# Impediments to mathematical creativity: Fixation and flexibility in proof validation

Abstract: Mathematical techniques in proof writing can be narrowed down to specific proof styles. Simply put, proofs can be direct or indirect- the latter using the Law of the Excluded Middle from logic as well as the axiom of Choice, to prove existence of mathematical objects. However, the thinking skills involved in writing indirect proofs are prone to errors, especially from novice proof writers such as prospective teachers. Creativity in mathematics entails the use of both direct and indirect approaches to determine the validity of a statement. In this article, I shed some light on this relationship, by reporting on some findings from a study on how students comprehend and validate direct and indirect proofs. Furthermore, I use the constructs of fixation and flexibility from creativity research to examine student approaches to direct and indirect proofs.

Keywords: Fixation; Indirect proof; proof validation; proof comprehension

## 1 Introduction

Creativity researchers have often conceptualized and investigated creativity as a de facto static phenomenon through fixed measures (see for instance Sriraman & Haavold, 2017). These approaches have contributed invaluably to our understanding of creativity, but they offer limited insight into the dynamic and multifaceted nature of creativity (Sriraman, 2009). Recently it has therefore been argued that a more dynamic, micro-longitudinal approach to studying creativity is needed – particularly with a focus on creativity in classrooms (Beghetto & Karawowski, 2019). In this study, I adopt this process-view of creativity and take a dynamic approach to investigate a key aspect of creativity – cognitive flexibility – in the setting of mathematical proofs.

Creative behavior always involve some form of uncertainty and surprise (Beghetto & Karawowski, 2019). Cognitive flexibility refers to our ability to switch between different mental sets, tasks and strategies in light of this uncertainty. It is considered a key characteristic of both human cognition and in models of creativity (Ionescu, 2012). Cognitive fixation, on the other hand, is the counterpart to flexibility. The notion of people struggling to come up with creative solutions because they fixate, or fail to abandon non-productive strategies, has its roots a long way back in psychological literature and features particularly in the writings of the Gestalt school (Haylock, 1987). This effect, commonly known as Luchin's Einstellung effect (1942), is seen when the subject continues to apply that method or process even when it becomes inappropriate, inefficient or unsuccessful. The capacity of breaking from mental sets overcoming fixations and mental rigidity are frequent themes in discussions of the creative process. Recently, Nijstaad et al. (2010) proposed a model that argues that creativity, as the generation of original and appropriate ideas, is the result of cognitive flexibility and cognitive persistence. According to Nijstaad et al. (2010), cognitive flexibility is a key element for achieving creative insights, problem solutions, or ideas through the use of flexible switching among categories, approaches, and sets, and through the use of remote (rather than close) associations.

Research on the teaching and learning of proof has been a focus of mathematics education research for more than four decades. Tall's (1979) study on university students' preferences for a particular type of proof of contradiction viz., a generic proof versus a standard proof for the irrationality of $\sqrt{2}$ suggested no significant difference in their preference. However when showed a different irrational number other than $\sqrt{2}$, the students preferred a generic approach. Researchers have argued that this observation applies to indirect proofs in general, as students both dislike and experience a lack of conviction from them (see for instance Harel & Sowder, 1998; Leron, 1985).

Several explanations have been proposed for this observation. Leron (1985) pointed to the non-constructive nature of indirect proofs. Understanding a proof depends on the construction of mental entities corresponding with the mathematical objects or symbols in the proof. However, in an indirect proof the learner works within a "false impossible world". According to Leron (1985) this leads to a detachment from a "real mathematical world" created within our minds, and it consequently creates a cognitive strain that makes indirect proofs particularly problematic. Antonini and Mariotti (2008), on the other hand, argued that students' struggled with indirect proofs at the level of logical theorems. To know truth in a mathematical sense requires both a mathematical and logical theory. Indirect proofs are unique in the sense that they require learners to reason within logical meta-theorems, such as $(\neg q \rightarrow \neg p) \equiv (p \rightarrow q)$.

In this article, I report on some findings from a study on how students comprehend and validate direct and indirect proofs in the context of cognitive flexibility and attempt to answer the following questions:

1. *How do students validate and comprehend direct and indirect proofs?*
2. *Are students able to adapt their approaches to proof validation and proof comprehension flexibly according to the nature of the proofs?*

When students read mathematical proofs, the process can be separated into two related purposes: comprehension and validation (Weber and Mejia-Ramos, 2011). For proof comprehension, the purpose is primarily to understand the content of the proof and learn from it. This positions the reader in the role of a traditional learner, where the purpose is to make sense of and understand a proof. For proof validation, on the other hand, the reader takes the position of a critic, who takes evaluative authority and judges the argument. Here, the purpose is to determine the correctness of the proof (Selden & Selden, 2017). While it is not clear how comprehension and validation are related, the latter requires the reader of the proof to be explicit about his or her ideas whereas comprehension is more of an internal process.

However, recent eye-movement analyses from Parse et al. (2018) indicate that proof comprehension and proof validation do not involve different reading behavior. Although these findings do not provide direct information about conscious experience, they do at least strongly suggest that comprehension and validation are closely related. In this study, I do not presume a specific relationship between proof comprehension and proof validation, but instead see them as two intertwined processes involved in the reading of proofs. By explicitly conceptualizing and investigating how students attempt to both comprehend and validate direct and indirect proofs, respectively, this study may be able to provide some more insight on whether students find indirect proofs less convincing than direct proofs.

