she needed to complete her argument involves examining the remainders of those addends when divided by three and this structure was not available to her.

This is important because some conceptual insights require more operators and operators that may not be familiar to the student. Perhaps if Nessa had returned to the domain and looked for different structures of the domain, she would have been able to identify another structure that allowed for her to construct a viable argument using a different conceptual insight. The structure available to a student dictates the conceptual insights they can form and thus influences their ability to construct viable arguments.

6.2.3 Generalization of a Result Versus Generalization of the Domain

When examining how students intend their arguments, in particular, are they describing or representing the domain in general, I found identifying some forms of empiricism were easier than others. Empiricism is defined as a student using a finite subset of examples to justify the truth of a general claim for a larger set. In classic examples, students say things like "the claim is true because it worked for these three examples." However, there is a form of empiricism that is tricker to identify. This is when students are using observations from their examples to justify the claim. For example, "the claim is true because when you divide the sum by three the answer is the middle number." In the interviews it became apparent that students who made this statement were observing a pattern, finding the existence of a pattern convincing, and justifying the truth of the claim by describing the existence of a pattern that they believe will extend to the whole domain. These students described the domain as sharing a property using the conceptual register without having generalized about the domain. However, it is also very conceivable that a student could make this argument and see the underlying structure and just not see that structure or transformation as a necessary part of their proof. The student is either generalizing about the objects in the domain or they are generalizing the result of a pattern.

The challenge of determining whether the student is generalizing an observed pattern or generalizing about the objects in the domain aligns with Pedemonte's (2007, 2008) findings. The interviews were crucial to making these decisions. When I examined student use of conceptual insight within the written arguments collected for the quantitative study five categories arose: 1) the structure is of the domain 2) the structure is of the domain after treatment, 3) the structure is in the result of treatment 4) the structure belongs to a much larger superset of the domain, and 5) the structure is unclear. In the case where the structure described is not directly describing the structure of the domain, it is unclear if the student has performed a treatment on the domain or is making an empirical observation from their exploration of examples. Here researchers must make an assumption about the student's understanding and intentions, and as proof is dependent on the context of the specific community those assumptions may not be appropriate for data sets that include arguments from different classrooms, universities, states. To address this in the future I suggest attending to the structure that students describe rather than seeking to make assumptions about student's intentions. I found it my interviews that when the structure was of the domain but after undergoing a treatment the student had still generalized about the domain. However, when the structure was in the result of the treatment the student was describing a pattern they observed in their empirical observations. By examining the structure that students describe or represent researchers can consider the conceptual insights that a student may have access too as well as the generalizations about the domain the student has made.

6.3 Constructing Structure

Primary and secondary mathematics educators are encouraged to support their students to "look for and make use of structure" (National Governors Association Center for Best Practices, 2010). My findings suggest that structure is not seen but rather constructed and the mechanism that supports the construction of structure is the generalizations made by students about the set of mathematical objects. This is evident in the student arguments in response to Task C. Several students given the algebraic definition did not adopt the notation to form a technical handle as I had anticipated. Instead, they used the notation to generate examples. Through the process of generating examples in the domain of the claim and confirming the claim for those examples, they found a conceptual insight. Then developed a generic example to communicate the conceptual insight instead of using the algebraic notation.

A possible explanation for how these students used the algebraic notation to develop examples but not to develop their argument is that they did not "see" the structure of the domain within the algebraic notation, $n, n+1, n+2$, making them not adopt the representation as technical handle to stand for all sums of three consecutive natural numbers. The students first saw the structure after generating examples and generalizing about the domain. The students process of developing an argument aligns with constructivist theory of learning wherein a student actively constructs knowledge by through a process of acting and reflecting (Mascolo & Fischer, 2005). The structure of the domain was constructed by the students as they acted on objects from the domain, observed outcomes, and generalized their understanding to the whole domain.

The structure of the domain available to the student determines the conceptual insights they can imagine and thus the possible paths available to develop a viable argument. To successfully develop a viable general direct argument, students, need to identify pertinent structure of the domain and depict or describe how that structure guarantees the conclusion holds for all objects in the domain. Seeing the structure of the domain implies a familiarity with the domain that students won't have unless they have constructed the structure of the domain previously or are in a position where they can construct that structure as part of their exploration of the claim.

6.4 Limitations

The results of this study are subject to some possible limitations. First, the sample size was not sufficient to determine with confidence that the variation in means, for both dependent variables, was not statistically significant. Second, the context of the study includes the mathematical content of the prompt used. While these findings might extend to proving situations outside of context of the specific claim used, this might not be the case. Third, the participants might not be representative of the general student population. These students may share specific mathematical experiences that have shaped the way they respond to proving tasks. An example of this is the conceptual insight “groups of three.” As part of the curriculum for math courses for elementary school teachers, students learn the definitions of multiplication and division. Their instruction may have primed students to be looking for groups of three after reading the conclusion of the claim, “divisible by three.” Furthermore, all students interviewed volunteered. This resulted in interviewing only one student who responded to Task C, and three students who responded to Task B. This limited the opportunities to understand the role of the representations for students who were given definitions.

6.5 Avenues for Future Research

The results of this study provoked questions about the structures of the mathematical objects students perceive when given a definition. Using the framework by Sandefur et al. (2013), a successful proof development entails discovering a technical handle and a conceptual insight that are aligned. The role of the technical handle is to allow access to the structure of the domain that is pertinent to the conceptual insight. From the data it appeared that some students did see structure in the symbolic definition but rejected the notation as a viable technical handle. On the other hand, other students saw the notation as something that could be manipulated but did not see structure to form a conceptual insight. This leads me to wonder what mechanisms allow students to see structure in the definitions they are given, and what mechanisms allow students to view a representation in a definition as a candidate for representing an infinite domain.

In my tasks, I gave students definitions for consecutive and divisibility by three. I wonder how results would be affected if instead, the students were given general representations of sums of three consecutive natural numbers. For instance, the statement "all sums of three consecutive natural numbers can be represented by $n + (n + 1) + (n + 2)$ where n is a natural number." Or a statement with "all sums of three consecutive natural numbers can be represented by…" and then include an informal representation. In my current study participating students who used the definition they were given had to coordinate the definition with the definition of sum to develop a representation of the domain. The results of this study suggest that students did not coordinate the given definition of consecutive into a technical handle they could use to access the structure of the domain. If they were given a general representation of the domain, would more students have used the representation? And if they used it, would they have used it to develop a "formulaic proof," a "procedural proof," or a "formulated proof"?

Additional studies are needed to better understand how definitions and representations influence the construction of general direct arguments. The findings of this study are particular to the context of prospective elementary school teachers and the mathematical content of basic number theory. Zaslavsky and Shir (2005) explored the roles and features that students ascribe to definitions. In my study I found that students did not convert from the conceptual register to the symbolic register to use the algebraic representation of the domain. I have hypothesized how generalization of the domain may be a factor influencing their choice to not adopt the given representation to develop a technical handle. It is also possible that student conceptions of mathematical definitions are also at play as students are making these decisions.

