A SCHEMA TO ANALYSE STUDENTS' PROOF EVALUATIONS

Kirsten Pfeiffer

School of Mathematics, Statistics and Applied Mathematics

National University of Ireland, Galway

Abstract. In this exploratory study I investigate first year mathematics undergraduates' practice of proof evaluation of alternative mathematical proofs. This paper describes the theoretical background which I chose as a basis for developing a schema to describe and explore students' proof evaluation performances. This schema is illustrated on students' evaluations on one particular example proof. Further I demonstrate insights arising from this study about to what degree the nature and purposes of proofs are visible *to the participating students.*

Key words: proof, proof evaluation, proof validation, artefacts, purposes of proofs.

INTRODUCTION

Interested particularly in the students' transition from school to university, I explore first year students' behaviour and knowledge when validating and evaluating mathematical proofs. Students' proof validation performances have been discussed in the research literature, for example by Selden & Selden (2003) and Alcock & Weber (2005). With Selden and Selden, I call the readings and considerations to determine the correctness of mathematical proofs and the mental processes associated with them *validations of proof*. In a mathematical community the process of accepting a proof involves more than its validation. Validation, the determination of the correctness of an argument, is a significant part of the process of accepting a proof, followed by a more extensive and open-ended process that involves a search for understanding as well as correctness, a desire for clarity and an alertness to the possibility of adaptation or extension. Seeing learning as assessing and participating in the practices of a community, I suggest to widen the context from *proof validation* to the notion of *proof evaluation*. With *proof evaluation* I mean two things: determining whether a proof is correct and establishes the truth of a statement (validation) and also how **good** it is regarding a wider range of features such as clarity, context, sufficiency without excess, insight, convincingness or enhancement of understanding. That is, *proof evaluation* includes assessment of the significance and merits of a proposed proof.

In her doctoral thesis, Hemmi (2006) developed a theoretical framework to describe how students encounter proof when studying mathematics at university level in Sweden. Her theoretical framework combines a sociocultural perspective with Lave and Wenger's (1991) and Wenger's (1998) social practice theories and with theories about proof obtained from the mathematical education research. In my study I adopt parts of Hemmi's theoretical framework and its terminology and combine it with some new ideas for investigating and describing how students validate and evaluate mathematical arguments. A significant aim of my study is to develop and test a schema to describe and analyse students' proof evaluation skills and habits.

In this paper I present this schema and use it in the interpretation of some excerpts of interview transcripts. The first part of this paper introduces the theoretical framework. I describe Hemmi's view of proof as an artefact in a community of mathematical practice, then explain how I employed her ideas in my development of a schema to describe and explore how students evaluate mathematical proofs. The second part of this paper describes the experiment, in particular one of the tasks I used in interviews held in 2009 with eight first year students. The third part of this article describes the students' behaviour when evaluating one of the proofs proposed in the interviews, using the specialized schema mentioned above. Observations arising from this study provide opportunities for researchers to learn about the students' views of mathematical proofs. In the final part of this paper I outline some of those findings and discuss the value of the suggested framework.

THEORETICAL BACKGROUND

Students as newcomers in a community of mathematical practice.

Influenced by Vygotsky's theories, Lave and Wenger (1991) established their notion of *legitimate peripheral participation*, a situated learning theory that argues that knowledge is distributed throughout a *community of practice* and can only be understood with the 'interpretive support' provided by participation in the community of practice itself. I consider the mathematical community, as Hemmi does, as a *community of mathematical practice* and the students as its *newcomers*. A fundamental concept of sociocultural theory is that mental activity is organized through culturally constructed *artefacts*. Becoming knowledgeable or *learning* means increasing membership in the practice which includes the ability to use and understand its artefacts. They provide learners with opportunities to enter a community.

Motivated by Adler (1999) and Hemmi (2006) I argue that proofs can be seen as intellectual *artefacts* in mathematical practice. Adler considers *talk* as an artefact in mathematical learning and Hemmi extends this idea to mathematical *proof*. Recognizing that validating and evaluating proofs are crucial activities in a mathematical community, I investigate novice students' habits when performing these activities. Considering Hilpinen's (2004) philosophical approach towards artefacts, I describe how artefacts can be evaluated in general and specialize this to the practice of proof evaluation throughout the mathematical community and by *newcomers* such as first year students. Observation of novice students' behaviour when validating and evaluating proofs will give us insights into their existing knowledge about the artefact proof.