# 2 Theory

## 2.1 Cognitive fixation and flexibility

Although cognitive flexibility seems like an intuitive concept, there is still a lack of a clear definition and comprehension of the phenomenon (Ionescu, 2012). However, the Handbook of Behavioral Neuroscience (2016) broadly outlines cognitive flexibility as the ability to adapt behaviors in response to changes in the environment. In a review of the literature, Ionescu (2012) identified several behaviors that are considered cognitive flexible: switching between tasks or multitasking; changing behavior in light of a new rule; finding a new solution to a problem; and creating new knowledge or tools. In this paper, I focus on one particular aspect of cognitive flexibility that is especially important to the field of mathematics. All mathematics educators will have experienced children or students who stubbornly stick to inappropriate methods or strategies (see for instance Haavold, 2011). One explanation for this is that they are subject to a mental set and their thinking is fixated along inappropriate lines

(Haylock, 1987). In this paper, flexibility is seen as the ability to break away from unproductive mental sets and fixations when solving mathematical problems. Krutetskii (1976) noted that this type of flexibility was a key component of general mathematical ability: "Mathematical ability appears in varied approaches to the solution of a problem and in easy and free switching from one mental operation to another. The talented student is able, when necessary, to leave the patterned stereotyped means of solving a problem and find a few different ways of solving it ... this is the real appearance of mathematical creativity." (p. 117).

In a review of the literature on fixations in mathematics, Haylock (1987) concludes that there are at least two important types of fixation that are applicable to working on mathematical problems: algorithmic fixation and content universe fixation. Algorithmic fixation is closely related to the Einstellung effect (Luchin, 1942), and it refers to when students continue to use an initially successful algorithm learnt beforehand or developed through the sequence of tasks themselves. The other type of fixation, content universe fixation, is in particular relevant for the study reported here. According to Haylock (1987), this type of fixation refers to situations where students' thinking about mathematical problems is restricted unnecessarily to an insufficient range of elements that may be used or related to the problem. To overcome this kind of fixation, and to allow the mind to range over a wider set of possibilities than might first come to one's conscious awareness, is an important aspect of problem solving and creativity.

## 2.1 Proof comprehension

Proof comprehension means understanding a proof (Selden & Selden, 2017). However, as mentioned, most research on reading of proofs have focused on whether or not students are convinced by mathematical arguments, and their corresponding rationale. Although several studies have proposed explanations for why students struggle with validating proofs, few have sought out to investigate how students try to understand proofs (Mejia-Ramos et al., 2012).

The reading comprehension model for high school geometry proofs, proposed by Yang and Lin (2008), was the first attempt to explicitly define and analyze what it could mean to understand mathematical proofs. In this model, the authors proposed that readers attempt to comprehend a text by integrating new information into pre-information, through a cyclic process of deduction, induction, abduction, and selection and memorization. Mejia-Ramos et al. (2017) argued that the model proposed by Yang and Lin (2008) is insufficient to probe students' understanding of proofs outside high school geometry. For instance, the model does not take into account how logical nuances in a proof by contradiction should be understood.

Furthermore, Yang and Lin's (2008) model does not assess whether students are able to summarize the main ideas behind a proof, or use them in other cases. Both of which are important to prospective teachers, as they must be able to summarize big ideas, and approach problems and solutions flexibly in their classrooms.

Mejia-Ramos et al (2012) proposed a model, built on the work of Yang and Lin (2008), which takes into account these issues. In this model, Mejia-Ramos et al. (2012) define proof comprehension as the types of understanding that were valued by mathematicians and mathematics educators in the literature. A proof can be understood either locally as a series of individual deductions, or holistically based upon the ideas or methods that motivate the proof in its entirety. In the model, there are seven dimensions to understanding a proof. These seven understandings are divided into local and holistic understandings. Local understandings of proof deal with aspects that can be garnered by reading a small number of statements in the proof. The local dimensions are:

1. Meaning of terms and statements: Understanding the meaning of terms and individual statements of the proof. This includes stating the definitions of terms used in the theorem statement and proof and identifying trivial implications of a given statement.
2. Justification of claims: Understanding why each claim made in the proof follows from previous ones, and being able to identify claims that follow from a given statement later in the proof. In this study, this applies in particular to algebraic operations, since the proofs are fairly simple.
3. Logical status of statements and proof framework: Understanding the logical relation between the assumptions and conclusions in a proof, identifying the proof technique being used, and conceptualizing the proof in terms of its proof framework.

Holistic understandings of proof focus on the big ideas of the proof, or synthesizing the entire or large parts of the proof into a coherent whole. The global dimensions are:

1. Identifying the modular structure: Understanding how a proof can be broken into mathematically independent parts or sub-proofs, and how these parts logically relate to one another
2. Illustrating with examples: Understanding how a sequence of inferences can be applied to verify that a general theorem is true for a specific example.
3. Summarizing via high-level ideas: Understanding the overarching logical structure of the proof and being able to summarize a proof in terms of these ideas.

4. Transferring the general ideas or methods to another context: Being able to use the ideas or methods in the proof to establish a different theorem.

