Title	Features and purposes of mathematical proofs in the view of novice students: observations from proof validation and evaluation performances
Author(s)	Pfeiffer, Kirsten
Publication Date	2010-12-22
Item record	http://hdl.handle.net/10379/1862

Features and purposes of mathematical proofs in the view of novice students: observations from proof validation and evaluation performances

by
Kirsten Pfeiffer

School of Mathematics, Statistics and Applied Mathematics
National University of Ireland, Galway

2011

A dissertation submitted in fulfillment of the requirements for the degree of Doctor of Philosophy (Mathematics)

Abstract

This thesis describes a comprehensive exploratory study of the approaches taken by novice students to the validation and evaluation of mathematical proofs. A theoretical framework based on sociocultural learning theories was considered suitable as a basis to develop a new terminology and schema for observations and interpretations of proof validations and evaluations. In this theoretical framework learning is seen as accessing and participating in the practice of an expert community. The theoretical considerations of this thesis build on Hemmi's conception of proofs as artifacts in the community of practice. Philosophical theories about the evaluation of artifacts are specialized to the case of mathematical proof. A result of these considerations is a schema that extends Hemmi's model of proofs as artifacts, and provides both a theoretical basis and an analytic tool for consideration of the practice of proof evaluation and for interpretation of specific instances of proof evaluation. The study is based on a series of tests and interviews with first year honours mathematics students at the National University of Ireland, Galway. The students were asked to evaluate and criticize numerous proposed (correct and incorrect) proofs of mathematical statements. The participants' written comments and the interview discussions on different and partly incorrect proofs give insights into their criteria for valuing a mathematical proof, their habits when performing proof validations and evaluations and their knowledge about features and purposes of mathematical proofs. This thesis describes the theoretical framework as well as the design, observations and findings of the written and oral exercises. It also includes a discussion about advantages and shortcomings of the schema that has been developed and used for the interpretation of evaluations of mathematical proofs and a discussion about possibilities for further research.

Acknowledgments

I would like to extend a sincere word of thanks to my supervisor Rachel Quinlan for her support and advice and the many helpful and inspiring conversations over the past three years. Sincere thanks also to Kirsti Hemmi for her valuable advice in the development of my research project.

I also wish to thank the lecturers, postgraduates and students at the School of Mathematics, Statistics and Applied Mathematics at NUI Galway for sharing their views and experiences with me.

Many thanks go to Götz Pfeiffer and Liam Naughton for proofreading this thesis.

Finally, I thank Götz and our children Ronja, Kieran and Robin for their support during the last three years.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This research project is about novice mathematics undergraduates' practice of proof validation and evaluation. The study of proofs is a major obstacle in the transition from school to university mathematics. Given the importance of argumentation and proof in the spectrum of mathematical activities, the incoming students' understanding, appreciation and knowledge of the nature and role of proof must be considered. Proof is a difficult mathematical concept for students. Research shows that most university students do not know what constitutes a proof and experience difficulties not only in constructing proofs but also in determining whether a proof is valid. This thesis describes an exploratory study of first year mathematics undergraduates' criteria when validating and evaluating mathematical proofs. A theoretical framework based on sociocultural learning theories was considered suitable as a basis to develop a new terminology and schema suitable for observations and interpretations of proof validations and evaluations. The theoretical framework and schema to analyse proof evaluations are introduced in the theoretical part of this thesis. The study is based on a series of tests and interviews with first year honours mathematics students at the National University of Ireland, Galway. The students were asked to evaluate and criticize numerous proposed (correct and incorrect) proofs of mathematical statements. The first year students' written comments and the interview discussions on different and partly incorrect proofs give insights into the students' criteria for valuing a mathematical proof, their habits when performing proof validations and evaluations and their knowledge about features and purposes of mathematical proofs. The design, observations and findings of the written and oral exercises are described in the practical part of this thesis. The concluding part discusses some advantages and shortcomings of the schema that has been developed and used for the interpretation of evaluations of mathematical proofs. This part also suggests inclusion of proof evaluation activities in small groups into learning activities for novice students and discusses potential for further research.

Section 1.1 of this introduction considers the significance of proof as well as proof validation and evaluation in the learning of mathematics.

Section 1.2 explains the concept of *proof validation*. Various aspects of this activity are considered, including how and why mathematicians validate mathematical arguments. This section further includes an explanation why the research focus shifted from *proof validation* to *proof evaluation* while observing the students' behaviour in the interviews.

Section 1.3 describes how the research project was developed.

Section 1.4 gives an outline of the thesis.

1.1 Background of this study: reflections on the significance of *proof*, *proof validation* and *evaluation* in the learning of mathematics

Properly formulating a precise definition for the notion of a mathematical proof may be an impossible task.[1] Nevertheless, its essential role in mathematics has often been acknowledged. For example Rav (1999, p. 6) emphasizes the fundamental importance of proofs in mathematics: "Proofs, I maintain, are the heart of mathematics." According to Fischbein (1982, p. 17) the proof of a theorem is for a mathematician "the absolute guarantee of the universal validity of the theorem. He believes in that validity." Dreyfus (1990, p. 126) asserts that "proving is one of the central characteristics of mathematical behavior and probably the one that most clearly distinguishes mathematical behavior from behavior in other disciplines." In mathematical practice the role of proof exceeds demonstrating the truth of a theorem and the reason why a theorem is true. Rav (1999) states that proofs do much more than verify mathematical claims, that they are actually bearers of mathematical knowledge and also necessary to the enhancement of that knowledge. Rav argues that the very act of devising a proof contributes to the development of mathematics, and sees proofs as the primary focus of mathematical interest. In his view proofs can not only yield new mathematical insights, giving them a value far beyond establishing the truth of new propositions, but can also convey new mathematical strategies and new methods for solving problems. Furthermore, proofs are important means in communities of mathematical practice. Bass (2009) emphasizes this in a special lecture on the activity of proving and accepting new proofs as certification of new knowledge and as activity of the community. "Mathematics has evolved a unique method – deductive proof – dating back to Greek antiquity." In his talk, Bass considers "proof, as a theoretical concept" and "proving, as a fundamental practice of the disciplinary community". Critical and sceptical reading of proofs are important activities in mathematical practice. Hanna (1991, p. 59) quotes the view of the russian logician Manin who had emphasized that the acceptance of a proof depends significantly on a social process: "A proof becomes a proof after the social act of 'accepting a

[1]Davis and Hersh (1981) construct a dispute between a student of philosophy and the Ideal Mathematician (I.M). The student asks the I.M.: "What is a mathematical proof?". The I.M. responds with examples, but the student wants a general definition. After further attempts, each challenged by the student, the Ideal Mathematician confides that "a proof is just a convincing argument, as judged by competent judges". 12 years later Hersh (1993) remarks that "no philosopher or mathematician has yet taken up the Ideal Mathematicians' challenge."

proof"' (Manin 1997, as cited in (Hanna 1991)). Manin concludes that a new proof needs to be accepted and approved by other mathematicians. Section 3.2 will continue the discussion of the role of proofs and of the process of accepting proofs for mathematicians.

From a socio-cultural viewpoint it is essential for novices in a practice to become familiar with activities in this practice and to get opportunities to participate in it. In Section 2.1 I explain why I regard a socio-cultural learning theory as particularly useful for investigations in students' knowledge during the transition from second level to university education and describe the chosen learning theory.

Alibert and Thomas (1991, p. 215) describe that generally undergraduates are presented with mathematics as a "finished theory" and proofs as developed along traditional 'linear' deductive lines. This is in opposition to how conjectures and proofs develop in professional mathematicians' practice. A mathematician considers and rejects hypotheses and proposes new findings to the mathematical community, where they are reflected by other mathematicians. A proof in this context is "a means of convincing oneself whilst trying to convince others", which in most cases does not happen in linear order. Therefore, Alibert and Thomas claim, the "conflict between the practice of mathematicians on the one hand, and their teaching methods on the other, creates problems for students. They exhibit a lack of concern for meaning, a lack of appreciation of proof as a functional tool and an inadequate epistemology." Given the importance of proof and proof validation in the spectrum of mathematical activities, in addition to the impact of a sense of belonging and familiarity with mathematical practice on students' learning processes, the incoming students' understanding, appreciation and knowledge of the nature and role of proof as well as their abilities to validate and evaluate proofs must be considered. Proof is a difficult mathematical concept for students. Research shows that most university students have difficulties not only in construction of proofs (Schoenfeld 1985, Senk 1985, Martin & Harel 1989, Hart 1994, Moore 1994, Bills & Tall 1998, Harel & Sowder 1998, Weber 2001), but also in knowing what constitutes a proof (Harel & Sowder 1998, Recio & Godino 2001). Furthermore research demonstrates that students have difficulties to determining whether a proof is valid (Selden & Selden 2003).

There is little or no existing research literature specifically addressing the learning of mathematical proof and proving in Ireland. Discussions with mathematics lecturers of various Irish universities indicate that Irish university students have similar difficulties with mathematical proof to those described in the studies mentioned above. Worrying tendencies in Irish second level students' abilities to construct proofs are indicated by the Chief Examiner's Report on the Higher Level Leaving Certificate Examination in Mathematics (2005, p. 72) which describes a "noticeable decline in the capacity of candidates to engage with problems that are not of a routine and well-rehearsed type. [...] Whereas procedural competence continues to be adequate, any question that requires the candidates to display a good understanding of the concepts underlying these procedures causes unwarranted levels of difficulty." More specifically the report notes the students' difficulties with proof by induction: "most [students] are completely unable to handle the substantive step" (NCCA 2005, p. 74).

Selden and Selden (1995, p. 140) see the "lack of validation skills as linked to beginning university students' well-documented inadequate conceptions of proof". Alcock and Weber (2004, p. 1) highlight the importance of the learning of proof validation skills: "The ability to validate proofs is a critical ability for students of mathematics to possess. In order for a student to be convinced of a theorem by reading a proof of a theorem, that student would need reliable methods for determining whether a proof is valid." Powers et al. (2010, p. 501) emphasize the importance of proof validating skills for everyone who is going to be involved in the teaching of mathematics: the "ability to validate proofs is a much-needed skill for future teachers and those who may be involved in instruction or training as a graduate student or supervisor."

1.2 Clarification of the terms *proof validation* and *proof evaluation*

The attempts of students to validate mathematical proofs first stimulated my interest when I was analysing their responses to a written exercise that was held at NUI Galway in May 2008, at the end of their first year at University. My research has focussed on that topic since. Therefore the design of further research instruments concentrated on proof validation. In the following chapters I consider how experienced mathematicians and first year students validate and evaluate mathematical proofs. Before that, I state Selden and Selden's definition of *proof validation* and distinguish it from other types of reading and from construction of proofs. I will then explain how and why my focus later expanded from *proof validation* to the wider notion of *proof evaluation*.

With Selden and Selden (1995), I call the readings and considerations to determine the correctness of mathematical proofs and the mental processes associated with them "validations of proofs". Validation, in comparison to the reading of non-mathematical texts, requires the reader to put some additional effort into understanding of the reasoning. Validation usually takes more time, the validator might consider the whole proof or parts of it several times and might be more inclined to write a few notes checking deductions, verifying justifications, etc. According to Selden and Selden (2003) the mental process when validating proofs can include for example asking and answering questions, constructing subproofs or recalling other theorems and definitions. It is well documented, for example in the articles cited above, that construction of proof is a major obstacle for students. Selden and Selden (2003) describe how the ability to validate proofs relates to the ability to construct them. On the one hand proof construction and proof validation are different. Proof construction requires 'the right idea' at the 'the right time'. The validation process can usually be managed in a linear order, unlike construction of proof. On the other hand proof construction and proof validation entail each other as one considers during the process of proof construction how that proof would be validated, and as validation of a proof is likely to require the construction of subproofs. Figure 1.1 summarizes this relationship.