Appendix A: Task A

Task A

Prove that the sum of any three consecutive natural numbers is divisible by three. In your work, show and explain why this is true.

Claim: The sum of any three consecutive natural numbers is divisible by three

Please indicate whether you are willing to discuss your work in a follow up interview.

○ I am willing to discuss my work in a follow up interview.

○ No thank you.

What is your preferred contact information?

____

How do you prefer to be addressed?

____

Thank you for your time and for sharing your thinking!

Appendix B: Task B

Task B

Definitions:

Definition: Two natural numbers are consecutive if one of the numbers is one greater than the other number.

Example: 4,5,6 are consecutive natural numbers because 5 is one greater than 4 and 6 is one greater than 5.

Definition: A number is divisible by three if there is no remainder when the number is divided by 3.

Examples: 6 is divisible by three because 6 divided by 3 is 2 with no remainder.

Prove that the sum of any three consecutive natural numbers is divisible by three. In your work, show and explain why this is true.

Claim: The sum of any three consecutive natural numbers is divisible by three

Please indicate whether you are willing to discuss your work in a follow up interview.

O I am willing to discuss my work in a follow up interview. O No thank you.

What is your preferred contact information?

____

How do you prefer to be addressed?

____

Thank you for your time and for sharing your thinking!

Appendix C: Task C

Task C

Definitions:

<u>Definition:</u>	Three consecutive natural numbers are numbers that can be represented as n, n+1, n+2 for some natural number n.
Example:	4, 5, 6 are three consecutive integers because they can be written as 4, 4+1, 4+2
<u>Definition:</u>	A number m is divisible by 3 if it can be written as $3 * k$ for some integer k
Example:	6 is divisible by three because $6 = 3 * 2$. This means $6 = 3 * k$ when $k = 2$.

Prove that the sum of any three consecutive natural numbers is divisible by three. In your work, show and explain why this is true.

Claim: The sum of any three consecutive natural numbers is divisible by three

Please indicate whether you are willing to discuss your work in a follow up interview.

O I am willing to discuss my work in a follow up interview. O No thank you.

What is your preferred contact information?

____

How do you prefer to be addressed?

____

Thank you for your time and for sharing your thinking!

Appendix D: Interview Protocol

Interview Protocol

Goal:

The goal of this interview is to explore how participants intend their descriptions and representations of the domain of the claim.

- According to the participant, for what do the representations stand?
- According to the participant, does the representation stand for a single object, a placeholder for any object to be placed, a representation of a set as an entity, or something else?
- Is the set the student envisioned/imagined equivalent to the domain?

Administration Guidelines:

Participants will be selected for the interview using stratified sampling to explore the meaning of a diverse selection of representations developed by participants given different tasks.

Part 0: For virtual interviews

- Orally present the consent form and ask participants if they consent.

Part 1: Understand the argument made

- Tell me about your thinking as you completed this task. (Clarify justification for each step.)
- Can you tell me more about your thinking when you wrote/drew this [point to their representation/description of the domain]?
- Does your work show that ***all*** sums of three consecutive numbers are divisible by three? How do you know?
    - If we thought of a super large sum of three consecutive numbers, 1,346+1,347+1,348, do we know from your work whether that sum will be divisible by three?

Part 2: Understand what the description/representation stands for

Choose A if the representation belongs to the conceptual register and B if the representation belongs to the symbolic register or is a generic example..

A) Representations are in the **conceptual register.**

- You stated "consecutive numbers can always ...." Tell me more about why that is? How do you know?
- Is it possible for this [describe the transformation or procedure they detail] to be applied to some other sum, a sum that is not the sum of three consecutive natural numbers, and still result in [describe the outcome they found in their argument e.g., three copies of the middle number/three copies of the first number plus 3]?

B) Representations are in the **symbolic notation or are generic examples.**

- What does this [point to ....] stand for or represent? How do you know?
- For which sums of three consecutive numbers does this represent? How do you know?
- Is it possible for there to be something that is described by this representation that is not a sum of three consecutive numbers? How do you know?

Appendix E: Initial Codes and Themes from Inductive Thematic Analysis

Themes and Codes	Participants	Number of Excerpts	Example Excerpt
Examples as evidence	8	43	
example	Marge, Lucy, Kendra, Nessa	7	So I just did 7, 8, 9 because those are three consecutive numbers and seven, eight and nine add up to an answer of 24 and then I divided 24 by three to get eight, so it divided evenly by three with no remainders, no like fractions or decimals or anything, just 8
selected examples	Audrey, Jess	3	I chose 3, 4 and 5 as one example, and then I chose 10, 11 and 12 as a different example just to show that single digit versus multi digit numbers would both work.
random example	Jess	1	I picked three random numbers
examples show claim is true	Jess	1	I guess like technically not my work shows that, but Umm, I did like a few other sets of three because I was like I kind of noticed that like oh this is kind of cool. Like I can't like this one was divisible by three. And so I kind of did a few like in my head afterwards and I've been like, I don't know, I thought about it for a while afterwards. And realize that that's, from the ones that I've tested, it's pretty accurate
more examples needed	Marge, Kendra	2	I think that if I wanted to show that it worked for all, I would probably do bigger numbers too, numbers that had like the ones and the tens place, ones, tens hundreds place, ones in the thousands. But there's a lot of numbers out there
more examples to confirm	Lucy	1	I wasn't sure if it was just going to be true with one set of numbers, so I tried to do a few to make sure.

example to demonstrate procedure	Nessa	1	So I just find three like random three consecutive members. And use the way that I saw like first I found the number that is, cause three consecutive is for sure we'll have at least one, that like we'll be able to divide it by three directly. Feel like I see a bigger number like 102, 103 and 104. So for this one 102, 102 is going to be able to directly divide it by three and so. So for 101 no, 101 + 103 equals 204 which is also going to be able to divide it by three
example to demonstrate algebra	Cleo	1	I mean they put seven in the example. Obviously that works with literally any number
Algebraic Representation	**3**	**13**	
algebra mode	Cleo, Winston	2	because it specifically gave it the it as n + 1 and +2 up there. I was already kind of in like algebra mode
algebra as a procedure	Winston	1	it's just a practice in algebra, you know. What is this way to simplify what they're saying, you know?
algebra representing any object in the domain	Winston	1	[n+(n+1)+(n+2) is] any three consecutive numbers starting with n
algebra representing only objects in the domain	Cleo, Winston	2	[n+(n+1)+(n+2) cannot be anything other than 3 consecutive numbers] because no matter what, you substitute the, the problem with that is that it there's no other way to interpret it specifically because of the parenthesis. It might have been different if I hadn't parenthesized it, but the fact that each parenthesis is there indicates that. It itself is that n + 1 is an entire unit number
description of algebra	Cleo, Nessa Winston	4	if you're always using X and then as the same number, then you have X, X + 1 is going to be the next number, and x + 2 will be the number following that one