## 2.2 Proof validation

Much of the research on students' work on proofs is based on asking students to construct proofs for specific conditional mathematical statements (Stylianides et al., 2017). Proof validation, on the other hand, is a less studied approach for investigating students' understanding of proof. Although proof construction and proof validation are both important to understanding proofs, research has shown that validating proofs is different from constructing proofs (Selden & Selden, 2017). When students attempt to validate proofs, they read and determine whether arguments and sub-arguments are correct, and this provides a different perspective on students' understanding of proofs.

It should be noted here that validating a proof is more complex than a simple top-down reading of the proof and claiming the proof is either valid or invalid. Nevertheless, in this study I limit my conceptualization of validation to what can be inferred from observations. I do this by utilizing Harel and Sowder's (1998) ideas about proof schemes. According to them, a person's proof scheme "consists of what constitutes ascertaining and persuading for that person" (1998, p. 244). In this study, proof validation is therefore conceptualized as the judgment of the validity of a proof or argument, and the corresponding explicit justification.

A common finding in the literature is that students often accept logically invalid deductions, confuse evidence and proof, focus too much on surface properties, and concentrate on algebraic manipulations (Knuth, 2002; Weber, 2010; Hodds et al., 2014; Selden & Selden, 2017). Selden and Selden (2003) for instance investigated how eight mathematics and mathematics education mathematics majors read and evaluated four student-generated proofs of a single theorem. They found that the students tended to focus on surface properties such as computations and algebraic notations, rather than underlying logical structures. Another common theme in the recent literature is that these problems are caused by either a confusion of the relationship between empirical evidence and deductive proofs, or because the students lack the skill or will to evaluate an argument properly (Inglis & Alcock, 2012). Much of the related research suggests that students find empirical arguments convincing and believe they constitute an acceptable form of mathematical proof (See for instance Recio & Godino, 2001; Knuth, 2002; Segal, 2000). However, the connection between the types of arguments that students construct and what they actually find convincing is not straightforward. Harel and Sowder (1998) interviewed students in advanced mathematics courses in various settings, and found that they regularly tried to prove by providing empirical examples. Weber (2010), on the other hand, found that few students believed empirical arguments could be valid proofs. Instead, many of the students judged invalid deductive arguments to be valid proofs. He concluded that the difficulties with proof validation were primarily related to skill at validating deductive arguments and not to the misconception that non-deductive arguments might constitute valid proofs. A possible explanation for this apparent contradiction is that sometimes students have no recourse but to use examples as means of justifications, simply because deductive arguments are too difficult.

### 2.3 Indirect Proofs

In this study indirect proofs refer predominantly to the method of reductio ad absurdum which can be traced back to Book I of The Elements. Indirect proofs refer to both proofs by contradiction as well as proofs by contraposition. Both proof by contradiction and that of contraposition have common elements. To prove *P=>Q* is true using a proof by contradiction one starts by assuming *P* and *not Q* and works towards a contradiction. The nuance here is that one does not know where the contradiction will occur. In a proof by contraposition, one assumes not Q and works towards the conclusion of not P, thereby contradicting some known statement. In the latter the contradiction is already known whereas in the former the contradiction is not known, making it more difficult to comprehend the chain of reasoning.

## 3 Methods

### 3.1 Participants and procedure

The participants in the study were 18 pre-service teachers enrolled in a five year pre-service teacher education program, specifically aimed at teaching in secondary school. The students were not mathematics specialists, but were expected to graduate with a Master's degree in mathematics education. Prior to the mathematics education course on proof and deductive reasoning, the students had 60 credits of mathematics (geometry, algebra, number theory, calculus, probability theory). All 18 students had above median grades in the previous mathematics courses and other subjects, and they had specific course modules on proofs in mathematics.

I collected the data for this paper through pairwise task based interviews (Goldin, 1997). The interviews lasted for about 60-70 minutes, and 18 students – nine pairs – participated. The pre-service teacher students were all volunteers, and they were asked to partner up with another student whom they worked and collaborated well. In the interview each pair of students were presented two theorems and six mathematical proofs and asked three open-ended questions: a) what do they think about the theorems, b) what do they think about the proofs, and c) do they have any other comments or thoughts. The students were then asked if the proofs were correct, and to work on each of the proofs as they normally would, without any help from the interviewer. This section of the interview followed a think aloud protocol (Kuusela & Pallab, 2000). During this part of the interview, the interviewer answered clarification questions, but deflected more task specific and content related questions back to the students.

The six proofs were separated into three proofs for each of two mathematical theorems (see table 1). Eventually, when the conversation between the two students came to a stop, the interviewer asked a series of questions from an interview guide based on the theoretical framework on proof comprehension (Mejia-Ramos et al., 2012).

### 3.2 Materials

In laboratory settings, cognitive flexibility is typically investigated using task-switching paradigms (Ionescu, 2012). In this paradigm, participants are required to alternate between two or more tasks. The materials used in this study were two proof validation tasks that consisted of two theorems, with three corresponding proofs for each of them, and an interview guide consisting of ten open-ended questions. Both the proof validation tasks and the interview guide was designed research questions in mind, and in accordance to the guide suggested by Mejia-Ramos et al. (2017). More specifically, I used the comprehension categories described in the framework and the associated template questions to generate proof tasks and open-ended questions that could reveal the students' understanding and conviction of proofs. This meant that I chose mathematical theorems that were appropriate for the students' academic level, and could be comprehended fairly quickly. The corresponding proofs were selected so they could elicit students' understanding of proofs at both the local and holistic level. I also selected both correct and incorrect proofs, at both a local and holistic level, so that the validation process would be genuine.

