Applying Proof Frameworks Is Harder for Students than We Think

Iris Gaber
School of Computer Science
Academic College of
Tel-Aviv Yaffo

Amir Kirsh
School of Computer Science
Academic College of
Tel-Aviv Yaffo

Romina Zigdon
School of Computer Science
Academic College of
Tel-Aviv Yaffo

In order to overcome difficulties many students in mathematics and computer science face with rigorous proofs, some higher education institutes present a "Mathematical Reasoning" or "Transition-to-Proof" course. In this paper, we analyzed data from the exams of 310 students in such a course, focusing on the way they approach a proof regarding set inclusion, and whether they use a particular "proof framework" that we taught them when doing so. We found that the students are able to use the proof framework, and that using it is beneficial to them, but that a surprisingly large percentage of them fail to use it consistently.

Keywords: Transition-to-proof courses, Proof frameworks, Operable interpretations of definitions

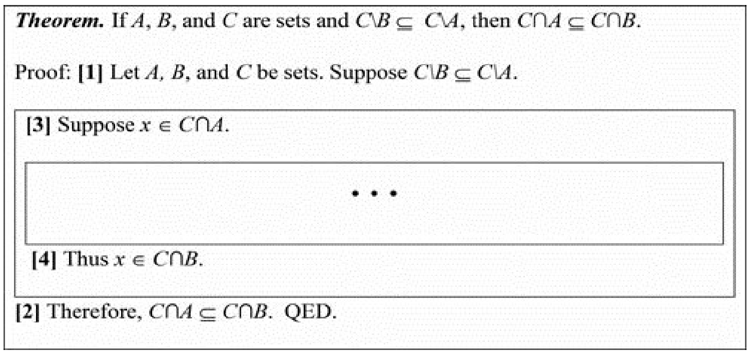
Introduction

One of the challenges mathematics instructors face is teaching their students to write correct rigorous proofs and to use the definitions properly. Between 1990 and 1999, more than 100 papers have been published on this subject (Hanna, 2000). The question of how students engage with proof and proving is continuously being studied (Reid and Knipping, 2011). These studies have suggested various explanations for the root of the problem, and proposed a variety of methods for addressing it, which have been tested with varying degrees of success. One common conclusion is that the transition from an intuitive answer to the rigorous formulation is one of the students' main sources of difficulty. Vinner (1991), for instance, claims that students tend to ignore the definitions necessary for a rigorous proof, preferring an intuitive "general explanation" answer. Moore (1994) supports this assertion, noting that, while attempting to write formal proofs, students do not necessarily understand the content of relevant definitions or how to use them.

One of the ways institutes deal with this issue is by having students take an "introduction to proof" course. Most such courses emphasize helping students learn the proof construction process (see in particular, Selden, Benkhalti and Selden, 2014 and Moore, 1994). This paper presents data collected in the context of the course "Mathematical Reasoning" - a mandatory course for first year computer science undergraduates. Specifically, our study focuses on the manner and extent to which undergraduate students employ - or fail to employ - a "proof framework" while proving a theorem. The concept of the proof framework was first introduced by Selden and Selden, 1995, and later expanded upon (2014). Similar ideas were presented by Scheinerman, 2013, who refers to it as a "proof template," and Hammack, 2013. In this paper we focus on one specific framework, used for proving set inclusion according to the definition. We analyzed 310 students' answers to questions that require this proof and discovered that a large percentage did not consistently use the proof framework, even when they had presented the ability to apply a very similar proof framework in an earlier part of the question, and despite the fact that this was the only technique they had learned for dealing with a question of that sort.

Proof Framework

Selden, Selden and Bankhalti (2018) define *proof frameworks* as "a way of structuring a proof, in which the student begins by writing the first and last lines of a proof and works towards the middle. It often consists of a first level, in which a student unpacks the logical structure of the theorem statement to write the first and last lines of the proof, and a second level, in which a student unpacks the definitions involved in the conclusion to write the second and second-to-last lines of the proof" (Selden, Selden and Bankhalti , 2018 p.4). To illustrate this point, they provide a sample proof framework for an elementary set theory theorem, which we have copied in its entirety here:


Theorem. If A, B, and C are sets and $C\backslash B \subseteq C\backslash A$, then $C \cap A \subseteq C \cap B$.

Proof: **[1]** Let A, B, and C be sets. Suppose $C\backslash B \subseteq C\backslash A$.

[3] Suppose $x \in C \cap A$.

• • •

[4] Thus $x \in C \cap B$.

[2] Therefore, $C \cap A \subseteq C \cap B$. QED.