Proof - an artefact in the mathematical practice

An *artefact* can be described as an object that has been intentionally made or produced for a certain purpose. The philosopher Hilpinen (2004) describes how artefacts can be evaluated. He distinguishes between the *intended character* of an artefact, its *actual character*, and its *purpose*, and evaluates on the basis of the relationships among those features. In the following sections I apply this philosophical approach to artefacts of the type *proof* and use it to describe how the interview participants evaluate proofs.

Figure 1 below describes how a *proof* can be evaluated, applying Hilpinen's description of evaluations of artefacts in general. A proof can be evaluated in relating the three features of an artefact, its *intended character*, its *actual character*, and its *purposes*.

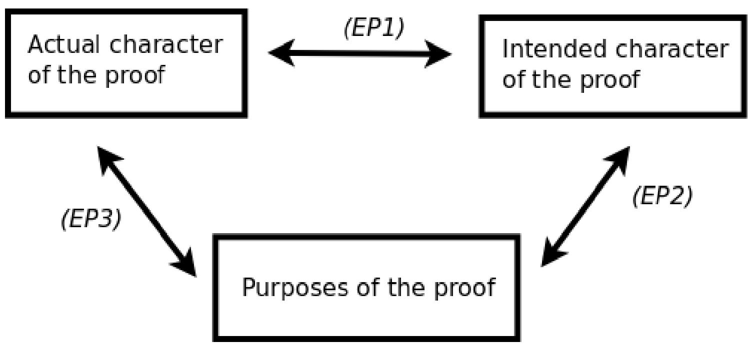

Fig1: Evaluation of the artefact proof

In this graphic the *actual character of the proof* means the actual realization of the author's intention, whereas the *intended character of the proof* designates this intention. *Purposes* (or functions) *of proofs* [1] have been widely discussed within the mathematical education literature in the last four decades, considering that a broader range of functions of proofs than that of establishing the truth of a statement should be recognised. De Villiers' (1999) suggested model for the functions of proof has been broadly accepted and applied within the mathematical education community. In his model functions of proofs include *verification* (concerned with the truth of a statement), *explanation* (providing insight into why it is true), *systematisatio*n (the organization of various results into a deductive system of axioms, major concepts and theorems), *discovery* (the discovery or invention of new results), *communication* (the transmission of mathematical knowledge) and *intellectual challenge (*the self-realization derived from constructing a proof). Expansions to this list of functions have been suggested. For example Hanna and Barbeau (2008) claim that the list “stopped short of stating that proof contains techniques and strategies useful for problem solving." Acknowledging that those purposes of proofs weigh differently, depending on preferences of authors and readers and also on the circumstances of the presentation of a proof, I consider in the context of proof evaluation particular *proofs*, not *proof* in general.

In **Figure 1** the relationships among the three features are labelled *(EP1)* to *(EP3)*, where '*EP'* symbolizes 'Evaluation of a Proof'. [2]

- *(EP1)* is concerned with how a proof is a successful realization of the author's intention, e.g. whether all steps of the proof are mathematically correct or whether the proof is clearly structured.
- *(EP2)* is concerned with how an intended proof, the author's idea of the proof, is suitable for its purposes. Is the idea appropriate to prove the mathematical statement?
- *(EP3)* is concerned with how the author was successful in proving the mathematical statement as claimed, establishing its truth, potentially convincing a mathematical community or regarding other purposes of proofs as suggested above.

In the interpretation of the transcripts of the conducted interviews I focus on the students' proof evaluating habits, in particular on whether and how they reflect on the relationships *(EP1)*, *(EP2)* and *(EP3)* among the actual and intended character and the purposes of a proof. **Figure 2** below demonstrates how the researcher might learn about the students' views of proofs through observations of their proof evaluation skills and habits.