Table 1:

Proof 1A establishes the converse relationship of theorem 1. The first proof is therefore incorrect. The purpose of the proof was to see if the students were able to notice the incorrect logical relationship. Furthermore, proof 1A was in superficially similar to proof 1B. Proof 1B is a contrapositive proof of the theorem in task 1. However, there is a small error error in the proof, so neither proof 1A nor proof 1B are entirely correct. These two flawed proofs were chosen to investigate more closely how the participants attempted to make sense of the both the local aspects and holistic aspects of the proofs (Mejia-Ramos et al., 2012).

Proof 1C in task 1 is a short, direct proof. The key observation is that if n is an integer, then the right side of the equation, n(n+1), must be an even number. The proof is purposely presented without a detailed explanation. The reason is that I wanted to see if the students were able to identify and utilize the abstract information in the proof, and flexibly approach the algebraic expression – in other words, comprehend both the local and the holistic aspects of the proof.

Proof 2A is a proof by contradiction and in many ways an algebraic variant of the Pythagoreans' first geometric proof of that the square root of two is irrational. The proof is also probably the most common proof of the irrationality of $\sqrt{2}$. However, the proof contains a small algebraic error in line eleven. The purpose was to see if the students focus on the overarching proof structure or the internal correctness of each mathematical inference in the proof. Proof 2B first demonstrate that $\sqrt{2}=1+1/(1+\sqrt{2})$, and then that the right side expression is an infinite continued fraction.

Proof 2C uses a similar approach as proof 2A, but the contradiction is demonstrated with the fundamental theorem of arithmetic, which says every integer greater than 1 has a unique factorization into powers of primes. While proof A emphasizes logical inexorableness, proof B "seems to reveal the heart of the matter" (Davis et al., 2011, p. 331), This proof was chosen to see if an aesthetic component, such as clarity and simplicity, influenced how the students attempted to validate and comprehend it.

### 3.3 Analysis

The interviews were analyzed retrospectively using qualitative content analysis (Mayring, 2015). First, I transcribed all nine interviews, and analyzed each interview separately. For each interview, I extracted all text components captured by the categories defined in the proof comprehension framework (Mejia-Ramos et al., 2012). Using the category system in the framework, I was able to assess and characterize each pair of students' way of comprehending the proofs according to the local and holistic dimensions. The purpose here was to get a clear description of which comprehension dimensions the students used, and even more importantly, which comprehension dimensions they did not use.

To identify how students validated the proofs, I built categories that summarized whether the students assessed the proof as valid or invalid, and their justification for this assessment (Toulmin, 1969). This was done by looking at the statements each student made for each theorem and proof, and summarizing the text excerpts into a specific category. I then worked through each of the other students similarly, and checked whether the proof validation fell under the same category. If not, I built a new category. This resulted in categories that stated whether the students viewed the proof as valid or invalid, and a description of the corresponding justification.

To answer the research questions, I first summed up the validation and comprehension categories used by the students for each proof. This provided a general overview of how students validated and comprehended direct and indirect proofs respectively. I then looked more closely at the pattern between the nature of the proofs and the comprehension and validation categories. This allowed me to see if the students' approached indirect and direct proofs flexibly. Finally, I isolated individual student statements to identify possible explanations for how the nature of the proofs might hinder or facilitate cognitive flexibility.

# 4 RESULTS

To illustrate the underlying analysis of the students' work, I first present some examples of how the analytical framework was used to categorize students' comprehension and validation of the proofs.

### 4.1 Students' work

*How students validated proofs*

Four categories of how the students justified their proof validation was identified – although not every category was observed for every proof. The first I called *because of logical conditions*, and it refers to cases where students rejected or accepted a proof by explicitly referencing the logical status of statements in the proof. An example of this category is found in proof 1B. Here, Sofie told Carrie: "this proof can't be correct, because they make a completely different assumption here. In the original statement, they say that if n squared is even." The students rejected the validity of the proof due to the logical status of the assumption in the proof. The second category is *because of empirical verification*, and it refers to cases where students accepted or rejected a proof based on numerical examples. In proof 1A, for instance, one student Sofie said "So proof A is correct, because we have calculated it." and Carrie answered "yes, it worked for all numbers we tried." The third category, because of *algebraic verification*, points to attempts to work through the proof line by line, by assessing the algebraic transformations. For proof 1A, Caroline stated for instance, to Christian that the proof was valid because: "we did the algebraic operations it and saw that the algebraic statements on each side of the equal sign was the same. Every algebraic manipulation was correct, so the proof is correct as well."

Each of the three categories was separated further into whether the students considered the proof valid or invalid. The fourth category is *proof is incomprehensible*, because the proof statements doesn't make sense. It was singled out as a separate category, as it covered incidents when the students explicitly stated that they were unable to assess a proof as either valid or invalid; the reason being that certain statements in the proof were incomprehensible to the students.