variable represents any	Cleo	2	Hopefully [my work shows that all sums of three consecutive natural numbers are divisible by three], uh, as I did use a variable which you can plug any number in to and still get the same result that it's divisible by three
leveraging structure from algebra	Cleo	1	Obviously, it was just like, well, if you have three numbers that the difference between the top number and the middle number and the bottom number altogether is 3, obviously you're gonna be able to divide that number by three
Patterns as a tool for generalization	**5**	**25**	
describing pattern leads to identifying general structure	Audrey, Jess	3	I think it would because the same pattern applies, that the numbers are all consecutive and so when you divide by three you're finding the average always ends up being that middle number for median
justifying pattern using average and median	Audrey, Freya	2	I guess, um, well, for both of my equations that was the case, but I think it's because it's the average of the three numbers. So when you take the average of the sum, you're always going to get the mean or the middle number, the median I guess. So I think that's why the pattern is consistent the whole way through
generalizing to all after identifying pattern for few	Audrey	1	I think that it's a pattern within numbers that's consistent throughout the numbers. So that's kind of where I made that educated assumption, I guess is what I would call it
pattern or general procedure	Nessa	1	So the work that I give here is first I find the like the number who can directly be like divided by three like 3, 9, 21 here. So like 3 divided by 3, 9 divided by 3, 21 divided by 3 and 1 + 2 = 3 and here is 7 + 8 = 15 and 22 + 23 = 45 which is also to be able to divide it by three so. Yeah, that's just how I thought.

discovered pattern	Audrey, Jess	2	Realized that, like the second number will always be the answer after it's divided by three, so I just kind of made a little note of that, but yeah, I don't know
looking for patterns	Freya	1	I like to kind of think about things in terms of patterns
result is a pattern	Lucy	1	it doesn't show that all numbers are, no, but it does prove that it is a pattern that can be seen at least with numbers extending up to 12
Shifting representation or procedure	**3**	**7**	
generic example	Freya	1	They're obviously all going to be like 9, 10 and 11 or one after the other, and so you can always take the top one which is 2 away from the bottom one. And take that extra one and put it over to 9 so they're all equal
new general objectification of domain	Audrey	1	I guess, um, well, for both of my equations that was the case, but I think it's because it's the average of the three numbers. So when you take the average of the sum, you're always going to get the mean or the middle number, the median I guess. So I think that's why the pattern is consistent the whole way through
new general procedure	Freya	1	I realized that when you would add three consecutive numbers, you could take the number, the larger numbers, like in 5, 6 or 7, you can take seven and you can move one of the numbers from 7 to the 5 to make them all equal. And then that would kind of, which is why we need to divide by three you're gonna always get the middle number.

new representation not connected to original structure	Nessa	1	Like I use like n is the first number and the second is n + 1 and then it's n + 2. So those three that are like consecutive and if I just add them together and plus n + 1 + n + 2. So it's 3n+3. And as we know that like any number which can like after multiply, not multiply, like adding up if they are like able to divide it by three, then that number is going to be able to divide it by three. So like 3n+3 no matter what n is, it will be able to divide it by three
conflict between new and old representation	Nessa	1	Wait hold up. So for 21, 22, 23 is actually different
pattern restricted to domain	Audrey	1	I think it's only with the sum when the three numbers are consecutive and it's divisible by three
Procedure can be applied to a bigger domain	Freya	1	I do think it's possible [to do a leveling thing with other sums that aren't consecutive natural numbers], but they would have to be like separated the same amount. So if you did like it would have to be like 1, 3 and 5 I think
What it takes to prove	**3**	**3**	
proof needed to show for all	Marge	1	I feel like you would have to probably come up with some sort of proof whether that be… And I don't know if maybe that would be like the last number in the equation like knowing that the numbers in the ones unit add up together and divide by three. But I don't know what it would be. I feel like you have to come up with some sort of proof to prove that any three numbers can do it
proof needed to show for all	Lucy	1	My mind immediately went to like proofs and stuff, so I just did like a few examples.
showing all is not possible	Audrey	1	So, the hard thing with that is it would be impossible to show every single example of that because there's infinitely natural numbers

Appendix F: Initial Codes and Themes from Theoretical Thematic Analysis

Initial Code	Participant	Task Version	Number of Excerpts	Excerpt
class representation	Winston, Cleo, Nessa	A,C	5	you have X, X + 1 is going to be the next number, and x + 2 will be the number following that one
Type 1 placeholder	Winston	C	1	That's any three consecutive numbers starting with n [referring to n+(n+1)+(n+2)]
Type 2 placeholder representation	Cleo	A,C	3	I did use a variable which you can plug number in two and still get the same result that it's divisible by three.
Empirical Representation	Kendra, Marge, Lucy, Jess, Audrey	A,B	5	they gave the example 4, 5, 6. So that's actually what I use for my first time up there. I added 4 + 5 + 6, that's what they provided. So then they added that together and I got 15 which is basic math and then it says that if it's divisible by three, there's no remainder, so I divided that by three, which I got five, so I put true
Empirical Representation (claim is accurate/correct)	Kendra	B	2	I just wanted to check with other numbers so I just did like kind of the basic numbers. So like 1, 2 and 3 is easy. So that was six. 6 divided 3 was two and then I knew that was true. I just wanted to verify again for the third time because three lucky numbers was like OK, let's do it 7, 8, 9 and I got 24. 24 divided by 3 was eight, simple math, which was true. So I I felt like it was pretty accurate if I could verify it three times

Empirical Representation (doesn't show all)	Marge, Lucy	A,B	2	I just did 7, 8 and, 9. That doesn't mean that 64, 65 and 66 will also add together and divide by three
general procdure with justification	Freya	A	1	I think it'll always work with three consecutive numbers because. They're obviously all going to be like 9, 10 and 11 or one after the other, and so you can always take the top one which is 2 away from the bottom one. And take that extra one and put it over to 9 so they're all equal.
general procedure	Nessa	A	1	I used the example like 1, 2, 3 or 7, 8, 9 or 21, 22, 23 or whatever numbers they're like consecutive. And so like the first two once they add up together like 21 + 22 equals 43, yes, and so 43 and then is 23, 43 + 23 = 69 so 69 you know is able to be divided by three. R: OK, so those are examples that you P: Wait hold up. So for 21, 22, 23 is actually different. So the work that I give here is first I find the like the number who can directly be like divided by three like 3, 9, 21 here. So like 3 divided by 3, 9 divided by 3, 21 divided by 3 and 1 + 2 = 3 and here is 7 + 8 = 15 and 22 + 23 = 45 which is also to be able to divide it by three so. Yeah, that's just how I thought.
general procedure/pattern	Freya	A	1	I realized that when you would add three consecutive numbers, you could take the number, the larger numbers, like in 5, 6 or 7, you can take seven and you can move one of the numbers from 7 to the 5 to make them all equal. And then that

				would kind of, which is why we need to divide by three you're gonna always get the middle number.
pattern	Audrey	A	1	So I discovered when doing both of these equations that my answer in the division. So I I added to find the sum of the three and then when I divided it by three it ended up being the, um the second number in the sequence, it was the same as the quotient from the division problem.
pattern	Jess	B	1	Realized that, like the second number will always be the answer after it's divided by three, so I just kind of made a little note of that, but yeah, I don't know