Fig. 1. Proof framework of set inclusion

Clearly, a theorem regarding set inclusion should start with an element that belongs to the included set, and by a sequence of steps ends with the result that the element belongs to the including set. In their study, Selden, McKee, and Selden, 2010 discuss the habits of mind that appear to drive many mental actions in the proving process, and discuss the way feelings can help cause mental actions and also arise from them. They offer the following example of a wrong "proof," in the form of Sofia's answer to the following theorem.

Theorem: *For any sets* A, B, C *if* $A \subseteq B$ *then* $(A \cap C) \subseteq (B \cap C)$.
Sofia's Proof: (Selden et al., 2010, p.13) Let A, B, C be sets
Suppose $x \notin A, x \in B \text{ and } x \in C$
Then $x \notin (A \cap C)$, but $x \in (B \cap C)$
Therefore $(A \cap C) \subseteq (B \cap C)$

We were curious **how common** the approach presented by Sofia might be - namely starting wrong, such that it is impossible to produce a proof from that starting point. These authors' study presented insightful qualitative findings, based on a deep analysis of a small sample. Our study is based on an analysis of hundreds of students' exams, which allowed us to produce, in addition to qualitative findings, some interesting quantitative results. Our primary goal was to understand whether students who fail to present a rigorous proof do so because they do not use the proof framework that we teach them, or because they use it incorrectly.

This information is important in order to improve teaching: Should we be placing more emphasis on the use of the proof framework (i.e. encouraging the students to use it), or putting more effort into the details of the framework itself (making sure that the students use the framework correctly)? It is also interesting to check whether students who do not use the framework (i.e. students who use other proof techniques) are doing well. If we see that using the proof framework does not result in better proofs, compared to other techniques, this might indicate that the effort invested in sticking to a proof framework is redundant, or even harmful.

Our Course

Much like other Transition-to-Proof courses elsewhere (e.g. Selden, Selden and Bankhalti , 2018), our reasoning course is based on students practicing the act of proving. We cover basic classical mathematical concepts and proving methods: The role of definitions, quantifiers, proof by contradiction, dealing with cases, and induction. We rehearse these concepts and methods throughout the course. The course contains only elementary mathematical concepts: arithmetic, basic set operations, and basic linear algebra. It consists of 3 hours of weekly lectures, with an emphasis on independent practice and homework. In every lesson, we solve the homework given in the previous lesson. The students are encouraged to present their proofs in class, and get feedback from the teacher. This usually takes the first half of the lesson, and the second half is dedicated to a method of proving, explained and followed by many examples. The students are first year, first semester, computer science undergraduate students, and the course is designed to help them develop the skill of proving, which is needed in later courses such as Linear Algebra, Calculus, Data Structures and Algorithms. In each genre of proof, we introduce the students to a proof framework (see Section II for details) as a way of getting started writing proofs.

The Study

We teach in one of the largest colleges in the country, and our students have high grades in the acceptance exam and in the matriculation exam in mathematics (significantly higher than the average).

Around 300 students are enrolled in the "Mathematical Reasoning" course every year. When dealing with **basic set theory** we stress that:

1. When proving $A \subseteq B$ one should start with 'Let $x \in A$' and end with 'then $x \in B$'.
2. When proving $A = B$ the technique is to prove that $A \subseteq B$ and that $B \subseteq A$, using the framework of set inclusion.

We chose a non-covid cohort, where all lectures and the exams were frontal (i.e. not via zoom). The class took place between October 2019 and January 2020 and was divided into 5 groups of around 60 students. We analyzed the students' responses to a multi-part question on basic set theory given in the course's final exam. Our goal was to analyze the use of the proof framework in students' answers, specifically: did students use the proof framework that we taught them? If so - did they use it correctly or did they merely write something that looks like the framework but is actually totally wrong? If not - were they able to present a rigorous and correct proof without depending on the proof framework?

The exam question that we analyzed began with a "definitions" section requiring the students to present definitions that were taught during the semester. Specifically, we were interested in the correct definitions of inclusion and equality of sets. The "definitions" section was followed by two "proof" sections, in which the students were asked to prove or disprove certain claims (that were actually both true). We will later refer to the two proof sections as the "first" and "second" proofs. Here is the question (summarized, for brevity):

1. (10 pts) For A, B sets define $A\backslash B,\ A \cap B,\ A \cup B,\ A \subseteq B,\ A = B$
2. (12 pts) Prove or disprove according to the definitions:
   For sets A, B, C, $(A\backslash B)\backslash C = (A\backslash C)\backslash(B \cup C)$
3. (12 pts) Prove or disprove according to the definitions:
   For sets A, B, C, $C \subseteq A$ if and only if $(A \cap B) \cup C = A \cap (B \cup C)$

The "definitions" section is important as it requires the students to recall the correct definitions needed in the proof sections. In the analysis, we set aside the exams of those who did not write the definitions correctly, based on the assumption that we cannot expect students who are not able to present the definition of inclusion correctly to prove inclusion correctly (only a handful of students failed to correctly answer the definitions section, likely because this section is repeated every year).