*How students comprehended proofs*

Although the students' use of comprehension dimensions varied greatly, all of the comprehension dimensions were identified in the analysis – except for holistic dimension 1 (H1). I provide here an excerpt here to illustrate how I categorized the students' use of proof comprehension dimensions local 1 (L1), local 2 (L2) and holistic 2 (H2). The other dimensions (L3, H3 and H4) were identified in a similar manner.

Eight students used dimensions L1, L2, and H2 in order to comprehend proof 1A. Sofie and Carrie approached the proof in the following way:

Sofie: What if we try a few numbers first, and see what happens? Let's first try a simple number.
Carrie: Let's use n equals ten. If n is an even number we can write two times a number.
Sofie: That means n, or 10, equals two times five. N squared is then two times five squared, which is 100.
Carrie: It'll work with other numbers as well. For instance six gives us 36.
Sofie: 100 gives us 10000. All of which are even numbers.
Carrie: we also know that two k squared is two k times two k, which is 4 k squared…
Sofie: which is two times two k squared. So this is also correct.
Carrie: agreed.
Sofie: So proof A is correct, because we have calculated it.
Carrie: yes, it worked for all numbers we tried.
Sofie: yes.

This exchange between Sofie and Carrie started with inserting specific even numbers for $n$ and then verifying that $n^2$ is also an even number. I classified this part of the exchange as both dimensions L1 and H2. The reason is that the students identified both examples that illustrated the assumption and conclusion statement of the proof, and also illustrated a sequence of inferences with a specific numerical example. Of course, in short proofs like this, it is difficult to draw a clear line between L1 and H2. In this study, I have classified any attempt at understanding individual and isolated terms and statements, including the use of examples to illustrate individual statements, as L1. However, if a numerical example was used to illustrate a sequence of inferences across several individual statements, I classified it as H2. In the second half of this exchange, Sofie and Carrie verified algebraically why each statement in the proof followed from the previous one, which is comprehension dimension L2.

### 4.2 Relationship between proofs and students work

Table 2:

There were 54 proof validation attempts of direct and indirect proofs respectively (table 2). For the indirect proofs, the students validated the proofs 34 times correctly and 20 times incorrectly. For the direct proofs, the students validated the proofs 30 times correctly and 24 times incorrectly. Immediately, it would seem that indirect proofs are not more difficult than direct proofs.

However, when the comprehension dimensions the students used and how they justified their validation attempts are taken into account, the pattern becomes more nuanced. There were two main differences in how students attempted to comprehend direct and indirect proofs. First, all, or nearly all, students employed both L1 and L2 for the direct proofs. For the indirect proofs, however, L2 is only used a total of 18 times. Second, the students rarely used the holistic comprehension dimensions, and when they were used, they were primarily used for direct proofs. When the students justified their validation attempts, they relied primarily on the logical conditions of the proofs when they validated indirect proofs. 42 proof validations were justified based on logical conditions of statements in the proofs, while 12 proof validations were justified based on algebraic verification. For direct proofs, the students based their justifications on logical conditions 12 times, algebraic verification 24 times, and empirical verification 8 times.

These findings indicate that students approach indirect and direct proofs differently. However, table 2 only shows correct uses of comprehension dimensions. In many cases, the students tried, but failed, to use certain comprehension dimensions when they validated the proofs. Furthermore, the students were given an unequal number of direct and indirect proofs for each theorem, and the proofs were both valid and invalid, with both local and holistic errors. To answer the research questions, I therefore have take into account the relationship between the students' justification and each proof, as well as the students' attempts at using other comprehension dimensions.

In proof 1A, only four students recognized the erroneous logical structure of the proof. The other 14 students concluded the proof was valid, after verifying it either algebraically or empirically. Six of the students concluded 1B was invalid, due to a small algebraic error. However, and more interestingly, none of the six students used L3 and considered the logical structure of the proof. An additional ten students said the proof was invalid, because of the logical conditions of the proof. These ten students attempted to use L3, but concluded that the proof was invalid because the proof proved a different logical relationship than the original statement. For proof 1C, ten students said the proof didn't make any sense because the assumption and conclusion were stated simultaneously in the proof. The ten students used dimensions L1 and L2. They also attempted to use L3, but failed to understand the logical conditions of the statements.

Six students concluded that proof 2A was invalid, after trying to verify it algebraically. These students used dimensions L1 and L2, but they never discussed the logical structure of the proof or used any other comprehension dimensions. Ten students said the proof was invalid because of the logical conditions of statements in the proof. These students used L1 and attempted, but failed, to use L3. For proof 2B, all 18 students concluded the proof was valid, after using L1 and L2. None of them considered the logical structure of the proof or individual proof statements. This was very different from proof 2C. Here, 16 of 18 students concluded the proof was invalid based on logical conditions. All 16 students attempted to use L3, but failed to understand the logical condition and meaning of proof statements.

### 4.3 General themes

In proof 1A, the relationship between the assumption and the conclusion is converse, and only four students' recognized this error. This lack of attention to or understanding of the logical structure of proofs, is also seen in the indirect proofs 1B, 2A and 2C. Proof 2C is a valid indirect proof, but only two students' concluded as such. As for proof 1B and 2A, 16 students assessed the validity of the proof correctly. However, it is important to point out that ten of the students, for each of the two proofs, rejected the proof erroneously based on logical conditions. Furthermore, six of the students, in each case, rejected the proofs based on algebraic manipulations, without ever considering the logical structure of the proof. It could be claimed that there was no need to evaluate the logical structure of proof 1B and 2A, as the students' validated the proofs correctly based on algebraic manipulations. However, if we compare proof 2A and 2C, then we see that this is probably not the case. Proof 2A and 2C are similar in structure, and proves the same theorem. The students who correctly rejected proof 2A based on algebraic errors, also concluded incorrectly that proof 2C was invalid.