Appendix G: Calculations of Cohen's kappa for Interrater Reliability Study

The calculation of Cohen's kappa, κ, for generality of representation code where P_A is the proportion of arguments on which the scorers agree on the score and P_C is the proportion of arguments for which agreement is expected by chance.

$$P_A = \frac{total\ number\ of\ responses\ where\ scorers\ agree}{total\ number\ of\ responses} = \frac{28}{30} = 0.90$$

$$P_C = \frac{total\ number\ of\ 1\ scores\ by\ 1st\ scorer}{total\ number\ of\ responses} * \frac{total\ number\ of\ 1\ scores\ by\ 2nd\ scorer}{total\ number\ of\ responses} + \frac{total\ number\ of\ 0\ scores\ by\ 1st\ scorer}{total\ number\ of\ responses} * \frac{total\ number\ of\ 0\ scores\ by\ 2nd\ scorer}{total\ number\ of\ responses}$$

$$= \frac{17}{30} * \frac{16}{30} + \frac{13}{30} * \frac{14}{30} \approx 0.50\overline{4}$$

$$\kappa = \frac{P_A - P_C}{1 - P_C} \approx \frac{0.9 - 0.504}{1 - 0.504} \approx 0.798$$

The calculation of Cohen's kappa, κ, for viability of the conceptual insight code where P_A is the proportion of arguments on which the scorers agree on the score and P_C is the proportion of arguments for which agreement is expected by chance.

$$P_A = \frac{27}{30} = 0.9$$

$$P_C = \frac{total\ number\ of\ 1\ scores\ by\ 1st\ scorer}{total\ number\ of\ responses} * \frac{total\ number\ of\ 1\ scores\ by\ 2nd\ scorer}{total\ number\ of\ responses} + \frac{total\ number\ of\ 0\ scores\ by\ 1st\ scorer}{total\ number\ of\ responses} * \frac{total\ number\ of\ 0\ scores\ by\ 2nd\ scorer}{total\ number\ of\ responses}$$

$$= \frac{4}{30} * \frac{7}{30} + \frac{26}{30} * \frac{23}{30} = \frac{626}{900} \approx 0.69\overline{5}$$

$$\kappa = \frac{P_A - P_C}{1 - P_C} \approx \frac{0.9 - 0.696}{1 - 0.696} \approx 0.671$$

Literature Cited

Ahmadpour, F., Reid, D., & Reza Fadaee, M. (2019). Students' ways of understanding a proof. *Mathematical Thinking and Learning*, *21*(2), 85-104. https://doi.org/10.1080/10986065.2019.1570833

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

Arcavi, A. (2003). The role of visual representations in the learning of mathematics. *Educational Studies in Mathematics, 52*, 215-241. https://doi.org/10.1023/A:1024312321077

Arzarello (Eds.), *Proceedings of the 6th congress of the European Society for Research in Mathematics Education* (pp. 2336–2345). Lyon, France: INRP.

Azrou, N., & Khelladi, A. (2019). Why do students write poor proof texts? A case study on undergraduates' proof writing. *Educational Studies in Mathematics*, *102(2),* 257-274. https://doi.org/10.1007/s10649-019-09911-9

Banerjee, R., & Subramaniam, K. (2012). Evolution of a teaching approach for beginning algebra. *Educational Studies in Mathematics, 80(3),* 351-367. https://doi.org/10.1007/s10649-011-9353-y

Balacheff, N. (1988a). Aspects of proof in pupils' practice of school mathematics. *In D. Pimm (Ed.), Mathematics, teachers and children* (pp. 216–235). London: Hodder and Stoughton.

Balacheff, N. (1988b). A study of students' proving processes at the junior high school level. *Second UCSMP International Conference on Mathematics Education*. https://hal.archives-ouvertes.fr/hal-01652045v2

Balacheff, N. (2013). cK¢, A Model to Reason on Learners' Conceptions. In Martinez, M. & Castro Superfine, A (Eds.). *Proceedings of the 35th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education.* Chicago, IL: University of Illinois at Chicago.

Ball, D. L., & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. *Multiple Perspectives on the Teaching and Learning of Mathematics*, *4*(1), 83-104. Westport, CT: Ablex.

Ball, D. L., & Bass, H. (2003). Making mathematics reasonable in school. In J. Kilpatrick, W.G. Martin, & D. Schifter (Eds.), *A research companion* to *Principles and Standards for School Mathematics* (pp. 27-44). Reston, VA:National Council of Teachers of Mathematics.

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, China: Higher Education Press.

Battista, M. T. (1999). Fifth graders' enumeration of cubes in 3D arrays: Conceptual progress in an inquiry-based classroom. *Journal for Research in Mathematics Education, 30(4),* 417–448. https://doi.org/10.2307/749708

Bell, A. (1993). Principles for the Design of Teaching. *Educational Studies in Mathematics, 24*(1), 5-34. https://doi.org/10.1007/BF01273293

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

Booth, L. R. (1988). Children's difficulties in beginning algebra. In A. F. Coxford (Ed.), *The ideas of algebra, K-12*, (pp. 20-32). Reston, VA: National Council of Teachers of Mathematics.

Booth, R. L. & Thomas, M. J. (2000). Visualization in Mathematics Learning: Arithmetic problem-solving and student difficulties. *Journal of Mathematical Behavior, 8(2)*, 169-190. https://doi.org/10.1016/S0732-3123(99)00027-9

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Brown, S. (2008). Exploring epistemological obstacles to the development of mathematics induction. In M. Zandieh (Ed.), *Proceedings of the 11th Conference for Research on Undergraduate Mathematics Education,* San Diego, CA. Retrieved from http://sigmaa.maa.org/rume/crume2008/Proceedings /S_Brown_LONG.pdf

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

Bruner, J. S. (1964). The course of cognitive growth. *American Psychologist*, *19(1),* 1–15. https://doi.org/10.1037/h0044160

Buchbinder, 0., & Zaslavsky, O. (2007). How to decide? Students' ways of determining the validity of mathematical statements. In D. Pita-Fantasy & G. Philippot (Eds.), *Proceedings of the Fifth Congress of the European Society for Research in Mathematics Education* (pp. 561-571), Larnaca, Cyprus: University of Cyprus. http://orlybuchbinder.com/publications/Buchbinder_Zaslavsky_CERME5.pdf

Bujang, M. A., & Baharum, N. (2017). Guidelines of the minimum sample size requirements for Kappa agreement test. *Epidemiology, Biostatistics and Public Health*, *14*(2). http://doi.org/ 10.2427/12267

Carraher, D., Schliemann, A., Brizuela, B., & Earnest, D. (2006). Arithmetic and Algebra in Early Mathematics Education. *Journal for Research in Mathematics Education, 37*(2), 87-115. https://doi.org/10.2307/30034843

Chazan, D., & Lueke, M. H., (2009). Exploring Relationships Between Disciplinary Knowledge and School Mathematics. In Stylianou, D. A., Blanton, M. L., Knuth, E. J., (Eds.) *Teaching and learning proof across the grades* (pp. 21- 39). New York, NY: Routledge

Coe, R., & Ruthven, K. (1994). Proof practices and constructs of advanced mathematics students. *British educational research journal*, *20*(1), 41-53. https://doi.org/10.1080/0141192940200105

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, *20*(1), 37-46. https://doi.org/10.1177/001316446002000104

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Thousand Oaks, CA: Sage Publications.