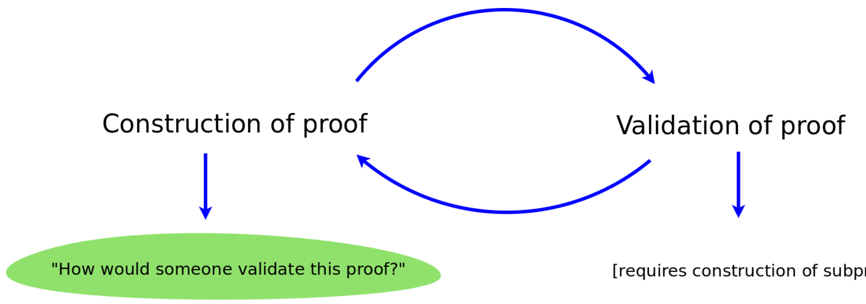


Figure 1.1: Relationship between construction and validation of proof

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. Mathematicians are concerned with qualitative features of proofs as well as with the question of their validity, and the individual and collective mental processes through which proofs are studied and refined in the mathematical community entail much more than validation in the narrow sense of the definition suggested above. It is common for a referee to propose improvements to a logically correct proof presented in a submitted paper, or to comment on connections to other literature. Also it is not unusual to find papers with titles like "A new proof of ..." offering an alternative proof of a known result that is preferable to existing ones in some way. Seeing learning as accessing 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.

Considering the broader range on the nature of proof evaluation, I extend the diagram about the relationship between proof construction and proof validation (Figure 1.1) to highlight the range of learning effects through the processes of both proof validation and evaluation, see Figure 1.2. The diagram indicates why in my view observation and encouragement of students' proof validations and evaluations are worth deeper considerations.

Experiences with oral evaluation exercises indicate that not only the tasks to construct but also to validate and evaluate proofs can be considerably challenging to incoming students. Moreover, it appeared that the participating students were not experienced in discussing mathematical proofs. The majority of them read the proofs more or less carefully and articulated relative brief comments and judgements. However, the task of comparing and ranking various proposed proofs prompted more

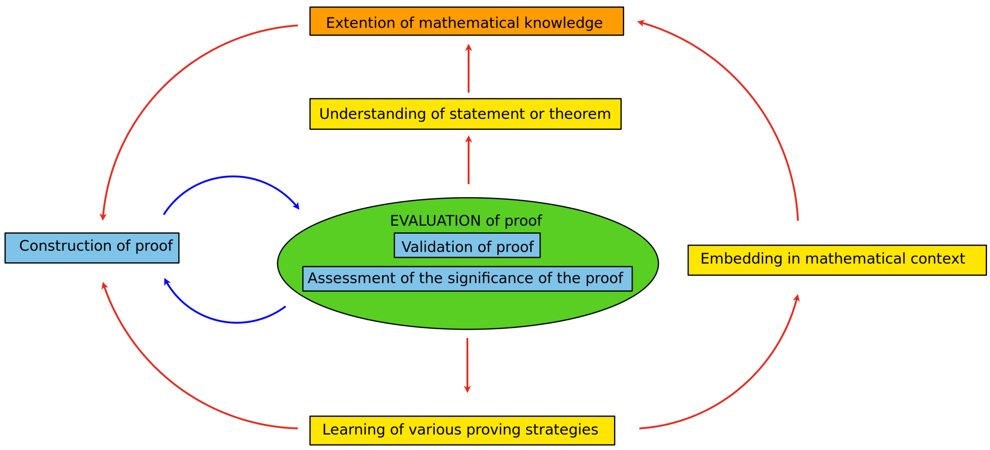


Figure 1.2: Learning effect through evaluation of proofs

discussion. Comparison of proof appeared to be a good way for newcomers to appreciate or experience what is involved in both proof validation and proof evaluation. Therefore most of the evaluation tasks that are used in this study do involve comparison. For the oral exercises which are based on the participants' willingness and ability to talk about their thoughts about the proofs, observation of students' proof evaluations initiated by comparison of a number of proofs was chosen. The students appear to engage more when they are asked to compare various proofs rather than when they are asked to evaluate just one proposed proof. Comparison of the proofs involves their evaluations and encourages discussions about evaluation criteria. The students' considerations of which of the proposed proofs they prefered and of their choices of criteria when evaluating supplied some insights into their knowledge and views about mathematical proofs.

The term *evaluation* corresponds well with the terminology in the theoretical framework. I will describe in Section 3.1 that I regard proof as an *artifact* in mathematical practice. Evaluation is significant in the discussion of artifacts. Every artifact can be evaluated. The notion of evaluation might be used when defining the term *artifact*. It is used in context with a wide range of artifacts including for example arts. In many cases an artifact is expected to serve many different purposes, so is proof in a mathematical community. *Validation* is used in context with certain purposes of the considered artifact, namely its instrumental quality. *Evaluation* is the relation of the artifact with its whole spectrum of purposes.

Moreover, in the light of the comments above, students' engagement with evaluation tasks may be worthy of consideration as a teaching method which leads to practice of proof validation and proof construction.

To summarize, I regard *proof validation* as the activity to determine whether a proof is successful in establishing the truth of a statement, whereas I see *proof evaluation* as an extension of *proof validation* through assessment of the significance and merit of a proof. Considering the crucial role of proof as well as proof validation and evaluation in mathematical practice, in combination with the influence of students' sense of familiarity and belonging to the practice on their learning of mathematics

- especially during the phase of transition from school to university, I investigate incoming students' abilities and habits in validating and evaluating proofs. These observations will also provide significant insights into the participating first year students' knowledge about features and purposes of mathematical proofs.

1.3 Chronicle of this study

This explanatory study has been developed as my PhD research project in the academic years 2007/08 until 2009/10.

A series of weekly workshops, held with first year mathematics students at the National University of Ireland, Galway, played a constituent role in the development of this research project. The theme of the workshops is the practice of basic mathematical skills and talking and teamwork about/on unusual mathematical problems. In these workshops a group of lecturers and myself work intensively in small groups with the students. During the first year of this project the students' attitudes toward mathematical proofs caught my attention.

In the workshops, various written and oral exercises were completed. These tasks were designed to focus on certain mathematical skills such as generalizing, formulating or justifying conjectures, and to investigate learning progress during the first year in University. Experiences with the early exercises influenced the design and focus of further research. In particular the students' responses to an evaluation task which was included in a written test at the end of the academic year 2007/08 caught my attention. Inspired by research papers of Selden and Selden (2003) as well as Alcock and Weber (2005), and the fact that there was otherwise relatively little research literature on students' proof validation performances, I chose to focus on this topic in my further research.

The students' responses to an evaluation task in one of the workshop tests (the so-called *End-of-the-Year-Test08*) were analysed and gave some understanding about the students' criteria when validating and evaluating mathematical arguments. The analysis of the written evaluations left some uncertainties and open questions.

To get broader as well as deeper insights, another test (*Diagnostic-Test08*) and semi-structured interviews were designed with the findings and open questions from the responses to the first written evaluation task in mind. One aim of the interviews was to achieve a clearer understanding of the students' views on mathematical proofs. Other aims were to investigate their proof validation skills and habits and whether practice of proof validation has the potential to advance learning processes. The *Diagnostic-Test08* and the semi-structured interviews were conducted in the academic year 2008/2009.

Considerations of the nature of *proof validation* as well as the fact that the processes observed in the interviews were as much evaluations (including decisions about what the students value in a proof) as validations of proofs, motivated a shift to the wider notion of *proof evaluation*.

Accompanying the acquiring of data, I studied pedagogical theories in order to find an appropriate frame for my study. Hemmi's (2006) theoretical frame seemed particularly suitable to use for describing learning processes in the context of mathematics especially when validating and evaluating proofs. In Lave and Wenger's (1991) learning theory learning means increasing participation in a community of practice and gaining familiarity with its artifacts. Hemmi considers communities of mathematical practice and proofs as artifacts in this practice.

I chose to investigate whether the practices of *proof validation* and *proof evaluation* can be considered to provide opportunities for students to access mathematical practice and to learn to understand and use the artifact proof.

I applied a philosophical approach to describe evaluations of artifacts as relating actual and intended character of the artifact and its purpose(s) to artifacts of the type *proof* in mathematical practice. Based on this description of proof evaluations I developed a schema and associated terminology to analyse students' proof evaluation performances.

This schema facilitates the formulation of some research questions for this study. The research questions are concerned with the students' proof evaluation and validation habits, with their views of what they accept as a mathematical proof, what the students regard as purposes of proof as well as whether learning effects are advanced through practice of proof evaluations.

The conducted interviews were transcribed and interpreted in the light of the suggested schema and the proposed research questions. The observations and findings from previously analysed written student evaluations were considered again with the suggested schema in mind. Due to time limitation for this study some of the acquired data have been postponed for possible later analysis.

Both experiments under consideration of the schema provide significant insights about the students' evaluation habits and their views on features and purposes of proofs.

1.4 Outline of the thesis

The thesis consists of this introduction and three parts, a theoretical, a practical and a concluding part.

The **Theoretical Part** outlines the theoretical framework and the schema that was developed to analyse proof evaluations.

Chapter 2 discusses why a sociocultural learning theory is considered particularly suitable for investigations of students' learning progress during the transition from school to university. Lave and Wenger's sociocultural learning theory is introduced to provide a basis to discuss and describe learning about nature and purposes of proofs in mathematics.

Chapter 3 describes how purposes of proofs are discussed in the fields of mathematics, mathematical education and mathematical philosophy.

Chapter 4 explains a schema and terminology to analyse students' proof evaluation performances, which has been developed as part of this study.

Chapter 5 summarizes the Theoretical Part of this thesis.

The **Practical Part** describes the design, observations and findings of the written and oral exercises conducted in this study.

Chapter 6 is concerned with practical aspects of the setting as well as with the evolution of the research project.

Chapter 7 includes a description of the evaluation task in the *End-of-the-Year-Test08* and of the students' responses as well as an analysis of the responses. The chapter also includes findings regarding the students' criteria when evaluating mathematical proofs. Relating these findings to the suggested schema provides some insights about the students' evaluation habits and their views on purpose(s) of proof.

Chapter 8 is concerned with the interviews held in March 2009. It includes considerations on the interview design as well as a description of the employed tasks and interview transcripts. This chapter also involves interpretations of the transcripts in the light of the suggested schema.

Chapter 9 summarizes the Practical Part of this thesis.

The **Concluding Discussion** considers some advantages and shortcomings of the proposed schema to interpret proof evaluations. It also suggests inclusion of proof evaluation activities in small groups into learning activities for novice students and discusses possibilities for further research.

Part I

Theoretical Part

Chapter 2

Theoretical Background

In this part of the thesis I describe the theoretical framework used in this study. The focus is on its use for describing learning processes in the context of mathematical learning especially when validating and evaluating proofs. A sociocultural learning theory is considered particularly suitable for observations on students' learning in the transition from second level to university education. Hemmi (2006) developed in her doctoral thesis a theoretical frame to describe how students encounter proof when studying mathematics at university level in Sweden. Her theoretical frame combines a sociocultural perspective with Lave and Wenger's (1991) and Wenger's (1998) social practice theories and theories about proof obtained from Mathematical Education research. In this study I adopt parts of Hemmi's theoretical framework and its terminology and combine it with some new ideas to investigate and describe how students validate and evaluate mathematical proofs. Another aim of this study is to develop and test a schema to describe and analyse students' proof validation and evaluation skills and habits.