The students' statements provide a plausible explanation for these findings. In proof 1B, 2A and 2C, a majority of the students explicitly stated that the proofs were confusing or invalid because they made assumptions that contradicted the original theorem. In proof 1B, for instance, ten students concluded that the proof was invalid because it proved a different logical relationship than the original theorem in task 1. Proof 1B is a contrapositive proof, and it could be that the students found it difficult to understand that "A ⇒ B" is equivalent to "not B ⇒ not A". Antonini and Mariotti (2008) argue that indirect proofs are unique, because they require learners to reason with theorems that are part of logical theory, or metatheorems. Weber (2010) observed similar behavior when students accepted invalid deductive proofs, as they neglected checking if assumptions and conclusions of the proof were aligned with assumptions and conclusions of the given theorem. However, the students in this study did not simply reject the indirect proofs because they were unable to link the assumptions and conclusions of the proof to the original theorem statement. In this study, the students rejected the proofs because the proofs made assumptions that contradicted the statements in the theorem. The same argument was seen in proof 2A and 2C, where a majority of the students expressed confusion or skepticism because the proof made an assumption that, in their view, contradicted the original statement.

A possible explanation for the students' problems is that their minds were set in an inappropriate direction, and they were too rigidly adhering to an approach that was not fruitful. The students' approach to the problems, were inflexible and too fixed on pre-determined attitudes (Krutetskii, 1976; Haylock, 1987). Haylock (1987) referred to this as content universe fixation. As mentioned earlier, this refers to situations where students' thinking about mathematical problems is restricted unnecessarily to an insufficient range of elements that may be used or related to the problem. A majority of the students' approach to proof 1C, which was direct, lend further support to this idea. Here, ten students expressed confusion as to how $n^2$ and $n$ could both be placed on the left side of the equation. One was the assumption and the other was the conclusion in the original statement. It seems that the students had pre-conceived ideas of how proofs should be constructed. A proof should make the same assumption as the statement which one is trying to prove, and through some step-by-step procedure arrive at the same conclusion as stated in the original conditional statement.

This also ties to the consistency hypothesis that argues that there are “rules” for governing what is accepted or rejected in a conceptual system within mathematics, and contradiction depends on meaning (Sierpinska, 2007). More specifically, if a statement is meaningless to a student, then the question of consistency becomes meaningless as well. In this study, a majority of students’ expressed confusion and skepticism as to how the indirect proofs could make assumptions that were different from the original statement. The cause of this needs to be further investigated. However, certain statements from the students allows us to propose a hypothesis. A majority of the students explicitly questioned the validity of the original assumptions in the indirect proofs. Working on 2C, for instance, Sandra asked “how can we assume something that is incorrect”. Similar phrases are found across a majority of the interviews with students when they were working on the indirect proofs. The students expressed doubts as to how the indirect proofs can make assumptions that differ from the statement they’re trying to prove. Furthermore, the interviews indicate that the students presume that the statement they’re trying to prove is correct, and therefore doubt how a proof can make assumptions that ostensibly contradict it.

This does, however, not explain why 14 of 18 students concluded proof 1A was correct. Proof 1A makes an initial assumption that is different from the original statement. But unlike the indirect proofs, the initial assumption in proof 1A is not the negation of the initial assumption in the original statement. It is possible that the students only rejected a proof if it made assumptions that clearly contradicted the original statement, and not proofs that simply made “different” assumptions.

## 5 Conclusion

On the surface, it would seem that indirect proofs are not more difficult than direct proofs. Students validate indirect and direct proofs correctly about the same rate. However, when the comprehension dimensions the students used, how they justified their validation attempts, and the validity of each proof, are taken into account, some differences between indirect and direct proofs appear. The students in this study tended to focus on the logical structure of the indirect proofs, while they focused more on algebraic and empirical verification of direct proofs. Furthermore, the students seemed to reject indirect proofs based on a preconceived idea of what a valid proof should look like. Based on the observations in this study, it seems the students think proofs should make the same assumption as the statement that we are trying to prove, and through some step-by-step procedure arrive at the same conclusion as stated in the original conditional statement. When the proofs diverged from this preconceived idea of how a proof should look like, the students' tended to reject it. I therefore propose that the students' approach to the problems were inflexible and too fixed on pre-determined mental sets related to mathematical proofs.

### 5.1 Limitations

Some limitations to this study are worth mentioning. First, the students were interviewed in pairs and not one-on-one. This was done to elicit natural talk between the students, to provide some insight into how the students were thinking. Although I did not observe that any of the pairs of students were particularly dominated by one of the students, it is possible that some students would have given different answers if they had been interviewed one-on-one. Second, it is possible that each student has his or her preferred way of reading proofs. However, how each student attempted to comprehend and validate the proofs varied across the proofs. This indicates that each student did not have a single preferred way of comprehending and validating the proofs, but instead adapted to each new situation. Lastly, in this study, some tasks were incorrect, some were correct. I cannot be sure about how students would have reacted to the same proofs if the situation had been altered. However, there are some indications of this, as some of the proofs were similar on a structural level, but dissimilar on a local level. For instance, by comparing the students' performance on proof 2A and proof 2C, it is possible to see how a local error in 2A influence the students' work.