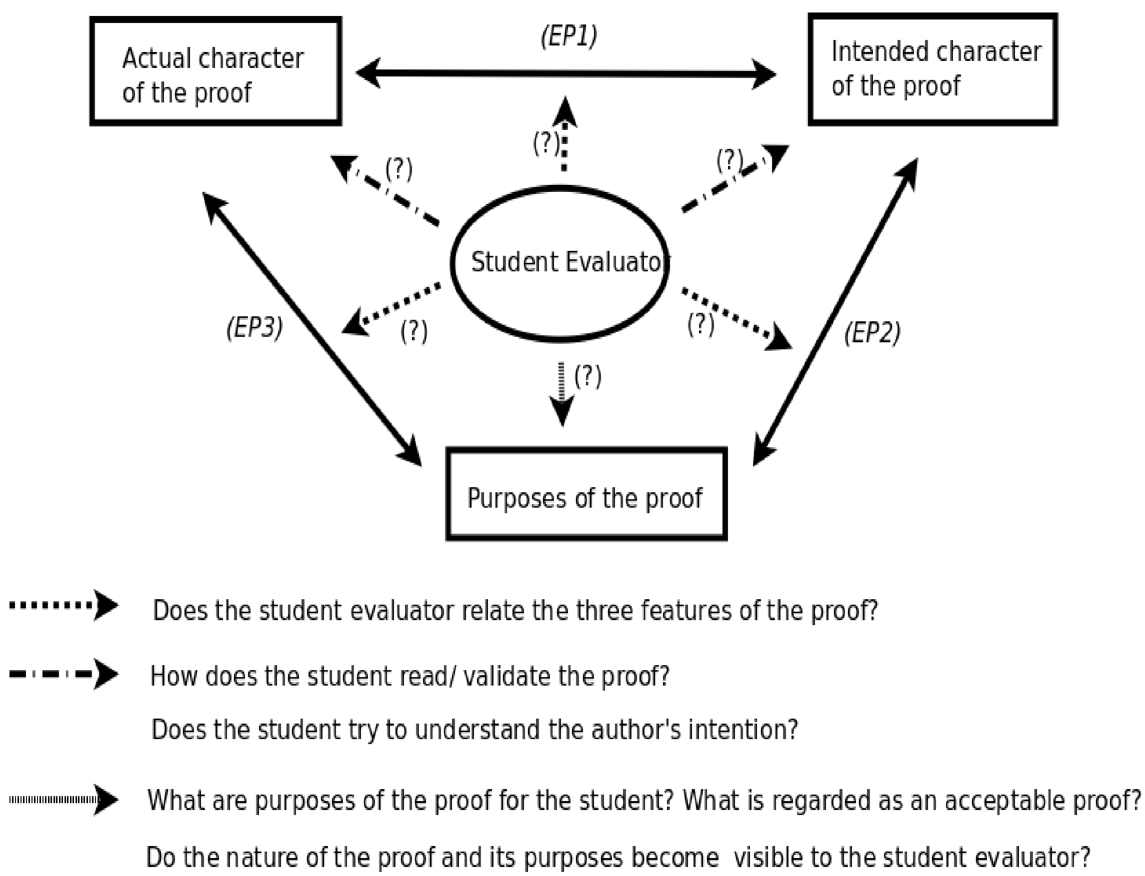

Fig2: Research questions: how does the student evaluate a mathematical proof?

THE EXPERIMENT

The study is based on a series of tests and interviews conducted with first year honours mathematics students at NUI Galway. In March 2009 interviews were held with eight students. Eighteen students who had attended a written exercise including an evaluation task in September 2008 as well, were invited to participate in a research project. They were chosen carefully in an effort to cover a wide spread of performances in the written experiment. All eight students who volunteered participated in the project. Each of the interviews took 30 to 45 minutes. Every interview was tape recorded and transcribed. The aim of the interviews was to get a deeper insight into students' opinions about valuable proofs, students' validation and evaluation processes and learning effects during the validation and evaluation processes. The students were presented with two mathematical statements and five or six proposed proofs of each statement and asked to evaluate and rank them. One proof of the first statement was purely visual, one consisted only of a fairly random assortment of examples, one was completely wrong but written in "algebraic" language, one was more general than required, another was written in text. [2] I will now present one of the proposed proofs and reflect on how an experienced reader might evaluate it. An interpretation of the students' evaluations and rankings of this proof during the interviews will demonstrate how the transcripts were used to learn about the students' evaluation habits and their knowledge about mathematical proofs. Finally I will outline the results of the analysis of the entire student evaluations of the six proposed proofs of *Statement I* [3].