This chapter is concerned with the view of learning within the theoretical framework.

Section 2.1 gives an introduction into sociocultural learning theories, beginning with an explanation of what Lave and Wenger mean by a *community of practice* and how they interpret *learning* within this theory (2.1.1). This is followed by considerations of the community of mathematical practice and reflections about how learning can be viewed in such a community, namely as participating in the practice (2.1.2).

Section 2.2 considers the nature and role of artifacts both from a philosophical and sociocultural viewpoint.

Section 2.3 introduces the notion of the *condition of transparency* of artifacts. Finding the right balance in the condition of transparency is regarded as crucial in teaching and learning processes.

2.1 Learning from a sociocultural view

The transition from second level education to university can be a particular challenge for students. The psychologists Bonica and Sappa (2008) describe two major reasons to consider the first year of university to be a significant and critical period of transition. Firstly, the transition exposes students to several discontinuities in the learning setting in terms of both the complexity of study activity and a new system of roles and relationships. The new academic environment requires students to improve their meta-cognitive skills in order to study a greater quantity of material, to monitor and organize their learning and study time more independently, and to develop the ability to critically analyse and integrate a variety of texts and theoretical approaches. Success in confronting this learning challenge depends on the extent to which students are able to improve these skills and the opportunities they are presented with to reinforce their self-confidence in facing these new learning tasks. Secondly, these challenges are faced within a new learning community that involves several roles and relational discontinuities. University classes are often larger than secondary school classes and relationships with classmates and teachers lack the closeness and familiarity that is common in the secondary school class setting. There is a greater degree of formality in the student-teacher relationship and the academic context requires the student to be more autonomous in studying and asking for help. Bonica and Sappa argue that students' initial response to the new context can influence their learning processes considerably. In order for students to take advantage of the collective context to support their learning, it is important that a sense of belonging to the new learning community is developed.

Based on the above mentioned aspects, a sociocultural theoretical framework is considered a particularly suitable perspective for observations on students' performance during the transition from school to university. Providing access for the learners to a community of practice is the key for progress in learning in Lave and Wenger's (1991) theory of *legitimate peripheral participation*. Their theory is influenced by Vygotsky's (1978) sociocultural theories. Vygotsky's approach is significant for a change of the view of learning in the field of education research: from mostly influenced by constructivistic views concerned with the development of individual learners (e.g. Piaget or Bruner) to more sociocultural approaches to learning. A key feature of the sociocultural view of human development is that learning develops out of social interaction. Vygotsky argues that a child's development cannot be understood by a study of the individual in isolation. The external social world in which that individual life has developed must also be examined. Another key concept in Vygotsky's sociocultural theory is that all human activity is mediated by tools. Vygotsky describes learning as being embedded within social events and occurring as a child interacts with people, objects, and events in the environment. Lave and Wenger's (1991) theory of *legitimate peripheral participation* is a situated learning theory that argues that knowledge is distributed amongst a community of practice and can only be understood with the 'interpretive support' provided by participation in the community of practice itself.

A *community of practice* is a group of people who share an interest, a craft, and/or a profession (Lave & Wenger 1991). The group can evolve naturally because of the members' common interest in a particular domain or area, or it can be created deliberately with the goal of gaining knowledge related to a specific field. Wenger (1998, p. 73) characterizes a *community of practice* by its three dimensions:

- mutual engagement (engaged diversity, doing things together, relationships, social complexity, community, maintenance),
- a joint enterprise (negotiated enterprise, mutual accountability, interpretations, rhythms, local response),
- a shared repertoire (stories, styles, artifacts, tools, actions, historical events, discourses, concepts).

Lave and Wenger (1991, p. 167) emphasize that a community of practice is not static, it is changing permanently. Over time, forms of participation and identities change as newcomers themselves become old-timers with respect to the next set of newcomers.[1] As newcomers move towards full participation, the community of practice itself changes, as do the power relations between newcomers and old-timers. Wenger (1998) describes how knowledge develops in a community of practice, using the example of a community of claim processors. In her doctoral thesis Hemmi (2006) considers a mathematics department as a community of practice.

2.1.1 Learning in a community of practice

In Wenger's (1998) view learning is about becoming, about participating in practices. It occurs through participation in activities and contributes to a growing identity within the community of practice. Consequently Wenger (1998, p. 56) regards participation within a community of practice as a 'source of identity'. In the process of participation, newcomer's identities change as they are increasingly recognized as belonging to and contributing to a community of practice. The concept of learning as participation therefore helps to explain 'the evolution of practices and the inclusion of newcomers [and][...] the development and transformation of identities' (Wenger 1998, p. 13). Identity is built around social engagement and is constantly being renegotiated as individuals move through different forms of participation (Jawitz 2009). Lave and Wenger's examples of learning in communities of practice are mostly located in 'real world' work situations, for instance in their examples of a group of claim processors in an insurance company and a group of members of the Alcoholics Anonymous organisation. These communities of practice differ from communities at universities, which are mostly defined by common or shared *knowledge* (in opposition to practical work) of its members and some researchers refer to them as *communities of knowledge*. In publications about the changing nature of academic work, academic identity is viewed as both 'individual and embedded in the community of primary importance to them' (Henkel 2000, p. 251). Henkel (2000, p. 17) refers to Polanyi's *"Republic of Science"*:

[1] The terminology is adapted from Wenger: *newcomers* are new entrants, *old-timers* experienced members in the community of practice.

> "Polanyi underlines the critical importance to science of the individual with a passionate commitment to truth. He notes that science has become increasingly fragmented into specialisms; but scientists, he says, are held together by common values and conceptions of their enterprise and by the ability of specialists to communicate with those in adjacent areas of inquiry. It is therefore a vision which claims for science both unity and diversity."

Learning or becoming knowledgeable in a scientific practice means becoming a full participant in the practice, which includes sharing the common values and conceptions of the enterprise and learning to communicate in the manner of the practice. In her analysis of enculturational practices of novice scientists Traweek (1988, p. x–xi) notes that

> "Novices must learn what sort of things they need to know to be taken seriously: they must become unselfconscious practitioners of the culture, feeling the appropriate desires and anxieties, thinking about the world in a characteristic way."

2.1.2 Communities of mathematical practice

Hemmi (2006) regards a Swedish mathematics department as a *community of mathematical practice*. She considers all people learning and doing mathematics at the department of mathematics as members of a community of practice of mathematics. There are mathematicians, doctoral students, teaching assistants and undergraduate students. "It is a dynamic practice and the joint enterprise for all participants is the learning of mathematics in a broad sense." The term *mathematical practice* involves both shared practice and knowledge of its participating members. Mathematical practice includes for example its "special language, symbols, tools, documents, specified criteria and well-defined roles." The mathematical practice then defines a community through the three dimensions suggested by Wenger, mutual engagement, a joint enterprise and a shared repertoire. Hemmi (2006) describes examples for each of those dimensions in a community of mathematical practice:

- mutual engagement: studying, teaching, learning, communicating maths,
- a joint enterprise: learning and developing the practice,
- a shared repertoire: courses, words, symbols, artifacts, computers, ...

In Wenger's theory about communities of practice the learning progress of *newcomers* is significantly affected by two factors: the newcomer's aim to become *old-timers* and the provided opportunities to have "broad access to arenas of mature practice" (Wenger 1998, p. 110). Both factors are problematic in a community of mathematical practice, such as a mathematics department. Firstly not all students aim to become mathematicians. That is different from the examples of communities of practice described by Wenger (claim processors or the community of Alcoholics Anonymous). For example Irish first year university students learning mathematics do not necessarily aim to become mathematicians and therefore full members of the community of mathematical practice. That is only the case for a small percentage of them. Most aim to enter other communities, to learn other professions. Secondly, undergraduate students at a department of mathematics usually do not have broad access to arenas of mature practice. During the first few years at university the ratio of mathematicians to students makes it almost impossible for the teaching lecturer to provide access to their own practice, including activities like their research, communication with colleagues, participating at conferences, reading mathematical texts. Typically students only see mathematicians in lectures. The mathematicians might tell their students about their work, and of course the students see them teaching. Other aspects of the mathematicians' work are not apparent to the students.

Considering the difficulties described above for newcomers in a mathematical practice to access the practice, finding and using opportunities for newcomers to participate is important. This study is not concerned with a description of a community of mathematical practice, the interaction between members, the mutuality of engagement, etc. A focus on learning within the community of mathematical practice motivates investigation of how students are encouraged to participate and provided with opportunities to participate. One question in this study is whether the specific activities of proof validation and evaluation can provide students with such opportunities to participate in a community of mathematical practice, and whether the described theories apply; more precisely, whether we can detect a learning effect caused by these activities.

2.1.3 Learning in a community of mathematical practice

Acknowledging Wenger's sociocultural view of learning, learning mathematics at university level is conceived as increasing participation in the community of mathematical practice. Learning or becoming knowledgeable in a mathematical practice includes learning to talk about mathematics, to use its language and symbols, and it also includes learning to prove, to talk about proofs and to validate and evaluate proofs.

Beginning university students have just left another community of practice, the second level school, which can be seen as a community of schooled individuals. Old-timers in this community usually are not mathematicians, physicists, etc.[2] Lave and Wenger (1991, pp. 99/100) describe the example of a group of school students engaged over a substantial period of time in learning physics.

> "What community of practice is in the process of reproduction? Possibly the students participate only in the learning of high school itself. But assuming that the practice of physics is also being reproduced in some form there are vast differences between the ways high school physics stu-

[2] This might be different in more practically orientated subjects at school, where teachers can provide students with more opportunities to participate in practice, activities like sport events, musical concerts, building wood or metal work, cookery, etc.

> dents participate in and give meaning to their activity and the way professional physicists do. The actual reproducing community of practice, within which school children learn about physics, is not the community of physicists but the community of schooled adults."

Lave and Wenger (1991, p. 100) suggest "rather than learning by replicating the performances of others or by acquiring knowledge transmitted in instruction, [...] that learning occurs through centripetal participation in the learning curriculum of the ambient community."

An essential difference between the learning experience at school and university is the opportunity for authentic participation in a community of disciplinary practice that is provided at university. Solomon (2006, p. 386) describes the situation for the students in transition from school to university.

> "For the novice, the crucial development is to see oneself as a 'legitimate peripheral participant' (Lave & Wenger 1991) who, while not yet a full participant, has the potential to become one. But as researchers in school mathematics have shown, learners can be, and often are, excluded from the negotiation of meaning or even the beginnings of it, developing instead an identity of non-participation and marginalisation."

Her interviews with twelve first year university students suggest that this identity does not change with the transition to university. The students described themselves as outside of the mathematics community.

> "Their relationship with the lecturers required them to engage only with mathematics as already created rather than with the disciplinary process of creation and validation of knowledge, and their experience of missing explanations and exclusivity placed them on the periphery of the community."

Regarding the learning of mathematics, the first few years at university can be seen as transition from a community of schooled individuals to communities of practice and/or communities of knowledge. Compared to first and second level teaching university students are being taught by mathematicians, in Wenger's terms *old-timers* in the community of mathematical practice. As argued in former reflections, university students do not get many opportunities to participate in the community of mathematical practice compared to apprenticeship situations. Considering the number of incoming university students and the fact that most of these students are not interested in becoming mathematicians, the traditional lecture style can not be reconstituted into an apprenticeship kind of teaching style. Reproduction is a very important tool within the learning of mathematics. Nevertheless, to avail of all potential to enhance learning, the undergraduate student should become gradually more and more involved in the mathematical community through participation. Wenger (1998, p. 100) notes that

> "Observation can be useful, but only as a prelude to actual engagement. To open up a practice, peripheral participation must provide

> access to all three dimensions of practice: to mutual engagement with other members, to their actions and their negotiation of the enterprise, to the repertoire in use. No matter how the peripherality of initial participation is achieved, it must engage newcomers and provide a sense of how the community operates."