## References


Antonini, S., & Mariotti, M. A. (2008). Indirect Proof: What is specific to this way of proving. ZDM – The International Journal on Mathematics Education, 40,401–412.

Beghetto, R. A., & Karwowski, M. (2019). Unfreezing creativity: A dynamic micro-longitudinal approach. In Dynamic Perspectives on Creativity (pp. 7-25). Springer, Cham.

Goldin, G. A. (1997). Chapter 4: Observing mathematical problem solving through task-based interviews. Journal for Research in Mathematics Education. Monograph, 40-177.

Haavold, P. Ø. (2011). What characterises high achieving students' mathematical reasoning?. In Sriraman, B. & Lee, K. (Eds.) The elements of creativity and giftedness in mathematics (pp. 193-215). Brill Sense.

Harel, G. & Sowder, L. (1998). Students' proof schemes. In E. Dubinsky, A. Schoenfeld, & J. Kaput (Eds.), Research in Collegiate Mathematics Education, Vol. III (234-283). Providence, RI: American Mathematical Society.

Haylock, D. W. (1987). A framework for assessing mathematical creativity in school children. Educational studies in mathematics, 18(1), 59-74.

Hodds, M., Alcock, L., & Inglis, M. (2014). Self-explanation training improves proof comprehension. Journal for Research in Mathematics Education, 45(1), 62-101.

Inglis, M., & Alcock, L. (2012). Expert and novice approaches to reading mathematical proofs. Journal for Research in Mathematics Education, 43(4), 358–390.

Ionescu, T. (2012). Exploring the nature of cognitive flexibility. New ideas in psychology, 30(2), 190-200.

Knuth, E. J. (2002). Secondary school mathematics teachers' conceptions of proof. Journal for Research in Mathematics Education, 33, 379–405.

Kuusela, H., & Pallab, P. (2000). A comparison of concurrent and retrospective verbal protocol analysis. The American journal of psychology, 113(3), 387.

Luchins, A. S. (1942), Mechanization in problem solving: The effect of Einstellung, Psychological Monographs 54(6).

Mayring P. (2015) Qualitative Content Analysis: Theoretical Background and Procedures. In Bikner-Ahsbahs A., Knipping C., Presmeg N. (eds) Approaches to Qualitative

Research in Mathematics Education. Advances in Mathematics Education. Springer, Dordrecht

Mejia-Ramos, J. P., Fuller, E., Weber, K., Rhoads, K., & Samkoff, A. (2012). An assessment model for proof comprehension in undergraduate mathematics. *Educational Studies in Mathematics*, 79(1), 3-18.

Mejia-Ramos, J. P., Lew, K., de la Torre, J., & Weber, K. (2017). Developing and validating proof comprehension tests in undergraduate mathematics. *Research in Mathematics Education*, 19(2), 130-146.

Nijstad, B. A., De Dreu, C. K., Rietzschel, E. F., & Baas, M. (2010). The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. European Review of Social Psychology, 21(1), 34-77.

Recio, A. M., & Godino, J. D. (2001). Institutional and personal meanings of proof. Educational Studies in Mathematics, 48(1), 83–99

Segal, J. (2000). Learning about mathematical proof: Conviction and validity. *Journal of Mathematical Behavior*, 18, 191–210.

Selden, A., & Selden, J. (2003). Validation of proofs considered as texts: Can undergraduates tell whether an argument proves a theorem? Journal for Research in Mathematics Education, 34, 4–36.

Selden, A., & Selden, J. (2017). A comparison of proof comprehension, proof construction, proof validation and proof evaluation. In Proceedings of the Conference on Didactics of Mathematics in Higher Education as a Scientific Discipline (pp. 339-345).

Sierpinska, A. (2007). I need the teacher to tell me if I am right or wrong. In J. H. Woo, H. C. Lew, K. S. Park, & D. Y. Seo (Eds.), *Proceedings of the 31st conference of the international group for the psychology of mathematics education* (vol 1) (pp. 45–64).

Sriraman, B. (2009). The characteristics of mathematical creativity. ZDM, 41(1-2), 13.

Sriraman, B., & Haavold, P. (2017). Creativity and giftedness in mathematics education: A pragmatic view. In first compendium for research in mathematics education. Reston: National Council of Teachers of Mathematics.

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

Tall, D. (1979). Cognitive aspects of proof, with special reference to the irrationality of √2. Proceedings of the third international conference for the psychology of mathematics education (pp. 206-207). Warwick.

Weber, K. (2010). Mathematics majors' perceptions of conviction, validity, and proof. Mathematical Thinking and Learning, 12(4), 306–336.

Weber, K., & Mejia-Ramos, J. P. (2011). Why and how mathematicians read proofs: An exploratory study. Educational Studies in Mathematics, 76(3), 329-344.

Yang, K.-L., & Lin, F.-L. (2008). A model of reading comprehension of geometry proof. Educational Studies in Mathematics, 67, 59–76.