The first proof is a set equality. This requires, by definition, a bi-directional inclusion. We expected the students to use the proof framework twice, once for each inclusion. The first starts with $x \in (A\backslash B)\backslash C$ and ends with the conclusion that $x \in (A\backslash C)\backslash(B \cup C)$, and the second does the same, but in the opposite direction. Since the question required the students to prove by definition, using set identities was not allowed. However, the students *could* also legitimately prove each of the directions by contradiction.

The second proof is an "if and only if" claim. One direction of the proof required students to prove an inclusion, and the other required them to prove a set equality. We expected the students to use the proof framework three times. Here, as well, proof by contradiction is considered "according to definitions", and was accepted as a correct answer, although it is not based on the proof framework. Note that there are altogether *five* inclusions to be proved: two of them for proving the first proof, and three for proving the second. The students saw at least four examples of very similar questions in class and solved five more similar questions in their homework assignments, which were solved in class after submission. In addition, every exam in the course from past years contained a question of that sort. When analyzing students' answers for the first and second proofs, we first focused on whether the student implemented the proof framework in the answer, to determine the extent of the phenomenon of students not applying the framework. We categorized the students' answers according to the following taxonomy:

1. **Correct use:** The student followed the proof framework correctly.
2. **Misuse:** Presenting incorrect use of the proof framework, (i.e., starting the proof of inclusion by inspecting an element which does not belong to the included set).
3. **No use - right:** Not following the framework and using another technique, however the solution was correct (e.g., proof by contradiction).
4. **No use - wrong:** The student did not follow the framework and used another technique, and the solution was incorrect.

If the student skipped the proof sections of the question or tried to disprove it (for both proof sections, the correct answer was proving), we marked this as well.

Findings

We analyzed 310 exams. We focused on the students who wrote the definitions correctly, meaning they knew that the definition of inclusion $A \subseteq B$ is that every $x \in A$ holds that $x \in B$, and that the definition of $A = B$ is that $A \subseteq B$ and $B \subseteq A$. 10 students failed to write the definitions correctly and were filtered out of the study. A perfect usage of the proof framework consisted of 5 uses, as the first proof required two inclusions to prove an equality and the second proof was an iff proof that required proving three inclusions. Therefore, within a single exam, it was possible for each student to sometimes use the framework right, sometimes use it wrong, and sometimes not use it at all. We started our analysis of the proof sections by inspecting the first proof, which asked the students to prove or disprove an equality (see Table I).

The results seemed promising. A large percentage of the students indeed used the framework we had taught them, and a very large percentage out of those received all the points for that part. It is interesting to note that only two students tried to prove by contradiction, which is allowed (it is considered as proving "according to the definitions"). However, their overall answer was not perfect and they did not receive all the points.

TABLE I
SUCCESS RATE FIRST PROOF

	Num	**percentage**
Failed on the definition part (out of)	10 (310)	3.2%
Disproved or wrote nothing	2 (310)	0.6%
Correct definition and attempted to prove	298 (310)	96.1%
Two correct uses	274 (298)	91.9%
Received all points (two correct uses)	224 (274)	81.7%

There were two conclusions we could draw right away:

1. Almost all students are **familiar** with the framework for set inclusion.
2. Almost all students know how to prove an equality between sets.

Having established that, and encouraged by the students' success in the first proof, we continued to check the second proof, where we found a drastic drop in framework use. Recall that the second proof is an iff claim with one side being an equality and the other side a single inclusion. Hence, we expected the proof to contain three inclusions. However, fewer than half of the students met this expectation by using the framework three times, as shown in Table II.