In this study I consider proof validation and evaluation as potential opportunities for novice students to participate on the periphery of the community of mathematical practice. Mathematicians validate and evaluate arguments not only when marking exam papers or discussing with students but also when reading mathematical books and articles, listening to conference talks, debating with colleagues, reviewing articles. Proof validation and evaluation therefore play important roles in a mathematician's work and are used in various aspects of a mathematician's life. (The role of proofs and their validations and evaluations in mathematical practice is discussed in more detail in Section 3.2.) It is also suitable to use as newcomer's tasks, which according to Wenger (1998, p. 110) should be "short and simple". Therefore one of the focuses in this study is on the learning effects during the processes of proof validation and evaluation.

2.2 Artifacts

The most fundamental principle of sociocultural theory is that the human mind is *mediated*. In opposition to the orthodox view of mind, Vygotsky (1978) argues that humans do not act directly on the physical world but rely, instead, on tools and labor activity, which allows us to change the world, and with it, the circumstances under which we live in the world. We also use symbolic tools, or signs, to mediate and regulate our relationships with others and with ourselves and thus change the nature of these relationships. *Physical* as well as *symbolic* or *intellectual tools* are artifacts created by human culture(s) over time and are made available to succeeding generations, which can modify these artifacts before passing them on to future generations. Included among symbolic artifacts are numbers and arithmetic systems, music, art, and language. In Vygotsky's view mental activity is organized through culturally constructed artifacts. Whether physical or symbolic, artifacts are generally modified as they are passed on from one generation to the next. Each generation reworks its cultural inheritance to meet the needs of its communities and individuals.

2.2.1 The role of artifacts for learning in a community of practice

Artifacts serve as *mediators* within a community. Participants use them to mediate the relationships among each other and, as Hemmi (2006, p. 38) explains, to mediate between the *social* and the *individual*: "We come to know the world and the culture by mediation through artifacts: for example meanings are known through language, which is [...] seen as an artifact."

In regard to a community artifacts are *products* of human activity (the practice, often work), and as well *tools* to use for activities within the community, successively being modified to increase their value and usability in the practice. Becoming knowledgeable or *learning* means increasing membership in the practice which includes the ability to use and understand its artifacts. They provide learners with opportunities to enter a community. Adler (1999, p. 49) states that

> "access to artifacts in the community both through their use and their significance is crucial. Artifacts [...] are often treated as givens, as if the histories and significance are self-evident. Yet artifacts embody inner workings that are tied up with the history and development of the practice and that are hidden. These inner workings need to be made available."

She explains this by reflecting about the example of an apprentice carpenter. If he or she aims to become a full participant in the practice of carpentry, Adler argues, it is not sufficient that he or she learns to *use* a particular cutting tool. He or she needs to *understand* the role of this tool in the overall context of the practice of carpentry as well as how and for what purpose it is used now.

2.2.2 Artifacts and their evaluations - a philosophical approach

Before considering proofs as artifacts in communities of mathematical practice, I suggest philosophical reflections on the notions of *artifacts* and their *evaluations*.

An *artifact* can be described as an object that has been intentionally made or produced for a certain purpose. Aristotle distinguished between things "that exist by nature" and "products of art" or "artificial products" (artifacts). An artifact, unlike a natural object, is a product of human actions, and therefore has a *maker* (author, designer, etc.) or a group of makers. Artifacts are often characterized in terms of functions and goals (e.g. 'hammer' or 'calculator'), a fact which refers to the essential role of the maker's and the user's intention. In general an artifact can be classified by its dual nature: its features and its purpose.

Hilpinen (2004) explains that we **evaluate** only artifacts: "It doesn't make much sense to say that a stone is good, except if it is used by a human to promote a practical end (say, to crack a nut)." Hence, an artifact's goodness simply **is** its instrumental value, its value of being a means to an end.

Hilpinen describes how artifacts can be evaluated. He distinguishes between the *intended character* of an artifact, its *actual character*, and a *purpose* F, and evaluates on the basis of the relationships among those facets:

($E1$) The degree or fit of agreement between the intended character and the actual character of an object,

($E2$) The degree of fit between the intended character of an object and the purpose F, in other words, the suitability of an object of the intended kind for the purpose F,

($E3$) The degree of fit between the actual character of an object and the purpose F, that is, the suitability of an artifact for F.

($E1$) determines whether an artifact is a successful embodiment of the maker's intentions, ($E2$) determines whether the character that the maker intends to give to an artifact is suitable for the purpose F, and ($E3$) tells whether the author has succeeded in making an object that is in fact suitable for the purpose F. The study of artifacts is intrinsically evaluative, since viewing an object as an artifact means viewing it in the light of intentions and purposes.

The purpose F on which the evaluation of an artifact and its design are based need not be the purpose that the maker had in mind; it can be any purpose for which the artifact might be used. The direction of evaluation may be reversed so that the maker or user of an artifact tries to find new uses for it.

After suggesting the consideration of proofs as artifacts and their role in mathematical practice in Chapter 3, I apply this philosophical approach to evaluations of the artifact proof and use it to describe how students evaluate proofs (Section 4.1).

2.3 The condition of transparency: the balance between *visibility* and *invisibility* of artifacts

The *condition of transparency* identifies the relation between using and understanding artifacts. According to Lave and Wenger both need to be learned in the process of gaining full participation in a practice. A carpenter needs to learn about the nature of a chisel or a drill machine as well as how to use it properly. When using it, the tool is *invisible* in the sense that the carpenter concentrates on the production of some wooden piece, and his attention focuses on the stages of this production; the tool does not get any attention, it is *hidden* in the carpenter's mind (as long as it is working properly). On the other hand, when learning about the nature and features of a tool, the tool is in the focus of the carpenter's attention, it is *visible*. *Visibility* and *invisibility* of an artifact and the balance between both is always to be seen in relation to a certain user or a group of users. *Visibility* of an artifact can be explained as being in the focus of the user's attention, *invisibility* as not being in the focus but necessary to a certain production, which *is* in the focus of the user's attention. Lave and Wenger (1991, p. 103) interpret *transparency* of resources as the interplay between the dual characteristics of *invisibility* and *visibility*: "invisibility in the form of unproblematic interpretation and integration into activity, and visibility in the form of extended access to information". *Visibility* and *invisibility* are not just two different attributes of artifacts. According to Wenger (1991, p. 108), one would not exist without the other.

> "Visibility and invisibility in transparency [...] are not mere opposite; neither are they two distinct ways of being in relation with the world. The two are always essential to each other."

The carpenter would not be able to use a tool without understanding it, and at the same time he or she would not understand the tool without the opportunity to use it. Not only for a single user, but also in a community of practice is transparency of the artifacts crucial. For Lave and Wenger, artifacts in a practice, like the carpentry tool, need to be "visible so that they can be noticed and used, and they need to be simultaneously invisible so that their attention is focused on the subject matter, the object of attention in the practice", as cited in (Adler 1999, p. 50).

Lave and Wenger (1991, pp. 101-103) call the balance between visibility and invisibility the *condition of transparency.*

> "Transparency refers to the way in which using artifacts and understanding their significance interact to become one learning process."

Lave and Wenger (1991) state that access to a practice requires its artifacts to be *transparent.* Consequently, seeing *talk* or *proof* as artifacts in a mathematical community, this idea has been applied in the context of mathematical education by Adler (1999), who reflects about transparency of talk in the teaching of mathematics, and by Hemmi (2008), who considers transparency of proof as crucial for the learning of mathematics. The *condition of transparency* of proofs in a mathematical practice is the balance between *visibility* and *invisibility* of proofs. Finding the right balance between focussing on a proof and using it to focus on the proven mathematical fact is a major challenge in the teaching of mathematics, named by Adler and Hemmi as 'the dilemma of transparency'. Devising or studying a proof might involve a focus on objects and properties particular to the mathematical topic (e.g. group theory, analysis or set theory). The nature of proof may not be in the forefront of the student's attention, i.e. may be *invisible.* At the same time, proving and its techniques can require explicit attention and therefore proof needs to be *visible.* Learners need to understand the significance of mathematical proof. These are the dual characteristics of a *transparent* artifact.

Regarding students' proof validation and evaluation skills and habits, the significance of *transparency* of artifacts, i.e. proofs, in the learning process leads to two questions:

1. Do features and purposes of proof(s) become *visible* to students through the practice of proof validation and evaluation?
2. Is the observation of students' behaviour when evaluating proofs an appropriate method to provide insights about to what degree features and purposes of proof are *visible* to the students?

In proposing a schema to interpret students' proof validation performances in Section 4.2, I include considerations on how features and purposes of proof become *visible* to students.

Chapter 3

Proofs in Mathematical Practice

Inspired by Adler (1999) and Hemmi (2006), I argue in Section 3.1 of this chapter that proofs can be seen as intellectual artifacts in mathematical practice.[1]

Section 3.2 considers purposes of the artifacts proofs and the role of proof validations and evaluations in communities of mathematical practice.

3.1 Proofs as artifacts

A key concept from sociocultural theory is that all human activity is mediated by tools or artifacts. In a mathematical practice artifacts can be either physical tools like texts, computers, maps, calculators, the document preparation system LaTeX, white boards or intellectual tools such as the specific language and symbols, discourses or systems of ideas. Hemmi (2006) argues that proof can be seen as intellectual artifact as well. If all human action is mediated by artifacts then one significant artifact in mathematical practice is proof. Hemmi (2006) took up Adler's idea to consider *talk* as an artifact in mathematical learning and applied it to mathematical *proof*. Based on Lave and Wenger's (1991) sociocultural perspective Hemmi describes how students and mathematicians approach proof in a mathematical practice.

[1]In the context of purposes of proofs I consider particular *proofs*, not *proof* in general, acknowledging that proofs serve a range of functions or purposes. These purposes weigh differently, depending on preferences of authors and readers and also on the circumstances of the presentation of a proof. For example a presentation of a 'visual proof' in addition with informal explanation might be preferred in a conference talk, whereas a 'formal' presentation of the proof would be preferred in an article in a scientific journal.

3.2 The role of proofs and their validations and evaluations in mathematical practice

Section 2.1.3 was concerned with how learning appears in a community of mathematical practice. Based on Wenger's theory, providing access for students to the community of mathematical practice was considered crucial for their learning of mathematics. Access to the community includes gaining familiarity with the artifacts of mathematical practice and participating in its activities. This chapter is about the significance and purposes of artifacts of the type *proof* and the role of the activities *proof validation* and *proof evaluation* in mathematical practice.

As described in Section 1.1, the essential role of proofs in mathematics is widely acknowledged and, as Mariotti (2006, p. 173) points out, "there seems to be a general consensus on the fact that the development of a sense of proof constitutes an important objective of mathematical education".

An important function of proofs is to establish the truth of mathematical statements. However, to see this as the only function of proof would be to take a limited view to explain why searching for new or improved proofs for already established statements is common practice in mathematical communities. The mathematician and philosopher Avigad (2006, p. 153) notices that

> "a much broader range of terms is employed in the evaluation of mathematical developments: concepts can be fruitful, questions natural, solutions elegant, methods powerful, theorems deep, proofs insightful, research programs promising."