> ***Statement I.*** Consider the following statement. The squares of all even numbers are even, and the squares of all odd numbers are odd.
>
> **Anna's answer:**
>
> Even numbers end in 0,2,4,6 or 8.
>
> $0^2 = 0$, $2^2 = 4$, $4^2 = 16$, $6^2 = 36$, $8^2 = 64$.
>
> When you square them the answer will end in 0, 4 or 6 and is therefore even.
>
> So it's true for even numbers.
>
> Odd numbers end in 1,3,5,7 or 9.
>
> $1^2 = 1$, $3^2 = 9$, $5^2 = 25$, $7^2 = 49$, $9^2 = 81$.
>
> Squaring them leaves numbers ending with 1,5 or 9, which are also odd.
>
> So it's true for odd numbers.

An experienced evaluator would probably identify that Anna's argument centres on her assertion that the last digit of the square of an integer is determined by the last digit of that integer itself. This assertion is correct. It certainly could be argued that the assertion needs some justification. If the evaluator is prepared to accept Anna's assertion, the actual character of this proof does coincide with the intention and therefore the argument does satisfy condition *(EP1)*. However, Anna's argument does not provide an essential explanation of WHY squaring an integer preserves parity (i.e. oddness or evenness). There is no reason to construct a modulo 10 argument (based on the last digit - the remainder on division by 10) for a problem in modulo 2 arithmetic (the problem is about remainders on division by 2). A reader may well complain that by using 10 cases where two would suffice, this proof misses the key point. The intended character of this proof, involving 10 different cases, is not a good fit to the purpose of explaining why squaring preserves parity. For that reason an experienced evaluator might regard Anna's proof not satisfactory concerning *(EP2)* and *(EP3)*.

Example: An Interpretation of some transcript excerpts.

The coding table below (**Table 2**) provides an overview about the participants' evaluations of Anna's proof; the codes in the bottom line refer to how the students placed Anna's proof in the ranking of all six proposed proofs. The heading line refers to the codes for the students [4] (students *T4* and T5 were interviewed together). **Table 1** describes the coding scheme.

Satisf	The student regards the answer as satisfying.
NotSatisf	The student regards the answer as **not** satisfying.
Proof?	The student is not sure whether the proposed approach is a sufficient proof of the statement or not.
NoProof	The student does not regard the answer as proof of the statement.
NotGeneral	The student criticises that the proposed approach is not applicable in general.

Table 1: Coding Scheme.

K19	***Ja4***	***T4/T5***	***Ja1***	***R9***	***T2***	***C3***
Proof?	NotSatisf NotGeneral	Satisf NoProof NotGeneral	NotSatisf NotGeneral	NotSatisf NotGeneral	Satisf Proof? NotGeneral	Satisf NoProof

first	fourth	third	fourth (jointly)	fourth (jointly)	fifth	second

Table 2: Coding Table: Students' evaluation of Anna's proof.

The coding table refers to three groups of comments.

- **Satisf/NoProof**: The student is happy with Anna's answer and considers approvingly that Anna is using examples. The student considers that Anna's answer is not a proof of the statement. *"It's not a proof, but it works"* or *"It's a good answer. (...) There is no kind of proof (...)"* are responses assigned to this group.
- **Satisf/Proof?**: The student likes Anna's answer because Anna *"gives examples"* or the answer *"is different"*. It is not clear from the interview conversation whether the student regards Anna's answer as a sufficient proof of the statement.
- **NotSatisf/NotGeneral**: The student does not accept the answer as a proof of the statement because it "is not general".

With the research questions listed in the above diagram (**Fig2**) in mind I suggest an interpretation of the students' evaluations of Anna's proof, occasionally using short transcript exemplars to advance certain claims.

How do the student evaluators relate the three features of Anna's proof?