TABLE II
SUCCESS RATE SECOND PROOF

	Num	**Percentage**
Failed on the definition part (out of)	10 (310)	3.2%
Disproved or wrote nothing	26 (310)	8.3%
Correct definition and attempted to prove	274 (310)	88.3%
Three correct uses	125 (274)	45.6%
Received all points (three correct uses)	72 (125)	57.6%
At least one by-contradiction (no-use-right)	34 (274)	12.4%
Received all points	12 (34)	35.3%

The results were surprising. We expected that students who were able to use the proof framework correctly in the first proof could do that again in the second proof as well. It is important to note that the numbers clearly show that using the proof framework is potentially beneficial, with a success rate of 81.7% for part one for those who used the framework (compared to 0% for those who did not use it), and a success rate of 57.6% for those who used the framework in part two (compared to 35.3% for those who did not use the proof framework in the second proof). Having noticed the difference in the percentage of framework use between the two proofs, we proceeded to count the number of students who had shown familiarity with the framework (i.e., used it at least once), but who did not use it all the way through the exam. This finding is presented in the first row of Table III. Then we counted the number of students who used the framework perfectly on the first proof (two "correct use"), but made at least one misuse in the second proof, starting with an element belonging to the set in the hypothesis rather than an element in the included set. This finding appears in the second row of Table III. Both numbers (33.5% and 28.5% respectively) seem to be very high, and point to a difficulty in assimilating and implementing the proof framework, even when you have shown that you know how to use it.

TABLE III
CORRECT AND INCORRECT USES

	Num	percentage
At least one correct use and at least one mistake (out of)	100 (298)	33.5%
Two correct uses in first part and at least one misuse in second	78 (274)	28.5%

TABLE IV
MISTAKES IN BOTH PROOFS

	Num	percentage
First proof At least one misuse (out of)	3 (298)	1%
At least one no-use-wrong (out of)	20 (298)	6.7%
Second proof At least one misuse (out of)	85 (274)	31%
At least one no-use-wrong (out of)	39 (274)	14.2%

The large difference in the number of correct usages of the framework between the first proof and the second is especially striking, as can be seen in the differences between Table I and Table II. This overwhelming difference drove us to look deeper into the answers to the second proof. Two findings arose from that inspection:

1. We first ruled out the effect of the "iff". It turned out that **all** the students who attempted to prove the second part but did not follow the framework wrote that there are two directions to be proved. This means that they correctly addressed the iff part of this proof.
2. The second proof, as opposed to the first one, had a hypothesis. Many students began their answer with an element belonging to the set in the hypothesis rather than to an element in the included set. This was found to be the most common "misuse" of the framework.

In order to conjecture about the reasons for the difference between the two proof sections, we asked ourselves how many students made the mistake of using the framework wrong ("misuse") or not using it at all ("no-use-wrong"). The results are presented in Table IV. The type of "misuse" employed by many students in the second part, which included a hypothesis, reflects a problem described in Selden et al., 2010 as "*Students who focus too soon on the hypothesis.*" There can be several explanations for this common phenomenon. Our conjecture is that the students have not yet acquired the habit of relying on the proof framework, and it does not come naturally to them, so they were distracted by the existence of a hypothesis. Another possibility is that the effort of being strict while solving the first proof was followed by a "loosening" later in the exam. The evident conclusion, however, is that whatever the reasons, applying the framework is harder for many students than we may think.

Examples

In this section, we present snippets from actual exams, focusing on the approach taken.

A. Adam

Adam's answer demonstrates a **correct use** of the framework, 5 times as expected:

1. Prove / disprove according to the definitions: For sets A, B, C, $(A\backslash B)\backslash C = (A\backslash C)\backslash(B \cup C)$.
   We will show a bi-directional inclusion. Let $x \in (A\backslash B)\backslash C$ [...] then $x \in (A\backslash C)\backslash(B \cup C)$.
   Now let $x \in (A\backslash C)\backslash(B \cup C)$ [...] so $x \in (A\backslash B)\backslash C$.
2. Prove / disprove according to the definitions: For sets A, B, C, $C \subseteq A$ if and only if $(A \cap B) \cup C = A \cap (B \cup C)$
   We will prove the two claims.
   - First, assume that $(A \cap B) \cup C = A \cap (B \cup C)$ and we will prove that $C \subseteq A$. Let $x \in C$ [...] and also $x \in A$.

- Second, assume that $C \subseteq A$. To prove $(A \cap B) \cup C = A \cap (B \cup C)$ we will prove bi-directional inclusion [...]
  - First, let $x \in A \cap (B \cup C)$ [...] it holds that $x \in (A \cap B) \cup C$.
  - Second, assume that $x \in (A \cap B) \cup C)$ [...] so $x \in A \cap (B \cup C)$.

We proved bi-directional inclusion so by the definition of set equality it holds that $(A \cap B) \cup C = A \cap (B \cup C)$. We proved both iff directions so this completed the proof.