He further describes that

> "we [mathematicians] often value a proof when it exhibits methods that are powerful and informative; that is, we value methods that are generally and uniformly applicable, make it easy to follow a complex chain of inference, or provide useful information beyond the truth of the theorem that is being proved".

Based on examples of well-known mathematical proofs, Avigad (2006, p. 105/6) shows that very often new proofs of an old theorem are valued. He deduces:

> "Put simply, the challenge is to explain what can be gained from a proof beyond knowledge that the resulting theorem is true. Of course, one sense in which a proof may be viewed as constituting an advance is that it may actually establish a stronger or more general statement, from which the original theorem easily follows. But even in cases like these we need to account for the intuition that the proof can also augment our understanding of the original theorem itself, providing a better sense of why the theorem is true."

Comments of other mathematicians also advance that the view of purposes of proofs needs to be wide.

> "Proofs are for the mathematicians what experimental procedures are for the experimental scientist: in studying them one learns of new ideas, new concepts, new strategies - devices which can be assimilated for one's own research and be further developed. [...] in studying proofs [...] we are also engaged in a cumulative collective verification process." (Rav 1999, p. 20)

Hanna and Barbeau (2008, p. 347) note the following comment by Zeilberger:

> "The value of a proof of an outstanding conjecture should be judged, not by its cleverness and elegance, and not even by its 'explanatory power', but by the extent in which it enlarges our toolbox."

Acknowledging that finding new proofs is an important activity in a mathematician's work, a broader range of functions or purposes of proofs than that of establishing the truth of a statement should be recognized. Functions[2] of proofs in mathematical practice have been widely discussed within the mathematical education literature in the last four decades. Bell (1976) distinguishes three functions of proofs (verification, illumination and systematization). His distinction has been expanded by De Villiers (1999, p. 3, italics in source):

- *verification* (concerned with the truth of a statement)
- *explanation* (providing insight into why it is true)
- *systematization* (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)
- *intellectual challenge* (the *self-realization/ fulfillment* derived from constructing a proof)

De Villiers' suggested model for the functions of proof has been widely accepted and applied within the mathematical education community, for example by Hanna (2000) and Weber (2002). De Villiers explicitly states that he does not claim his list of functions to be complete. Expansions have been suggested, such as an *aesthetic* function, and a function of *memorization* and *algorithmization* as mentioned in (De Villiers 1999, p. 11). Hanna (2000, p. 8) suggests adding *construction of an empirical theory*, *exploration of the meaning of a definition or the consequences of an assumption* and *incorporation of a well-known fact into a new framework and thus viewing it from a fresh perspective*.

Hanna and Barbeau (2008, p. 348) refer to the first five bullets of the list (De Villiers' original model) and claim that "it stopped short of stating that proof contains techniques and strategies useful for problem solving." They are inspired by Rav's paper (1999, p. 20, italics in source) in which he claims that

[2]With *function* of proof I mean with (De Villiers 1999, p. 1) "meaning, purpose and usefulness" of proof.

> "*proofs rather than the statement-form of theorems are the bearers of mathematical knowledge*. Theorems are in a sense just tags, labels for proofs, summaries of information, headlines of news, editorial devices. The whole arsenal of mathematical methodologies, concepts, strategies and techniques for solving problems, the establishment of interconnections between theories, the systematization of results - the entire mathematical know-how is embedded in proofs."

With Rav's considerations in mind, De Villiers' suggested list of functions of proof can be expanded by:

- *mediation of problem solving competencies* (the transmission of techniques and strategies useful for problem solving).

This function of proofs is strongly connected with that of *communication*. Similar connections can also be found among the other functions of the list. De Villiers (1999, p. 11) explains that though the suggested functions of proofs can be distinguished from one another, "they are often interwoven in specific cases. In some cases certain functions may dominate others, while in some cases certain functions may not feature at all."

In Section 2.2.2 evaluation of an artifact was described as relating the actual and intended character of the artifact with each other and also with its purposes. Considering proofs as artifacts in mathematical practice, the suggested functions of proofs also describe purposes of proofs within the mathematical practice. Thus, proof evaluation means relating actual and intended character of the proof with its purposes as listed above. A 'good' proof establishes the truth of the statement and is likely to also meet a number of other purposes from De Villiers' list. *Proof evaluation* in mathematical practice will be described in more detail in Section 4.1.

The considerations above about the extended view of the role of proofs in mathematical practice are compatible with the shift of research focus from *proof validation* to *proof evaluation* as explained in Section 1.2. *Proof evaluation* embraces both the question of 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 its relation to the functions of proof as listed above.

Thurston's considerations highlight the crucial role of the mathematical community in the context of *proving* (see also Figure 1.1).

> "The people who see the way to proving theorems are doing it in the context of a mathematical community; they are not doing it on their own. They depend on understanding of mathematics that they glean from other mathematicians. Once a theorem has been proven, the mathematical community depends on the social network to distribute the ideas to people who might use them further." (Thurston 2006)

> "Mathematics is indeed done in a social context, but the social process is not something that makes it *less* objective or true: rather the social process *enhances* the reliability of mathematics, through important checks

> and balances." (Thurston 1994)

If proving is seen in the context of a mathematical community, the colleague mathematicians' critical and sceptical reading of proofs must be regarded as important activities in mathematical practice. Acknowledging that access to a practice and gaining familiarity with its artifacts is crucial for learning progress, practice of the activities *proof validation* and *proof evaluation* is considered beneficial for learning of mathematics at university.

In the next chapter I will suggest a schema to interpret students' proof evaluations based on Hilpinen's description of evaluation of artifacts and the purposes of proofs considered above.

Chapter 4

A Schema to interpret Proof Evaluation

A philosophical description of evaluations of artifacts was introduced in Section 2.2.2. In this philosophical view, *evaluating* an artifact means relating its three facets: its actual character, its intended character and its purpose.

The suitability of an artifact for its purpose(s), in the context of evaluation, has been emphasized in other research areas as well. For example the cognitive educational psychologist Bereiter (2002, p. 476) stresses the significance of the relation between an artifact (he calls it 'tool') and its purposes.

> "The one idea I have been advocating that seems to be at least potentially acceptable to a wide range of critics is the idea of regarding theories and the likes as tools. [...] The tool idea is not hard to grasp and it would seem to be safe from the worst excesses of relativism. No one would claim that every tool is as good as every other. The value of a tool is relative. But always relative to some purpose."

In the first part of this chapter Hilpinen's (2004) model of evaluation of artifacts is applied to describe the evaluation of the artifact proof.

The second part of this chapter suggests a schema to interpret students' proof evaluations.

4.1 Proof evaluation

Figure 4.1 below describes how artifacts of the type *proof* can be evaluated, applying Hilpinen's description of evaluations of artifacts in general. A proof can be evaluated by relating the three facets of an artifact, its *intended character*, its *actual character*, and its *purposes*.

In this graphic the relationships among the three facets are labelled E_{AI} to E_{AP},

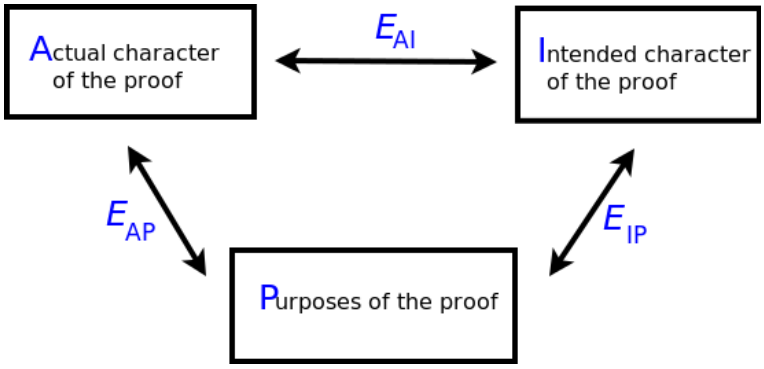


Figure 4.1: Evaluation of the artifact proof

where E symbolizes 'Evaluation', A the 'Actual character', I the 'Intended character' and P the 'Purpose(s)' of the proof.

- E_{AI} is concerned with how the 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.
- E_{IP} is concerned with how the intended proof, the author's idea of the proof, is suitable for its purposes. Is the idea appropriate to prove the mathematical statement?
- E_{AP} is concerned with how the author was successful in proving the mathematical statement as claimed, establishing its truth, potentially convincing a mathematical community, in transmission of problem solving competencies or regarding other purposes of proofs as suggested in Section 3.2.

4.2 A schema to interpret student evaluations

In the interpretation of the transcripts of the interviews conducted in 2009 (Section 8.3.2), I focus on the students' proof evaluating habits, in particular on whether and how they reflect on the relationships E_{AI}, E_{IP} and E_{AP} among the actual and intended character and the purposes of a proof.

Figure 4.2 below demonstrates how the researcher might learn about the students' views of proofs through observations of their proof evaluation performances. An exploration of students' proof evaluation habits is documented in Part II of this thesis. The results have a potential to provide insights into

- the students' validation and evaluation habits: do they distinguish between actual features of the proof, its purposes and the author's intention and relate them to each other? (Red arrows)
- the students' proof reading habits: how do they try to understand the proof? (Green arrows)

- the students' views of what they would accept as a mathematical proof: what do they regard as features and purposes of the proof? Do features and purposes of a proof become *visible* to students through the practice of proof validation and evaluation? (Brown arrow)

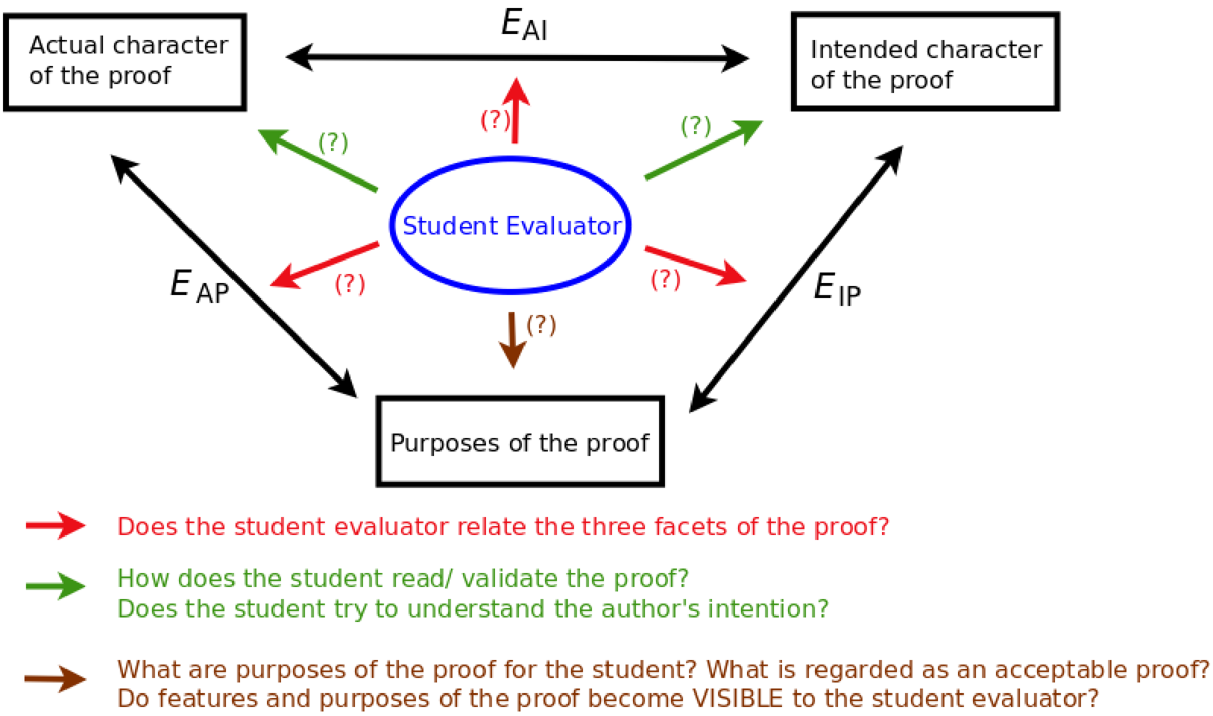


Figure 4.2: Research questions: how does the student evaluate a mathematical proof?