Table 1: Proof validation tasks

| Task 1 | Task 2 |
|---|---|
| Theorem: Suppose $n$ is an integer. If $n^2$ is even, then $n$ is even. | Theorem: $\sqrt{2}$ is an irrational number. |
| Proof A: If $n$ is even, then we can write $n = 2k$. We then see that $n^2 = (2k)^2 = 4k^2 = 2 \cdot 2k^2$. Therefore, $n^2$ is even. | Proof A: Suppose that $\sqrt{2}$ is a rational number. Then we can write it as an irreducible fraction $\frac{a}{b} = \sqrt{2}$. |

| | |
|---|---|
| | We square both sides and see that $\frac{a^2}{b^2} = 2 \Longrightarrow a^2 = 2b^2$.<br>$2b^2$ is even, and therefore $a^2$ must also be even. It follows that $a$ is even, and we can write $a$ as $a = 2k$.<br>We substitute $a = 2k$ into $a^2 = 2b^2$, and see that $(2k)^2 = 2b^2$.<br>We then see that $4k^2 = 2b^2 \Longrightarrow 2a^2 = b^2$.<br>Therefore, $b^2$ must be even, and it follows that $b$ must be even as well.<br>We now have that both $a$ and $b$ are even. But that means $\frac{a}{b}$ is not irreducible, which contradicts our assumption. We therefore have to conclude that $\sqrt{2}$ is irrational. |
| Proof B: Suppose $n$ is not even. Then it is odd, and we can write $n = 2k + 1$.<br>We then see that $n^2 = (2k+1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + k) + 1$.<br>That means $n^2$ is also odd. We therefore have to conclude the theorem in task 1 is correct. | Proof B: We can write $\sqrt{2} = 1 + (\sqrt{2} - 1) = 1 + \frac{1}{1+\sqrt{2}}$.<br>Because $\frac{(\sqrt{2}-1)(\sqrt{2}+1)}{1(\sqrt{2}+1)} = \frac{1}{(1+\sqrt{2})}$.<br>It follows then that $\sqrt{2} = 1 + \frac{1}{1+\sqrt{2}} = 1 + \frac{1}{1+\left(1+\frac{1}{1+\sqrt{2}}\right)} = 1 + \frac{1}{2+\frac{1}{1+\sqrt{2}}}$,<br>However, we can again make the same substitution, and this expression is an infinite continued fraction:<br>$1 + \frac{1}{2+\frac{1}{1+\sqrt{2}}} = 1 + \frac{1}{2+\frac{1}{2+\frac{1}{1+\sqrt{2}}}} = 1 + \frac{1}{2+\frac{1}{2+\frac{1}{2+\frac{1}{1+\sqrt{2}}}}} =$<br>So $\sqrt{2}$ is therefore irrational. |

| | |
|---|---|
| Proof C: $n^2 + n = n(n+1)$ is even. Since $n^2$ is even, then $n$ must also be even. | Proof C: Suppose that $\sqrt{2}$ is a rational number. Then we can write it as an irreducible fraction $\frac{a}{b} = \sqrt{2}$.<br>We square both sides and see that $\frac{a^2}{b^2} = 2 \Longrightarrow a^2 = 2b^2$.<br>Every integer can be factored into primes, and we suppose this has been done for $a$ and $b$. Thus in $a^2$ there are certain number of primes doubled up. And in $b^2$ there are a certain number of doubled-up primes. But, in $2b^2$ there is a 2 that has no partner. We have a contradiction and must conclude that $\sqrt{2}$ is irrational. |

Table 2: Proof validations and proof comprehension dimensions for each proof

| | N | L1 | L2 | L3 | H1 | H2 | H3 | H4 |
|---|---|---|---|---|---|---|---|---|
| **Direct proofs** | | | | | | | | |
| Proof 1A | | | | | | | | |
| *Invalid logical* | 4 | 4 | 4 | 4 | | 4 | 4 | |
| *Valid empirical* | 8 | 8 | 8 | | | 8 | | |
| *Valid algebraic* | 6 | 6 | 6 | | | | | |
| Proof 2B | | | | | | | | |
| *Valid algebraic* | 18 | 18 | 18 | | | | | |
| Proof 1C | | | | | | | | |
| *Valid logical* | 8 | 8 | | 8 | | 8 | 8 | |
| *Incomprehensible* | 10 | 10 | 10 | | | | | |
| Total | 54 | 54 | 46 | 12 | | 20 | 12 | |
| **Indirect proofs** | | | | | | | | |
| Proof 1B | | | | | | | | |
| *Valid logical* | 2 | 2 | | 2 | | 2 | 2 | |
| *Invalid algebraic* | 6 | 6 | 6 | | | | | |
| *Invalid logical* | 10 | 10 | | | | | | |
| Proof 2A | | | | | | | | |
| *Valid logical* | 2 | 2 | | 2 | | 2 | 2 | |
| *Invalid algebraic* | 6 | 6 | 6 | | | | | |
| *Invalid logical* | 10 | 10 | | | | | | |
| Proof 2C | | | | | | | | |
| *Valid logical* | 2 | 2 | | 2 | | 2 | 2 | 2 |
| *Invalid logical* | 16 | 16 | 6 | | | | | |
| Total | 54 | 54 | 18 | 6 | | 6 | 6 | 2 |