Two of the five students whose opinions belong to (Satisf/NoProof) or (Satisf/Proof?) do not seem to focus on relations *(EP2)* or *(EP3)* as there is no evidence to suggest that they are considering the purposes of mathematical proof. Even though the other three of those students (Satisf/NoProof) express the opinion that Anna's answer is not a valid proof of the statement and in particular that the argument is not applicable in general they acknowledge the unusual approach and the internal correctness and rank Anna's proof relatively highly (second or third out of six). Their responses to Anna's answer suggest that *(EP1)* may be more important to them than the relations *(EP2)* or *(EP3)*: *"It's not a proof, but it works."*(*T4/T5*) or *"There is no kind of proof, it's just ---. But it does make sense"*(*C3*). Internal correctness seems to be considered more important by this group of students than the purpose of establishing the general truth of the statement.

The three (NotSatisf/NotGeneral)-students do relate the actual proof not only with the author's intention but with its purposes and therefore do consider relations *(EP2)* and *(EP3)* as well as *(EP1)*. They consider at least one purpose of proof, namely its general applicability, criticize the poor relation between the actual or intended proof and its purposes, and therefore regard Anna's proof as unsatisfactory, which is indicated by their ranking of this proof. These three students seem to regard relations (*EP2) and/*or *(EP3)* as being at least as important as *(EP1)*.

Do the student evaluations of Anna's proof indicate what the students consider purposes of proofs?

Five students criticize a lack of general applicability in Anna's proof, which indicates that they consider this as one purpose of mathematical proof. One of the (NotSatisf/NotGeneral)-students (*Ja1*) finds the level of justification insufficient *"She doesn't prove that 'When you square, the answer will end in 0,4,6...' If she'd proved that, it would be ok.*" Certainly for this student justification of intermediate steps is a necessary ingredient of mathematical proof. This student seems to see Anna's answer as an attempt at a general argument about the last digit, that could be improved to a proof. This is similar to how an experienced evaluator is likely to see it, namely as more than a collection of examples. In an experienced evaluator's view the examples that are included in Anna's proof are not intended as examples but as items in an exhaustive list that covers all cases. The other students who complain that Anna's answer is not general and consists of *"just examples"* interpret this in a different way: student *R9* for example seems to see it as basically the same as another proposed proof which consists of a collection of examples, just *"up to ten numbers"*. Student *Ja4* like student *R9* considers Anna's answer as just a selection of examples: *"She took the numbers from 1 to 9, but what about all the other numbers? (…) Nice example, but that's about it."*

Do the student evaluations of Anna's proof indicate a learning process? Do nature or purposes of mathematical proof become *visible* to the student evaluator?

Student *T2*'s reaction indicates that a learning process is initiated by the task. Her first reaction (*"Very cool"*,*"different")* indicates that she admires the unusual approach (*"I could never think of anything like that, (...) the way she writes it down (...)"*). After careful prompting by the interviewer a reflection process is initiated and the student is getting more and more unsure, until at some point she almost decides that this is not a proof, but is never really sure about this. Student *T2'*s comments do not show clearly what she considers as a valid or valuable proof, but she certainly thinks about it.

The second proposed proof (*Benny's answer*) consists of a collection of ten examples. Interestingly seven of the eight students commented in the interviews on how they compare Benny's answer to Anna's, even though they weren't asked to do so. Four students regard the answers as very similar. Two students approve the fact that Benny includes examples of negative integers in his answer. Five students, all agreeing that neither answer proves the statement sufficiently, mention that Anna's answer is more like a proof than Benny's. They identify two aspects of proof more present in Anna's than in Benny's answer:

- the description of general patterns: *"She has this --- with the endings"* (*T4/T5*), *"In* [Anna's answer] *there is more thinking in it. She saw this fact, if you square an even number, that there is a 0,2,4,6,8 at the end of each one."* (*Ja1*)
- Anna's answer includes some attempts to explain **why** the statement is true. *"She says why the squares are odd, because they end in that. He* [Benny] *just presumes that they are odd numbers."* (*R9*)

Considering Benny's answer in comparison to Anna's, some of the students who have interpreted Anna's proof as list of randomly chosen examples when discussing Anna's answer now identify some potential in Anna's answer to provide a general proof: student *R9* states that *"Anna's is more of a proof* [than Benny's]. *She says why the squares are odd, because they end in that"*. Likewise student *Ja4* regards Benny's answer as *"more example than proof than Anna's was"*. These changes in some of the students' opinions about Anna's proof indicate a learning effect about proofs through the comparing process. It seems that some purposes of mathematical proof became *visible* to these students.