Czocher, J. A., & Weber, K. (2020). Proof as a cluster category. *Journal for Research in Mathematics Education*, *51*(1), 50-74. https://doi.org/10.5951/jresematheduc.2019.0007

Dawkins, P. C., & Karunakaran, S. S. (2016). Why research on proof-oriented mathematical behavior should attend to the role of particular mathematical content. *The Journal of Mathematical Behavior*, 44, 65-75. https://doi.org/10.1016/j.jmathb.2016.10.003

Ding, M., & Auxter, A. E. (2017). Children’s strategies to solving additive inverse problems: a preliminary analysis. *Mathematics Education Research Journal*, *29*(1), 73-92. https://doi.org/10.1007/s13394-017-0188-4

Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational Studies in Mathematics*, *61*(1-2), 103-131. https://doi.org/10.1007/s10649-006-0400-z

English, L. D. (2004). Promoting the development of young children's mathematical and analogical reasoning. In L. D. English (Ed.), *Mathematical and analogical reasoning of young learners.* (pp. 201-213). Mahwah, NJ: Lawrence Erlbaum.

Esty, W. W., & Esty, N. C. (2009). *Proof: Introduction to Higher Mathematics* (5th ed.). Self-published.

Forman, E. A., Larreamendy-Joerns, J., Stein, M. K., & Brown, C. (1998). “You’re going to want to find out which and prove it”: Collective argumentation in mathematics classrooms. *Learning and Instruction*, *8(6)*, 527–548. https://doi.org/10.1016/S0959-4752(98)00033-4

Gattegno, C. (1963). *For the teaching of mathematics*. *Vol. 3, Elementary mathematics.* Educational Explorers Limited.

Goetting, M. M. (1995). *The college student's understanding of mathematical proof* (Order No. 9539653). Available from ProQuest Dissertations & Theses Global. (304216478). Retrieved from https://www.proquest.com/dissertations-theses/college-students-understanding-mathematical-proof/docview/304216478/se-2

Gokkurt, B., Soylu, Y., & Sahin, O. (2014, December). Analysis of the mathematical proof skills of students of science teaching. *Educational Research Quarterly*, *38*(2), 3-22.

Goldin G. A. (2003). Representation in school mathematics: A unifying research perspectives. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A Research Companion to Principles and Standards for School Mathematics* (pp. 275-285). Reston VA: National Council of Teachers of Mathematics.

Goldin, G. A., & Shteingold, N. (2001). Systems of representation and the development of mathematical concepts. In Cuoco, A. A. & Curcio, F. R. (Eds.), *The role of representation in school mathematics* (pp. 1- 23). Reston VA: National Council of Teachers of Mathematics.

González, G., & Herbst, P. (2013). An Oral Proof in a Geometry Class: How Linguistic Tools Can Help Map the Content of a Proof. *Cognition and Instruction, 31(3)*, 271–313. https://doi.org/10.1080/07370008.2013.799166

Goulding, M., Rowland, T., & Barber, P. (2002). Does it matter? Primary teacher trainees' subject knowledge in mathematics. *British Educational Research Journal,* 28,689-704. https://doi.org/10.1080/0141192022000015543a

Goulding, M., & Suggate, J. (2001). Opening a can of worms: Investigating primary teachers' subject knowledge in mathematics. *Mathematics Education Review,* 13, 41-44.

Gray, E., Pitta, D., & Tall, D. O. (2000). Objects, actions, and images: A perspective on early number development. *Journal of Mathematical Behavior, 18*, 401-413. https://doi.org/10.1016/S0732-3123(00)00025-0

Haciomeroglu, E. S., Aspinwall, L., & Presmeg, N. C. (2010). Contrasting cases of calculus students' understanding of derivative graphs. *Mathematical Thinking and Learning, 12*, 152-176. https://doi.org/10.1080/10986060903480300

Hanna, G. (1990). Some pedagogical aspects of proof. *Interchange, 21(1),*6-13.

Hanna, G., & Jahnke, H. N. (1996). Proof and proving. In A. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 877-908). Dordrecht, The Netherlands: Kluwer.

Harel, G. (2001). The development of mathematical induction as a proof scheme: A model for DNR-based instruction. In S. Campbell & R. Zaskis (Eds.), *Learning and teaching number theory: Research in cognition and instruction* (pp.185-212). Dordrecht, The Netherlands: Kluwer.

Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. *American Mathematical Society*, *7*, 234-283.

Hayes, J. R., & Hatch, J. A. (1999). Issues in measuring reliability: Correlation versus percentage of agreement. *Written Communication*, *16*(3), 354-367.

Healy, 1., & Hoyles, C. (2000). Proof conceptions in algebra. *Journal for Research in Mathematics Education, 3*1(4), 396-428. https://doi.org/10.2307/749651

Hoch, M., & Dreyfus, T. (2004). Structure sense in high school algebra: The effect of brackets. In M. J. Hoincs & A. B. Fuglcstad (Eds.), *Proceedings of the 28th conference of the international group for psychology of mathematics education (Vol 3)* (pp. 49-56). Bergen: Bergen University C

Hodds, M., Alcock, 1., & Inglis, M. (2014). Self-explanation training improves proof comprehension. *Journal for Research in Mathematics Education, 45(1),* 62-101. https://doi.org/10.5951/jresematheduc.45.1.0062

Hollebrands, K., Conner, A., & Smith, R. C. (2010). The nature of arguments provided by college geometry students with access to technology while solving problems. *Journal for Research in Mathematics Education*, *41(4)*, 324–350. https://doi.org/10.5951/jresematheduc.41.4.0324

Hunting, R. (2003). Part-whole number knowledge in preschool children. *Journal of Mathematical Behavior*, *22*, 217-235. https://doi.org/10.1016/S0732-3123(03)00021-X

Iannone, P., &Inglis, M. (2010). Self efficacy and mathematical proof: Are undergraduates good at assessing their own proof production ability? In *Proceedings of the 13th Conference for Research in Undergraduate Mathematics Education.* Raleigh, North Carolina. Retrieved from http://sigmaa.maa.org/rume /crume2010/Archive/Iannone%20&%20Inglis.pdf

Inglis, M., & Alcock, L. (2012). Expert and novice approaches to reading mathematical proofs. *Journal for Research in Mathematics Education*, *43(4),* 358-390. https://doi.org/10.5951/jresematheduc.43.4.0358

Inglis, M., Mejia-Ramos, J. P., & Simpson, A. (2007). Modelling mathematical argumentation: The importance of qualification. *Educational Studies in Mathematics*, *66*, 3–21. https://doi.org/10.1007/s10649-006-9059-8

Jahnke, R.N., &Wambach, R. (2013). Understanding what a proof is: A classroom-based approach. *ZDM- The Inter- national Journal on Mathematics Education, 45*, 469-482. https://doi.org/10.1007/s11858-013-0502-x

Janvier, C. (1987). Translation process in mathematics education. In C. Janvier (Eds.), *Problems of representation in the teaching and learning of mathematics* (pp. 27-32). Hillsdale, NJ: Lawrence Erlbaum.