To show how the suggested interpretation schema can be applied I introduce the first of the proposed proofs which was presented to first year students in interviews conducted for this study. With the description of proof evaluation in mind I will reflect on how an experienced reader might evaluate it. An interpretation of the students' evaluations and rankings of this proof during the interviews will illustrate how the transcripts were used to learn about the students' evaluation habits and their knowledge about mathematical proofs.

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 E_{AI}. 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 right explanation. 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 E_{IP} and E_{AP}.

An interpretation of some transcript excerpts of first year students' evaluations of Anna's answer is suggested below. The main focus of the observation is how the students relate actual and intended character and the purposes of Anna's proof and how they view features and purposes of proof.

The transcripts of the participants' evaluations of Anna's proof refer to three groups of comments. (A presentation of more detailed interview transcripts can be found in Section 8.3.2.)

- 'Correct/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.
- 'Correct/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.
- 'NotCorrect/NotGeneral': The student does not accept the answer as a proof of the statement because it *"is not general"*.

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

Two of the five students whose opinions belong to 'Correct/NoProof' or 'Correct/Proof?' do not seem to focus on relations E_{IP} or E_{AP} as there is no evidence to suggest that they are considering the purposes of mathematical proof. Even though the other three of these students ('Correct/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 E_{AI} may be more important to them than the relations E_{IP} or E_{AP}: *"It's not a proof, but it works."*(*Students C* and *D*) or *"There is no kind of proof, it's just —. But it does make sense"*(*Student H*). Internal correctness seems to be considered more important to this group of students than the purpose of proof to establish the general truth of the statement.

The three 'NotCorrect/NotGeneral'-students do relate the actual proof not only with its intention but with its purposes, and therefore do consider relations E_{IP} and E_{AP} as well as E_{AI}. They consider at least one purpose of proof, namely its general applicability, criticize the poor relation between actual or intended proof to 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 E_{IP} and/or E_{AP} as being at least as important as E_{AI} in this instance.

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 'NotCorrect/NotGeneral'-students (*Student E*) 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 F* 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 B* like *Student F* 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 features or purposes of mathematical proof become *visible* to the student evaluator?

Student G'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 G*'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 were not 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"* (*Students* C and *D*), *"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."* (*Student E*).
- 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."* (*Student F*).

Considering Benny's answer in comparison to Anna's, some of the students who have interpreted Anna's proof as a collection of randomly chosen examples when discussing Anna's answer now identify some potential in Anna's answer to provide a general proof: *Student F* 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 B* describes Benny's answer as *"more example than proof than Anna's"*. 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.

Part II of this thesis includes a detailed description of the interviews, including tasks and full interview transcripts as well as suggested interpretations and findings regarding the research questions described above. The interviews were analysed under consideration of the interpretation schema suggested in this chapter.

Chapter 5

Summary of the Theoretical Part

In the introduction to this thesis the importance of proof in mathematical practice was emphasized. The crucial role of proof is widely acknowledged in the research literature of Mathematical Education, and more focus on proof and proving has been proposed for second level curricula in several European countries.

The introduction also clarified what is meant by *proof validation* and *proof evaluation* in this thesis and why the research focus of this study shifted from the former to the latter.

The theoretical part was concerned with a theoretical framework and development of a schema to analyse students' proof evaluations. A sociocultural theory was considered particularly suitable for investigations of students' learning progress during the transition from school to university. In the suggested learning theory, learning means increasing participation in a community of practice and gaining familiarity with its artifacts.

Considering communities of mathematical practice and proofs as artifacts in this practice, providing students with opportunities to access mathematical practice and learn to understand and use the artifact proof was regarded worthy of further investigation. The role of artifacts of the type *proof* as well as the role of *proof validation* and *proof evaluation* in mathematical practice was discussed.

I described how purposes of proofs are discussed in the fields of mathematics, mathematical education and mathematical philosophy. The research literature of the last four decades discusses how the purpose of proof is not only to establish the truth of a statement. Purposes of proofs also include enhancement of understanding of the statement and its mathematical context, convincing of the mathematical community, transmission of mathematical knowledge as well as techniques and strategies for problem solving, amongst other functions.

A philosophical approach to describing evaluations of artifacts as relating the actual and intended character of the artifact and its purpose(s) was applied to the artifact proof in mathematical practice. Using this description of proof evaluations, I developed a schema and terminology to interpret students' proof evaluation performances.

The newly introduced schema and terminology was applied to the consideration of students' evaluations of one particular example proof.

Part II

Practical Part

Chapter 6

Research Setting

This thesis includes descriptions and outcomes of a written evaluation task carried out in May 2008 and interviews conducted in March 2009. The practical part of the thesis covers design and findings of both experiments.

Chapter 6 is concerned with practical aspects of the research setting as well as with the evolution of the research project.

Chapter 7 includes a detailed description of the evaluation task in the so-called *End-of-the-Year-Test08* and of the students' responses as well as an analysis of the responses. The chapter also includes findings regarding the students' criteria when evaluating mathematical proofs. Relating these findings to the schema suggested in Chapter 4 provides some insights about the students' evaluation habits, i.e. how they relate the three facets, actual character, intended character and purpose(s) of a proposed proof, and their views on features and purposes of proofs.

Chapter 8 is concerned with the interviews held in March 2009. It includes considerations on the interview design as well as a description of the employed tasks and interview transcripts. This chapter also involves interpretations of the transcripts in the light of the schema suggested in Chapter 4.

Chapter 9 summarizes the findings of the practical experiments.

6.1 The Experiments: An Overview

The study is based on several tests and interviews with first year mathematics students at NUI Galway. The students' post-secondary experience in mathematics at this stage comprises two theoretically focussed introductory courses in Algebra and Analysis and a weekly workshop to practise basic mathematical skills and talking and teamwork about/on unusual mathematical problems. In these workshops a group of lecturers and myself work intensively in small groups with the students. One of our main impressions is that at the end of the students' first year at University many of them are not able to recognize a satisfying mathematical proof.

During the academic years 2007/08 and 2008/09 various written and oral exercises were conducted. Experiences with the early exercises influenced the design and focus of further research.

- In the early weeks of the academic year 2007/08 a so-called *Diagnostic-Test07* was carried out. The aim of this test was to get a picture about the students' approaches to solving mathematical problems. 82 students attended the Diagnostic Test.

- At the end of the same academic year (in May 2008) a final exercise was carried out as one of the last workshops of that year. The so-called *End-of-the-Year-Test08* was attended by 37 students.[1] This test was designed to analyse certain mathematical skills such as generalizing, formulating conjectures or justifying and to investigate learning progress during the first year in University.

- At the same time first pilot interviews were held with four students aiming for similar insights.

The students' approaches to validation and evaluation of mathematical proofs caught my interest when analysing responses to the tests and interviews described above, in particular one evaluation task which was part of the *End-of-the-Year-Test08*. The undergraduates' responses to a number of proposed proofs of a relatively simple mathematical statement indicate what they find essential in a good proof. Chapter 7 of this thesis is concerned with the students' performance in this evaluation exercise. The design of the next exercises for use in the interviews in 2009 focussed on proof validation and evaluation.

- The *Diagnostic-Test08* for the new incoming students included questions designed to get insights about proof validation skills. 103 students participated. The analysis of this test is not included in this thesis.

- In March 2009 a questionnaire was tested with seven local mathematicians from the School of Mathematics at NUI Galway, on a pilot basis. The aim of this survey was to investigate whether, or under which circumstances, mathematicians accept visual representations as proofs. The mathematicians' responses influenced the design of research instruments.

- Based on the findings of the analysis of the written exercises interviews were designed to be held with a smaller number of students. One aim of the interviews was to achieve a clearer understanding of the students' views on mathematical proofs. Other aims were to investigate their proof validation skills and habits and whether practice of proof validation has the potential to advance learning processes. The interviews were held with eight undergraduates in March 2009. Chapter 8 of this thesis is concerned with the students' performance in these interviews.

[1]The comparatively poor attendance of participants can be explained by two factors: some of the students had changed to a lower level (Pass) after the first Semester and also the attendance in the workshops was not as good as in the beginning of the year.

Some of the questions used in the written exercises were adapted (with permission) from the *Longitudinal Proof Project* (Küchemann & Hoyles 1999–2003).

6.2 Evolution of the Research Project

This study can be regarded as an *exploratory research* project. Exploratory research is a type of qualitative research conducted for a problem that has not been clearly defined. The sociologist Schutt (2009, p. 47/48) cites Brewer and Hunter's (1989) description of this research method.

> "The researchers begin by observing social interaction or interviewing social actors in depth and then developing an explanation for what has been found. The researchers often ask questions such as 'What is going on here?', 'How do people interpret these experiences?' or 'Why do people do what they do?'. Rather than testing a hypothesis, the researchers are trying to make sense of some social phenomenon. They may even put off formulating a research question until after they begin to collect data – the idea is to let the question emerge from the situation itself."

Exploratory research helps determine the research design and data collection methods. It should draw definitive conclusions only with extreme caution. The results of exploratory research can provide significant insight into a given situation. Although the results of qualitative research can give some indication about the "why", "how" and "when" of some occurrence, it cannot tell us "how often" or "how many".

This methodology is also at times referred to as a grounded theory approach to qualitative research or interpretive research, and is an attempt to unearth a theory from the data itself rather than from a presupposed hypothesis. One objective of exploratory research is to gather preliminary information that will help define problems and suggest hypotheses.

An exploratory study is undertaken when not much is known about the situation at hand or limited information is available on how similar problems or research issues have been tackled in the past. Exploratory studies are also necessary when some facts are known but more information is needed for developing a viable theoretical framework.

As the research literature on students' proof validations and evaluations is not extensive, exploratory research methods were chosen for this study.

Students' responses to a written evaluation task were used to develop a research focus. The written evaluation task was part of a wider written activity carried out in May 2008 in a weekly workshop. The responses to the evaluation task caught my attention, and I analysed them in detail. Some workshop experiences involving students' approaches to mathematical proof and, more specifically, the students' responses to the evaluation task can be seen as an initiation to a grounded theory approach. *Grounded theory* was developed by the two sociologists Glaser and Strauss (1967). Schutt (2009, p. 379) explains the goal of grounded theory as

> "to build up inductively a systematic theory that is 'grounded' in, or based on, observations. The observations are summarized into conceptual categories, which are tested directly in the research setting with more observations. Over time, as the conceptual categories are refined and linked, a theory evolves. [...] As observation, interviewing, and reflection continue, researchers refine their definitions of problems and concepts and select indicators. They can than check the frequency and distribution of phenomena [...] "

The written experiment was **not** designed in anticipation of certain research questions. Nevertheless, its outcome does provide insights into students' proof evaluation criteria and was used as a basis to design further research instruments.