B. Liam

Liam's proof for the first part was perfect, but in the second part, trying to prove equality between sets, he *focuses too soon on the hypothesis*. This is an example of a **misuse**.

[...] Let us prove that if $C \subseteq A$ then $(A \cap B) \cup C = A \cap (B \cup C)$.
According to the definition of equality we will show a bi-directional inclusions.

1. We prove that $(A \cap B) \cup C \subseteq A \cap (B \cup C)$ -- Let $x \in C$ [...]
2. We prove that $A \cap (B \cup C) \subseteq (A \cap B) \cup C$ -- Let $x \in C$ [...]

C. Sean

Sean, like many others, knew that to prove inclusion he should start with an element that belongs to the included set (he did this perfectly in the first part). However, instead of starting with $x \in C$ he started with an x that belongs to one of the sets in the hypothesis:

[...] Assume $(A \cap B) \cup C = A \cap (B \cup C)$ and prove that $C \subseteq A$. Let $x \in (A \cap B) \cup C$ [...]

D. Rita

Rita's proof of the first part was also perfect. However, in the second part she omitted the framework altogether. Her answer is an example of **no use - wrong**:

Let A, B, C be sets. We want to prove that if $C \subseteq A$ then $(A \cap B) \cup C = A \cap (B \cup C)$. By the hypothesis that $C \subseteq A$ it holds that every element that belongs to C also belongs to A. Let us look at the set $(A \cap B) \cup C$ the set of elements that belong to A and also to B, or belong to C. Now let us look at the set $A \cap (B \cup C)$ This set contains the elements that belong to A and to B or the elements that belong to A and to C. Combining the two, and according to the hypothesis, it holds that $(A \cap B) \cup C = A \cap (B \cup C)$.

Conclusions

Our examination showed that most students are able to use a proof framework correctly and that using a proof framework is potentially beneficial (i.e., significantly raises the chances of presenting a correct answer). However, we also discovered that students who were able to present a correct use of the proof framework in one question suddenly failed to use the same framework in a subsequent question. Analyzing examples of answers, we saw that many students neglected the framework that they had been using in the previous question, while others continued to use it, but incorrectly (starting with an element that belongs to the wrong set). This can be attributed to several reasons. At first glance it may seem that the 'iff' added a measure of difficulty to the question, but our analysis suggests that this was not the actual obstacle. Other possible explanations may be that the heavier "load" of the second proof caused students to lose their way, or perhaps they tired as they progressed through the test, thereby dropping the framework and going back to their bad "natural" habits. Finally, when a question contains a hypothesis, it may distract the students, who focus on it too soon. Making recommendations on how to better make use of proof frameworks may be premature at this preliminary stage. However, we can say that employing proof frameworks can indeed be beneficial for students, but that doing so consistently is harder than it may first appear. It appears that additional practice of the framework's more complicated usage scenarios is required. It may also help to present students with actual mistakes to avoid. It is clear, however, that students are able to apply proof frameworks, and we should aim to help them do so in the more complicated proofs.

References

Hammack, R. (2013). *Book of Proof*.

Hanna, G. (2000). Proof, explanation and exploration: An overview. *Educational Studies in Mathematics*, *44*, 5-23.

Moore, R. C. (1994). Making the transition to formal proof. *Educational Studies in Mathematics*, *27*(3), 249-266.

Reid, D. A., & Knipping, C. (2011). *Proof in Mathematics Education: Research, Learning and Teaching*. Sense Publishers.

Scheinerman, E. A. (2013). *Mathematics: A Discrete Introduction*. Third Edition. Boston, MA; Brooks/Cole

Selden, J., & Selden, A. (1995). Unpacking the logic of mathematical statements. *Educational Studies in Mathematics*, *29*(2), 123-151.

Selden, J., Benkhalti, A., & Selden, A. (2014, September). An analysis of transition-to-proof course students' proof constructions with a view towards course redesign. In T. Fukawa-Connolly, G.Karakok, K. Keene, & M. Zandieh (Eds.), *Proceedings of the 17th Annual Conference on Research in Undergraduate Mathematics Education* (pp. 246-259). Denver, Colorado

Selden, A., McKee, K., & Selden, J. (2010). Affect, behavioural schemas and the proving process. *International Journal of Mathematical Education in Science and Technology*, *41*(2), 199-215.

Selden, A., Selden, J., & Benkhalti, A. (2018). Proof frameworks: A way to get started. *Primus*, *28*(1), 31-45.

Vinner, S. (1991). The role of definitions in the teaching and learning of mathematics. In *Advanced mathematical thinking* (pp. 65-81). Dordrecht: Springer Netherlands.