Kieran, C. (1988). Two different approaches among algebra learners. In A. F. Coxford (Eds.), The ideas of algebra, K-12. 1988 Yearbook of the National Council of Teachers of Mathematics (pp. 91-96). Reston, VA: National Council of Teachers of Mathematics.

Kieran, C. (1992). The learning and teaching of school algebra. In D. Grouws (Eds.), *Handbook of research on mathematics teaching and learning* (pp. 390-419). New York: MacMillan

Kieran, C. (2018) Seeking, using, and expressing structure in numbers and numerical operations: A fundamental path to developing early algebraic thinking. In Kieran, C. (Ed.) *Teaching and Learning Algebraic Thinking with 5- to 12-Year-Olds: The Global Evolution of an Emerging Field of Research and Practice*, (pp. 79–105). Dordrecht: Springer.

Kitcher, P. (1984). *The nature of mathematical knowledge.* New York, NY: Oxford University Press.

Knuth, E. J. (2002). Teachers' conceptions of proof in the context of secondary school mathematics. *Journal of mathematics teacher education*, *5(1),* 61-88. https://doi.org/10.1023/A:1013838713648

Knuth, E. J., Choppin, J., & Bieda, K. (2009). Middle school students' productions of mathematical justification. In M. Blanton, D. Stylianou, & E. Knuth (Eds.), *Teaching and learning proof across the grades: A K*-16 *perspective* (pp.153-170). New York, NY: Routledge.

Knuth, E. J., Choppin, J., Slaughter, M., & Sutherland, J. (2002). Mapping the conceptual terrain of middle school students' competencies in justifying and proving. In D. S. Mewborn, P. Sztajn, D. Y. White, H. G. Weigel, R. 1. Bryant, & K. Nooney (Eds.), *Proceedings of the* 24*th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol.4, pp.1693-1670). Athens, GA: Clearinghouse for Science, Mathematics, and Environmental Education.

Ko, Y. Y., & Knuth, E. J. (2013). Validating proofs and counterexamples across content domains: Practices of importance for mathematics majors. *The Journal of Mathematical Behavior*, *32*(1), 20-35. https://doi.org/10.1016/j.jmathb.2012.09.003

Krummheuer, G. (1995). The ethnography of argumentation. In P. Cobb, & H. Bauersfeld (Eds.), *The emergence of mathematical meaning: Interaction in classroom cultures* (pp. 229–269). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kuehl, R. O. (2000). *Design of experiments: statistical principles of research design and analysis* (2nd ed.). Brooks/Cole.

Lamon, S. (1996). The development of unitizing: Its role in children's partitioning strategies. *Journal for Research in Mathematics Education*, 27, 170-193. https://doi.org/10.5951/jresematheduc.27.2.0170

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174. https://doi.org/10.2307/2529310

Lavy, I. (2006). A Case Study of Different Types of Arguments Emerging from Explorations in an Interactive Computerized Environment. *The Journal of Mathematical Behavior*, *25*, 153–169. https://doi.org/10.1016/j.jmathb.2006.02.006

Leron, U., & Zaslavsky, O. (2013). Generic proving: Reflections on scope and method. *For the Learning of Mathematics*, *33*(3), 24-30. http://www.jstor.org/stable/43894858

Linchevski, L., & Livneh, D. (1999). Structure Sense: The Relationship between Algebraic and Numerical Contexts. *Educational Studies in Mathematics, 40*(2), 173-196. Retrieved from http://www.jstor.org/stable/3483083

Mainali, B. (2021). Representation in Teaching and Learning Mathematics. *International Journal of Education in Mathematics, Science and Technology*, *9(1),* 1-21.

Malle, G. (1993). Didaktische Probleme der elementaren Algebra. [Didactic problems in elementary algebra.] .J. Braunschweig: Vieweg

Mamona-Downs, J., & Downs, M. (2009). Necessary realignments from mental argumentation to proof presentation. In *Proceedings of the Sixth Congress of the European Society for Research in Mathematics Education* (pp. 2336-2345). Lyon, France: Institut National de Recherche Pédagogique.

Martin, J. (2013). Differences between experts' and students' conceptual images of the mathematical structure of Taylor series convergence. *Educational Studies in Mathematics, 82*(2), 267-283. http://www.jstor.org/stable/23434872

Martinez, M. V., & Pedemonte, B. (2014). Relationship between inductive arithmetic argumentation and deductive algebraic proof. *Educational Studies in Mathematics*, *86*(1), 125–149. http://www.jstor.org/stable/43590064

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

Mariotti, M.A. (2000b). Justifying and proving in the Cabri environment. *International Journal of Computers for Mathematical Learning,* 6, 257-28l. https://doi.org/10.1023/A:1013357611987

Mariotti, M. A. (2006). Proof and proving in mathematics education. In A. Gutierrez & P. Boero (Eds.), *Handbook of research on the PME: Past, present and future* (pp. 173-204). Rotterdam, The Netherlands: Sense.

Mark Hodds, Lara Alcock, & Matthew Inglis. (2014). Self-Explanation Training Improves Proof Comprehension. *Journal for Research in Mathematics Education*, *45(1),* 62–101. https://doi.org/10.5951/jresematheduc.45.1.0062

Mascolo, M. F., & Fischer, K. W. (2005). Constructivist theories. Cambridge Encyclopedia of Child Development (pp. 49-63).

Mason, J., Stephens, M., & Watson, A. (2009). Appreciating mathematical structure for all. *Mathematics Education Research Journal, 21*(2), 10-32. doi:http://ida.lib.uidaho.edu:2153/10.1007/BF03217543

Mason, J., & Pimm, D. (1984). Generic examples: Seeing the general in the particular. *Educational Studies in Mathematics*, *15*(3), 277-289. https://doi.org/10.1007/BF00312078

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, *22(3),* 276-282. https://hrcak.srce.hr/89395

Moore-Russo, D., Conner, A., & Rugg, K. I. (2011). Can slope be negative in 3-space? Studying concept image of slope through collective definition construction. *Educational Studies in Mathematics*, 76, 3–21. https://doi.org/10.1007/s10649-010-9277-y

Morgan, C. (2014). Understanding practices in mathematics education: Structure and text. *Educational Studies in Mathematics, 87*(2), 129-143. Retrieved from http://www.jstor.org/stable/43589871

Morris, A. K. (1999). Developing Concepts of Mathematical Structure: Pre-Arithmetic Reasoning versus Extended Arithmetic Reasoning. *Focus on learning problems in mathematics*, *21*(1), 44-72.