Data used for this thesis consist of these responses to this evaluation task, and of task-based semi-structured interviews that were conducted later. *Interviews* were chosen as a method to gather more data because interviews seemed most suitable to gain a *deeper* insight into the students' proof validation and evaluation skills and habits. Unlike questionnaires or other written experiments, interviews facilitate opportunities to probe or ask follow-up questions. Also, interviews are generally easier for the respondent, especially if what is sought is opinions or impressions. As interviews can be very time consuming and resource intensive, I decided to conduct interviews with a small number of students (fewer than ten).

Lindlof and Taylor (2002, p66/67) describe a typical qualitative study as

> "*cyclical* in its basic movements: that is, most qualitative studies cycle many times through the same steps (e.g. with researchers doing scouting, data collection, data review and/or analysis, and interpretation, then doing them all over again). The process goes on until the researcher 'gets it right' – until an insightful interpretation has been achieved."

Figure 6.1 below describes how this cyclical element is significant in this study as well, starting with the inner cycle arrow at *End-of-the-Year-Test08*. The interpretation of the written evaluations influenced the research focus and with that the design of further research instruments, i.e. the interviews. The first analysis of the interviews instigated the development of a schema and associated terminology to interpret proof evaluations. The schema enabled a sharpening of the research focus. With the new schema in mind, both the interviews and the written student evaluations were interpreted again. This second analysis and interpretation led to a further narrowing of the research focus, some findings about students' proof evaluation habits, and about what criteria and purpose(s) of proof they consider when evaluating mathematical proofs.

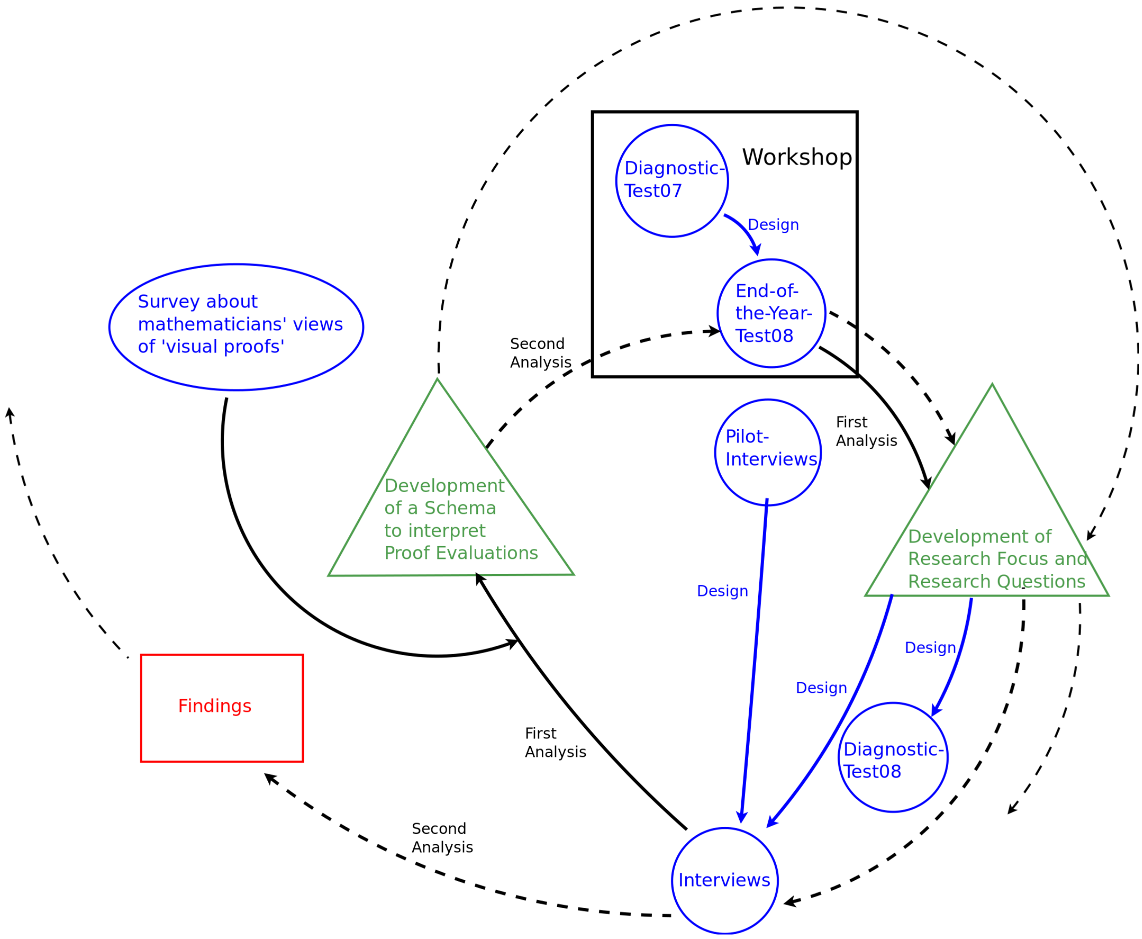


Figure 6.1: Evolution of the Research Project

The writing of this thesis is part of the cycle described in the diagram above, and gathers its strands together without necessarily marking the end of its evolution. I see the interpretation of the findings as a non-static process, which will be further developed and improved with additional studies and reflections.

Chapter 7

Written Evaluation Task

In May 2008 a final exercise was handed out in the workshops, the *End-of-the-Year-Test08*, which 37 students attended.[1] The aim of the exercise was to access and analyse certain mathematical skills such as generalizing, formulating conjectures or justifying and to investigate learning progress during the first year in university.

One of the questions in the test was an evaluation task, which is described in Section 7.2. This section also includes reflections on how an experienced evaluator might consider the proposed proofs.

The methodology used to manage the acquired data is described in Section 7.1.

Section 7.3 gives an overview of the acquired data and an interpretation of the students' responses to the evaluation task in relation to their proof evaluation criteria. This section also includes how those findings relate to the proposed schema to interpret proof evaluations.

Section 7.4 summarizes the findings of the written experiment.

7.1 Methodology

Earlier experiences with questionnaires and interviews as well as similar outcomes in the research literature indicate that asking directly what is important in a mathematical proof might not provide valuable information about the students' criteria when evaluating proofs.[2] However, the task of evaluating a variety of different and

[1] In the beginning of the academic year 2008/09 (September 08) we asked the new incoming students to perform similar exercises. This time 103 students attended. The results of the second test (*Diagnostic-Test08*) are not included in this analysis.

[2] Selden/Selden (2003, p. 27) asked their interview students how they read proofs after letting them validate four different arguments. They conclude "what students say about how they read proofs seems to be a poor indicator of whether they can actually validate proofs with reasonable reliability. They tend to 'talk a good line'. They say that they check proofs step by step, follow arguments logically, generate examples and make sure the ideas in a proof make sense. However, their first reading judgments yield no better than chance results, suggesting that they cannot

partly incorrect proofs of a not too difficult mathematical statement as one question 'hidden' in a series of questions, was considered suitable to provide some 'realistic' insights into the students' proof validation and evaluation criteria.

The proof evaluation task described in this thesis was Question No. 5 in a series of seven questions, carried out in one of the last workshop sessions of the academic year 2007/2008. The students had about 40 minutes to attempt all seven questions. Most of them finished earlier, so time pressure did not seem to be an issue. The evaluation task was designed to assess how the students validate proofs after their first year at university. It was **not** particularly designed in view of this research project. Nevertheless, the responses caught my attention and therefore I chose to investigate further proof validation habits of students. Data from 37 students seemed a good basis for a deeper analysis.

The students' written responses were categorized, coded and analysed by using the open source database MySQL. Analysis, categorization and coding of the data were partly accomplished in parallel order. That is, in analysing the data, new categories emerged and prompted introduction of new codes. The entire data analysis then had to be revised in terms of the new categories. During this coding and analysing process the research focus evolved.

For example one category was concerned with the question whether a proposed proof was regarded as sufficient or not in order to establish the truth of the statement. Using the term most commonly used by the students to express these opinions, they were coded as follows:

Code	**Description of response**
PC	The student regards the proposed proof as *"correct"*.
NoPr	The student does not regard the proposed argument as a *"correct"* proof of the statement.

In the course of the data analysis some responses did not fit into either of the above categories. Another category was included as follows:

PrNotEnough	The student regards the approach as not enough to be a proof.

The Coding Scheme can be found in Appendix B.

After the data was coded, the frequency of each category was determined. The investigation concentrated on the variety of approaches and attitudes towards the proposed proofs. During the analysis in the early stages of the research project, attention was directed at the criteria that the students appear to use in their proof evaluations. After development of the schema to interpret proof evaluations, a revision of the analysis also focussed on the students' evaluation habits, in particular on how they relate these proofs to an assumed intention and purposes of proof, and on what the students view as purposes of proof.

reliably implement their intentions."

7.2 Question Design

The undergraduates were presented with attempts of six fictional students Aoife, Barry, Cathy, Dan, Eve and Finn, to prove whether the following statement is true or false.

When you add any two even numbers, your answer is always even.

For each response, the students were asked to give a mark out of five and a line of advice. The proposed 'proofs' (labelled by fictitious names in alphabetical order) were chosen to provide a variety of different features to allow identification of the students' criteria for accepting or valuing a proof. Aoife's proof is an example of a correct, algebraically presented proof of the statement. Barry's answer consists of a collection of examples verifying the statement. Cathy's proof is correct and written in text. Dan's answer is an example of an unusual approach to justify the statement. Eve's answer is wrong, but presented algebraically. Finally Finn's proof is an example of visual reasoning[3]. To distinguish 'old-timers' from 'newcomers' (see p. 15) in the context of proof evaluation I introduce the terms *Experienced Evaluator*[4] and *student evaluator*. In the following each of the given answers is provided, followed by a comment about how an *Experienced Evaluator* might consider the proposed proof. Each of these comments consists of a direct reaction to the proposed proof, followed by considerations how this reaction fits into the schema as suggested in Section 4.2. (The task can also be found in Appendix A.)

***Aoife's answer*:**
a is any whole number.
b is any whole number.
$2a$ and $2b$ are any two even numbers.
$2a + 2b = 2(a + b)$.
So the statement is true.

***Experienced Evaluator*'s view of Aoife's answer:**
Aoife's answer is a correct proof of the statement. Her idea is appropriate for the purpose of explaining why the statement is true; thus her intention is well matched to the purpose of her proof. Aoife's proof could be criticized on the grounds that

[3]The term *visual reasoning* is used here as Dreyfus (1999, p. 105f) suggests: "to refer to arguments based on analysis of a diagrammatic situation". Dreyfus adds that "such reasoning may include analyzing, acting on and transforming images, mental or external ones, and drawing conclusions about mathematical relationships from these actions." He further claims that the "question under what conditions, and according to which criteria, visually based explanations can and should be accepted has received little attention".

[4]The view of a representative *Experienced Evaluator* has been developed in cooperation with my supervisor Rachel Quinlan, who is an experienced mathematician herself. The various proofs were discussed by mathematics lecturers and postgraduate students during a departmental seminar. In the comments of the *Experienced Evaluator* we try to reflect all the different views that were expressed by these experienced mathematicians, even though they were not all consistent with each other.

rather than starting with *any* two even integers, she starts with *any two* integers and asserts without explanation that the even integers obtained by doubling these represent an arbitrary choice of two even integers. Thus Aoife's written proof is not an entirely successful realization of her intention, with the result that the actual character of her proof is not perfectly matched to its purpose.