SUMMARY AND FURTHER OBSERVATIONS

Consideration of the interview data with the relationships *(EP1)*, *(EP2)* and *(EP3)* among the actual and intended character and the purposes of a proof in mind led to the following observations. The students' evaluations of all six proposed proofs of *Statement I* indicate which purposes of proofs they consider relevant. In student *K19*'s evaluation the most relevant consideration is how the statement's plausibility is verified and demonstrated by the proposed answer. Consequently she favors answers consisting of examples. Generality of a proof is an important evaluation criterion to most of the students. Five students mention at some point during the interviews that they appreciate considerations and explanations about *why* the statement is true in certain proposed answers. Some of the students take into account whether the proposed proof emphasizes some mathematical contents or general patterns. Some students consider sufficiency without excess in their proof evaluations. Some students appreciate a didactical value in a proof, which includes how well a reader's interest is stimulated or how well both statement and proof are being explained to the reader. The proof idea or method does not seem to play a significant role in the students' evaluations of proofs of the first statement, which is indicated by three observed phenomena. Firstly, a proposed visual approach is liked least considering the ranking of all eight students together. The intrinsic idea behind this approach seems unimportant to the students. Secondly the fact that one of the proposed arguments proves a more general fact than the facts of *Statement I*, is rarely being recognized and not appreciated by the students. The third surprising fact indicating poor appreciation of proof ideas or methods is the relatively high ranking score of an irredeemably wrong approach. While some of the students noticed errors in this proof, none questioned the basic strategy.

Interpretations of oral and written proof evaluation exercises so far suggest that the developed conceptual framework and schema to interpret student-evaluations are beneficial to gain some understanding about students' knowledge and skills about proofs and proving. The schema appears to be in particular useful to identify students' criteria to accept or value a mathematical proof and also to what extent and how first year students consider purposes of mathematical proofs. P*roof evaluation* as an important activity in mathematical practice might carry some potential to provide students with opportunities to enter the practice. The suggested schema to interpret student-evaluations is appropriate to determine whether that is the case, i.e. to what degree *proof evaluation* performances support learning effects. However, the suggested method seems to be less effectual regarding observations about students' proof reading habits. I did not gain a lot of noteworthy informations about how the students try to understand a proposed proof.

Overall, results of this study indicate that, considering importance and challenges in the teaching and learning of mathematical proof, exploration and practice of incoming students' proof evaluation skills and habits are worthy of further attention.

NOTES

1. By *function* of proof I mean with DeVilliers(1999) “meaning, purpose and usefulness” of proof.

2. Hilpinen (2004) introduced the notation *(E1) – (E3)* for the relationships of the three features of artefacts, relating to three aspects of evaluations, where 'E' symbolizes 'Evaluation'.

3. Detailed descriptions of the tasks, interview questions and transcripts can be found in my forthcoming PhD Thesis.

4. The code names consisting of one or two letters and a number were assigned to students according to their tutorial groups and have no particular significance.

REFERENCES

Adler, J. (1999). The dilemma of transparency: seeing and seeing through through talk in the mathematics classroom. *Journal for Research in Mathematics Education* Vol. 30(1): 47-64.

Alcock, L. and Weber, K. (2005). Proof validation in real analysis: Inferring and checking warrants. *Journal of Mathematical Behaviour*, Vol. 24: 125-134.

De Villiers, M. (1999). The role and function of proof in mathematics with sketchpad. *Rethinking Proof with Sketchpad*, Key Curriculum Press.

Hanna, G. and Barbeau, E. (2008). Proofs as bearers of mathematical knowledge. *ZDM Mathematics Education,* Vol. 40: 345-353.

Hemmi, K. (2006). *Approaching proof in a community of mathematical practice. Doctoral Dissertation.* Department of Mathematics, Stockholm University.

Hilpinen, R. (2004). Artifact. *Stanford Encyclopedia of Philosophy*.

Lave, J. & Wenger, E. (1991). *Situated Learning: Legitimate peripheral participation.* Cambridge University Press, Cambridge.

Selden, A. and Selden, J. (2003). Validations of proofs written as texts: Can undergraduates tell whether an argument proves a theorem? *Journal for Research in Mathematics Education*, Vol. 34 (1), 4-36.