Morris, A. K. (2002). Mathematical reasoning: Adults' ability to make the inductive-deductive distinctions. *Cognition and Instruction, 20(1),* 79-118. https://doi.org/10.1207/S1532690XCI2001_4

Morris, A. K. (2007). Factors affecting preservice teachers' evaluations of the validity of students' mathematical arguments in classroom contexts. *Cognition and Instruction, 25(4),* 479-522. https://doi.org/10.1080/07370000701632405

Morris, A. K. (2009). Representations that Enable Children to Engage in Deductive Argument. In Stylianou, D. A., Blanton, M. L., Knuth, E. J., (Eds.) *Teaching and Learning Proof Across the Grades* (pp. 87-101). National Council of Teachers of Mathematics.

Mulligan (2002). The role of structure in children's development of multiplicative reasoning. In B. Barton, K. C. Irwin, M. Pfannkuch, & M. O. Thomas (Eds.), *Mathematics Education in the South Pacific Proceedings of the 25th annual conference of the Mathematics Education Research Group of Australasia*, Auckland, NZ, Vol. 2, (pp. 497-503). Sydney: MERGA.

Mulligan, J., & Mitchelmore, M. (1997). Young Children's Intuitive Models of Multiplication and Division. *Journal for Research in Mathematics Education, 28*(3), 309-330. doi:10.2307/749783

Mulligan, J., & Mitchelmore, M. (2009). Awareness of pattern and structure in early mathematical development. *Mathematics Education Research Journal, 21*(2), 33-49. doi:http://ida.lib.uidaho.edu:2153/10.1007/BF03217544

Mulligan, J., Watson, J. (1998). A developmental multimodal model for multiplication and division. *Mathematics Education Research Journal, 10*(2), 61-86. doi:http://ida.lib.uidaho.edu:2153/10.1007/BF03217343

National Governors Association Center for Best Practices, Council of Chief State School Officers (2010). *Common Core State Standards Mathematics*. Washington, DC: National Governors Association Center for Best Practices, Council of Chief State School Officers,

Nardi, E., Biza, I., & Zachariades, T. (2012). Warrant" revisited: Integrating mathematics teachers' pedagogical and epistemological considerations into Toulmin's model for argumentation. *Educational Studies in Mathematics*, *79*, 157–173. https://doi.org/10.1007/s10649-011-9345-y

National Council of Teachers of Mathematics [NCTM]. Principles and standards for school mathematics (2000). In *National Council of Teachers of Mathematics*. Retrieved September, 2022.

Ott, R. L., & Longnecker, M. T. (2015). *An introduction to statistical methods and data analysis*. Cengage Learning.

Pedemonte, B. (2007). How can the relationship between argumentation and proof be analysed? *Educational Studies in Mathematics, 66*, 23-4l. https://doi.org/10.1007/s10649-006-9057-x

Pedemonte, B. (2008). Argumentation and algebraic proof. *ZDM*, *40(3),* 385-400. https://doi.org/10.1007/s11858-008-0085-0

Pedemonte, B., & Balacheff, N. (2016). Establishing links between conceptions, argumentation and proof through the ck¢-enriched Toulmin model. *The Journal of Mathematical Behavior*, *41*, 104-122. https://doi.org/10.1016/j.jmathb.2015.10.008

Pitta-Pantazi, D., Gray, E., & Christou, C. (2004). Elementary school students' mental representations of fractions. In Haines, M.J & Fuglestad, A. (Eds.), *Proceedings of the 28th annual conference of the International Group for the Psychology of Mathematics*

Presmeg, N. C. (2006). Research on visualization in learning and teaching mathematics. In A. Gutiérrez & P. Boero (Eds.), Handbook of research on the psychology of mathematics education (pp. 205–235). Rotterdam: Sense Publishers.

Polya, G. (1981). *Mathematical Discovery on Understanding, Learning and Teaching Problem Solving, Volumes I and II*. Wiley.

Ron, G., & Dreyfus, T. (2004). The use of models in teaching proof by mathematical induction. *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 113-120). Bergen, Norway: Bergen University College.

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

Raman, M., Sandefur, J., Birky, G., Campbell, C., & Somers, K. (2009). Is that a proof? Using video to teach and learn how to prove at the university level. In *Proceedings of the ICMI study 19 conference: Proof and proving in mathematics education* (Vol. 2, pp. 154-159).

Reed, S. D. (2021). *Connecting Logic and Proof Techniques: Identifying Learning in an Introduction to Proofs Course* (Doctoral dissertation, Middle Tennessee State University)

Reid, D. A., & Knipping, C. (2010). *Proof in mathematics education: Research, learning and teaching*. Brill.

Rüede, C. (2013). How secondary level teachers and students impose personal structure on fractional expressions and equations—an expert-novice study. *Educational Studies in Mathematics, 83*(3), 387-408. Retrieved from http://www.jstor.org/stable/23434899

Rø, K., & Arnesen, K. K. (2020). The opaque nature of generic examples: The structure of student teachers' arguments in multiplicative reasoning. *The Journal of Mathematical Behavior*, *58*, 100755. https://doi.org/10.1016/j.jmathb.2019.100755

Sandefur, J., Mason, J., Stylianides, G., & Watson, A. (2013). Generating and using examples in the proving process. *Educational Studies in Mathematics*, *83*(3), 323-340. https://doi.org/10.1007/s10649-012-9459-x

Schifter. (Eds.), *A* research *companion to Principles and Standards for School Mathematics* (pp. 22-44). Reston, VA:National Council of Teachers of Mathematics.

Selden, J., & Selden, A. (1995). Unpacking the logic of mathematical statements. *Educational Studies in Mathematics,* 29(2), 123-15l. https://doi.org/10.1007/BF01274210

Selden, A., & Selden, J. (2003). Validations of proofs considered as texts: can undergraduates tell whether an argument proves a theorem? *Journal for Research in Mathematics Education*, *34*(1), 4-36. https://doi.org/10.2307/30034698

Senk, S. L. (1989). Van Hiele levels and achievement in writing geometry proofs. *Journal for Research in Mathematics Education, 20,* 309-32l. https://doi.org/10.5951/jresematheduc.20.3.0309

Sfard, A. (1992). Operational origins of mathematical objects and the quandary of reification-the case of function. *The concept of function: Aspects of epistemology and pedagogy*, *25*, 59-84.

Shield, M., & Dole, S. (2013). Assessing the potential of mathematics textbooks to promote deep learning. *Educational Studies in Mathematics, 82*(2), 183-199. Retrieved from http://www.jstor.org/stable/23434868

Simpson, A., & Stehlíková, N. (2006). Apprehending Mathematical Structure: A Case Study of Coming to Understand a Commutative Ring. *Educational Studies in Mathematics, 61*(3), 347-371. Retrieved from http://www.jstor.org/stable/25472075

Sowder, L., & Harel, G. *(2003).* Case studies of mathematics majors' proof understanding, construction, and appreciation. *Canadian Journal of Science, Mathematics and Technology Education, 3*, 251-267. https://doi.org/10.1080/14926150309556563

Steele, M. D. (2005). Comparing knowledge bases and reasoning structures in discussions of mathematics and pedagogy. *Journal of Mathematics Teacher Education*, *8(4),* 291–328. https://doi.org/10.1007/s10857-005-0854-4

Steffe L. P. (1994) Children’s multiplying schemes. In: Harel G. & Confrey J. (eds.) The development of multiplicative reasoning in the learning of mathematics. SUNY Press, New York: 3–39.

Stemler, S. E., & Tsai, J. (2008). Best practices in interrater reliability: Three common approaches. *Best practices in quantitative methods*, 29-49.

Stephan, M., & Rasmussen, C. (2002). Classroom mathematical practices in differential equations. *Journal of Mathematical Behavior*, *21*, 459–490. https://doi.org/10.1016/S0732-3123(02)00145-1

Stranieri, A., & Zeleznikow, J. (1999). A survey of argumentation structures for intelligent decision support. In R. Burstein (Ed.), *Proceedings of 5th international conference of the international society for decision support systems ISDSS'99*. Melbourne: Monash University.

Stylianides, A. J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education, 38(3),* 289-32l. https://doi.org/10.2307/30034869

Stylianides, A. J. *(2016). Proving in the elementary mathematics classroom.* Oxford, United Kingdom: Oxford University Press.

Stylianides, A. J., & Stylianides, G. J. (2009). Proof Constructions and Evaluations. *Educational Studies in Mathematics*, *72*(2), 237–253. http://www.jstor.org/stable/40284620

Stylianides, G. J., Stylianides, A. J., & Philippou, G. N. *(2007).* Pre service teachers' knowledge of proof by mathematical induction. *Journal of Mathematics Teacher Education, 10,* 145-166.

Stylianides, G. J., Stylianides, A. J., & Weber, K. (2017). Research on the teaching and learning of proof: Taking stock and moving forward. *Compendium for research in mathematics education*, (pp. 237-266).

Stylianou, D. A. (2010). Teachers' conceptions of representation in middle school mathematics. *Journal of mathematics Teacher education*, *13*, 325-343. https://doi.org/10.1007/s10857-010-9143-y

Stylianou, D. A. (2011). An examination of middle school students' representation practices in mathematical problem solving through the lens of expert work: Towards an organizing scheme. *Educational Studies in Mathematics*, *76(3),* 265-280. https://doi.org/10.1007/s10649-010-9273-2

Tapan, M. S., & Arslan, C. (2009). Preservice teachers' use of spatio-visual elements and their level of justification dealing with a geometrical construction problem. *US-China Education Review, 6(3),* 54-60.

Thomas, N., Mulligan, J. T., & Goldin, G. A. (2002). Children's representations and cognitive structural development of the counting sequence 1-100. *Journal of Mathematical Behavior, 21*, 117-133. https://doi.org/10.1016/S0732-3123(02)00106-2

Toulmin, S. E. (1958). The use of arguments. Cambridge: University Press.

Vergnaud G. (1992) Conceptual Fields, Problem Solving and Intelligent Computer Tools. In: De Corte E., Linn M.C., Mandl H., Verschaffel L. (Eds.) *Computer-Based Learning Environments and Problem Solving.* NATO ASI Series (Series F: Computer and Systems Sciences), vol 84. Springer, Berlin, Heidelberg

Vergnaud, G. (1998). A comprehensive theory of representation for mathematics education. *Journal of Mathematical Behavior, 17*, 167-181. https://doi.org/10.1016/S0364-0213(99)80057-3

Warren, E. (2003). The role of arithmetic structure in the transition from arithmetic to algebra. *Mathematics Education Research Journal, 15*(2), 122-137. doi:http://dx.doi.org/10.1007/BF03217374

Warren, E., & Miller, J. (2013). Young australian indigenous students' effective engagement in mathematics: The role of language, patterns, and structure. *Mathematics Education Research Journal, 25*(1), 151-171. doi:http://dx.doi.org/10.1007/s13394-013-0068-5

Weber, K. (2006). Investigating and teaching the processes used to construct proofs. *Research in Collegiate Mathematics Education, 6*, 197-232.

Weber, K. (2008). How mathematicians determine in an argument is a valid proof. *Journal for Research in Mathematics Education*, *39*(4), 431-459. https://doi.org/10.5951/jresematheduc.39.4.0431

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

Weber, K., & Alcock, L. (2005). Using warranted implications to understand and validate proof. *For the Learning of Mathematics*, *25(1),* 34–38. https://www.jstor.org/stable/40248484

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

Weiss, M., & Herbst, P. (2007). Every single little proof they do, you could call it a theorem": Translation between abstract concepts and concrete objects in the Geometry classroom. *annual meeting of the American Educational Research Association*, Chicago, IL,

Wilkie, K. J., & Clarke, D. M. (2016). Developing students' functional thinking in algebra through different visualisations of a growing pattern's structure. *Mathematics Education Research Journal, 28*(2), 223-243. doi:http://ida.lib.uidaho.edu:2153/10.1007/s13394-015-0146-y

Yackel, E. (2001). Explanation, Justification and argumentation in mathematics classrooms. In M. Van den Heuvel-Panhuizen (Ed.), *Proceedings of the 25th Conference of the International Group for the Psychology of Mathematics Education PME-25* (vol. 4) (pp. 33–40).

Yackel, E., & Rasmussen, C. (2002). Beliefs and norms in the mathematics classroom. In G. Leder, E. Pehkonen, & G. Toerner (Eds.), Beliefs: A hidden variable in mathematics education? (pp. 313–330). Dordrecht, The Netherlands: Kluwer.

Yackel, E., & Hanna, G. (2003). Reasoning and proof. In J. Kilpatrick, W.G. Martin, & D.

Young-Loveridge, J. (2002). Early childhood numeracy: Building an understanding of part-whole relationships*. Australian Journal of Early Childhood, 27(4),* 36-42.

Yopp, D. A. (2020). Eliminating counterexamples: An intervention for improving adolescents' contrapositive reasoning. *The Journal of Mathematical Behavior*, *59*, 100794. https://doi.org/10.1016/j.jmathb.2020.100794

Yopp, D. A., & Ely, R. (2016). When does an argument use a generic example? *Educational Studies in Mathematics*, *91*(1), 37-53. https://doi.org/10.1007/s10649-015-9633-z

Yopp, D. A., Ely, R., Adams, A. E., & Nielsen, A. W. (2022). Proof in the Context of Middle Grades: Can We Label Middle School Arguments as Proof with a Capital P?. In *Conceptions and Consequences of Mathematical Argumentation, Justification, and Proof* (pp. 109-127). Cham: Springer International Publishing.

Zandieh, M., Roh, K.H., & Knapp, J. (2014). Conceptual blending: Student reasoning when proving "conditional implies conditional" statements. *The Journal of Mathematical Behavior,* 33, 209-229. https://doi.org/10.1016/j.jmathb.2013.11.007

Zaslavsky, O., & Shir, K. (2005). Students' conceptions of a mathematical definition. *Journal for Research in Mathematics Education*, *36*(4), 317-346. https://doi.org/10.2307/30035043

Zazkis, R., & Liljedhal, P. (2004). Understanding the primes: the role of representation. *Journal for Research in Mathematics Education, 35*,164-168. https://doi.org/10.2307/30034911

Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science, 2*, 179-217. https://doi.org/10.1016/S0364-0213(99)80022-6