***Barry's answer*:**

$2+2=4$	$4+2=6$
$2+4=6$	$4+4=8$
$2+6=8$	$4+6=10$

So it's true.

***Experienced Evaluator*'s view of Barry's answer:**
Barry demonstrates the truth of the statement by listing some examples. The examples are correct, so the actual character of his argument can be considered to be consistent with the intended character. However, since Barry only demonstrates the truth of the statement for a selection of examples and offers no further explanation, neither the intended character nor the actual character of his argument is appropriate for any purpose of proof.

An evaluator might suspect that Barry has a misconception about the purpose(s) of proof, perhaps considering that it is merely to provide evidence in support of the truth of the statement. Another possible view is to consider that Barry's systematic choice of examples indicates a possible intention to prove the statement by induction; if so, this intention is not realized in the actual character of Barry's answer.

***Cathy's answer*:**
Even numbers are numbers that can be divided by 2. When you add two numbers with a common factor, 2 in this case, the answer will have the same common factor. So the statement is true.

***Experienced Evaluator*'s view of Cathy's answer:**
Cathy begins by describing a property of even numbers. Her description of this property ("can be divided by 2") is a slightly clumsy and imprecise execution of her intention to give a characterization of even numbers, but a friendly reader is likely to acknowledge that this intention is clear in the context.

Cathy then appeals (without explicit mention) to the distributive laws to explain why the statement is true. A possible view is that Cathy's statement that "the answer will have the same common factor" is no more than a reformulation of the problem, and that a justification for this claim is needed. According to this view the actual character of Cathy's proof is not appropriately matched to the purpose of providing a reason for the truth of the statement, and Cathy's answer is inadequate regarding the relation E_{AP}. Another possible view is that Cathy's claim that "the answer will have the same common factor" does not need further justification, and the observation that even numbers are exactly divisible by 2 is the key step. In this context an evaluator might consider that Cathy's proof achieves not only the immediate purpose of establishing the truth of the statement but also places the statement in a wider mathematical context by pointing out that it is a special case of a more general fact. Such an evaluator might feel that Cathy's intention, while perhaps not perfectly realized in the actual character of her argument, is well matched to a range of purposes of mathematical proofs as discussed in Section 3.2.

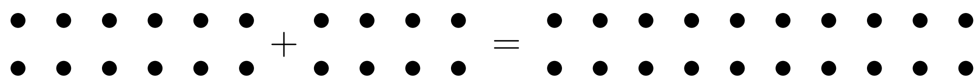
***Dan's answer*:**
Even numbers end in $0, 2, 4, 6$ or 8. When you add any two of these the answer will still end in $0, 2, 4, 6$ or 8. So it's true.

***Experienced Evaluator*'s view of Dan's answer:**
Like Cathy, Dan starts with a characterization of even numbers. The characterization that he proposes, while it is correct (in base 10 arithmetic), would not typically be regarded as a *definition* of evenness, and so it arguably should be justified. Dan's idea is to reduce the infinite domain to which the statement applies to a finite number of cases, by restricting his attention to the last digit, and by noting that the last digit of the sum of two integers depends only on their last digits. Dan's intention is appropriate at least for the purpose of establishing the truth of the statement that the sum of two even integers is even, though perhaps less appropriate than Cathy's for the wider purposes of expanding mathematical understanding and interpreting the statement in a wider context. However, whether Dan succeeds in fully demonstrating his intention in the actual character of his proof is questionable. An evaluator might consider that both Dan's characterization of even integers and his assertion about the last digit of the sum of two integers should be further explained.

***Eve's answer*:**
Let x = any whole number, y = any whole number.
$x + y = z$
$z - x = y$
$z - y = x$
$z + z - (x + y) = x + y = 2z$.
So the statement is true.

***Experienced Evaluator*'s view of Eve's answer:**
Eve's answer is wrong. If it were correct it would show that the sum of *any* two whole numbers is even, which is clearly false. Eve does not make use of the hypotheses of the statement that she is trying to prove, and it is not clear what her initial intention is in introducing two arbitrary integers x and y and considering their sum. This action does not appear to be appropriate to the problem of considering the sum of two even numbers; in any case Eve makes no mention of the terms "even" or "odd" throughout. Neither her intention nor the actual character of her argument appear to be connected to her purpose.

***Finn's answer*:**



So the statement is true.

***Experienced Evaluator*'s view of Finn's answer:**
Finn's answer admits a number of possible interpretations. What he writes is minimal and the reader is left to decipher his possible intentions.

- A reader might consider that Finn's answer consists merely of the statement "$12 + 8 = 20$", represented pictorially. Such a reader might consider that Finn has demonstrated the truth of the statement for this particular example, but done no more. If this is the extent of Finn's intention, then his intention is adequately realized in the actual character of his answer, but neither the intention nor its actualization is adequate for any purpose of proof.

- A reader inclined towards a more generous interpretation of Finn's intentions may be prepared to regard his answer as a demonstration, using a particular example, that *every* even number can be represented by a row of vertical pairs of dots as indicated, and that the geometric operation of concatenating two such rows corresponds to addition and will always result in a row of vertical pairs of dots, representing an even number. While this interpretation is generous given the limitations of what is written, it is probably more likely to be accurate, since an author who intends only to provide a demonstration consisting of one example would not have a reason to represent the example pictorially. The relation between the intended character and actual character of Finn's proof can certainly be criticized. However, an evaluator taking this interpretation of Finn's intention might consider that the intention is indeed well suited to the purpose(s) and possibly even suggests a more general statement, although less explicitly than Cathy's proof.

The following section describes how the undergraduate students who participated in the *End-of-the-Year-Test08* evaluated the proposed proofs.

7.3 Interpretation of the data

Section 7.3.1 consists of an overview of the students' responses to each of the proposed proofs.

Some of the students' comments warrant further attention. Section 7.3.2 includes reflections on a number of themes that were considered remarkable.

Section 7.3.3 considers criteria that the students appeared to have used in their written evaluations.

Section 7.3.4 discusses how those criteria indicate what these undergraduates considered to be significant purposes of mathematical proofs.

7.3.1 Overview

This section includes an overview of the students' responses to each of the proposed proofs. The students' comments are grouped in respect of their views. The frequencies of occurrence of various attitudes towards each of the proposed proofs are listed in the tables below. Each of the tables is followed by an interpretation of its contents.

- To each of the proposed proofs two types of response are distinguished:
  - Some comments reveal information about what the student approves or criticizes in the particular proof. This information indicates that they do relate the proposed proof to some purposes of proofs.
  - Other comments lack information about what is being valued or disapproved and therefore regarded as the students' evaluation criteria. These responses provide less information about the students' evaluation habits and/or views on the purpose(s) of proof.
- Most students explain in their comments what they criticize or approve about the particular proof. These topics and their frequencies are listed.
- The corresponding view of the *"Experienced Evaluator"* covers a range of opinions on how the particular proposed proof may be considered by a mathematician. These views are related to the students' responses.

The detailed responses as well as the students' markings and each coding can be found in Appendix C. The coding scheme is outlined in Appendix B.

Students' responses to *Aoife's* answer.

Marks	*Frequency*
0	3
1	0
2	2
3	4
4	11
5	17

Code of response	*Frequency*
VG	6
NoUnd	3
PC	5
NoPr	1
InclCrit	27
NotInclCrit	10
NoDef	6
Cl	4
NotClEnough	1
NoExpl	2
WellExpl	2
MExpl	1
NoEx	1
PrAsEx	3
Gen	1
IncorrectCrit	3
NoComment	6

Table 7.1: Coding table - Students' responses to *Aoife's* answer

Interpretation. The comment about how an *Experienced Evaluator* might have considered Aoife's answer indicates that it is a correct proof of the statement, which could be improved by a more careful introduction of two arbitrary even numbers. The majority of the participating students agrees with that opinion. About 75% (28 students) give Aoife's approach 4 or 5 marks out of 5.

Ten students do not explain what they criticize or appreciate about Aoife's proof. Therefore their evaluations do not give any information about their evaluation criteria or how they relate Aoife's proof to her intention or to any purposes of the proof. The other 27 students give some information on what they value or disapprove in Aoife's answer, which shows that they do relate Aoife's answer to some purposes of proofs. Those comments indicate what criteria the students find relevant to value a proof.

- Four of the students appreciate the clearness in Aoife's proof, for example:

  "Aoife has a very clear and straightforward answer",
  "Very clear",
  "Aoife is using clear and simple language to get her answer across (...)".

- Two other students approve the quality of explanation in Aoife's proof.

  "Well explained answer",
  "I think this explains it well (...)".

- One student approves the fact that the argument is applicable in general.

  "(...) it will work for all possible numbers."

- Six of the students express the opinion that Aoife's answer could be improved by a definition of an even number, e.g.

  "Define an even number before them in the proof."
  "(...) Should also point out that a number is even if divisible by 2 *more clearly."*
  "She should have described what an even number is."

  The students' views about the role of definitions in a proof is discussed in Section 7.3.2.

- A quarter of the students (9) regard Aoife's proof as relatively poor. Three of them criticize poor explanation, one a lack of examples and another student expresses the opinion that Aoife's answer is not clear enough.

Three students use arguments which are not mathematically correct to criticize Aoife's proof. One of them as well as two other students do not seem to understand Aoife's proof. Six students do not comment on Aoife's answer at all.

Students' responses to *Barry's* answer.

Marks	*Frequency*
0	3
1	9
2	18
2.5	1
3	5
4	0
4.5	1
5	0

Marks	*Frequency*
NoPr	2
PrbyEx	3
InclCrit	34
NotInclCrit	3
MExpl	1
NotGen	32
Conv	2
NoComment	1

Table 7.2: Coding table - Students' responses to *Barry's* answer

Interpretation. Only three students do not reveal what they criticize or approve about Barry's answer. The remaining 34 students offer reasons for their evaluations.

- Most students (32) recognize a lack of generality in Barry's answer. The majority conclude that Barry's approach is not a proof. Typical remarks on Barry's answer are

  "He has just given examples, this is not a proof" or
  "Not a general solution".

  However, only 12 students (32 %) give this answer zero or just one mark.

- The fact that 24 students (64 %) give Barry 2 to 3 marks suggests that they in some way appreciate Barry's collection of examples. The comments of three students indicate that "proof by example" might be seen as another type of proof, just not as good as one including a general argument:

  > *"It's not bad. But it's only proved by example"*,
  > *"Although this does prove the statement, it only does so for a few egs (...)"*.

  The students' views about the role of examples in the context of mathematical proof, as indicated by the findings of this study, will be considered in Section 7.3.2.

- Two students, while they criticize the lack of generality in Barry's answer, acknowledge that his examples might convince a well disposed reader of the truth of the statement:

  > *"Barry's answer is not answering the question he has simply 'shown' a friend he has yet to prove anything"*,
  > *"I gave him a few marks just because he's my friend but his enemy will never believe him because he only chose 6 examples out of an infinite amount."*[5]

  Other comments also indicate that convincingness of an argument is a criterion for some of the students, see Section 7.3.3.

The view of the representative *Experienced Evaluator* embraces three ways to interpret Barry's answer. As seen, the majority of students agree with the opinion that Barry's collection of examples is not appropriate for any purpose of proof. However, some students seem to take the second suggested perspective. They consider provision of evidence in support of the truth of the statement to be a major purpose of proof. None of the participating students seems to interpret Barry's answer as an attempt to prove the statement by induction.

[5]Some of the students had been introduced to Mason's (1982) idea of convincing a friend/enemy in the workshops.

Students' responses to *Cathy's* answer.

Marks	*Frequency*
0	2
1	0
2	1
2.5	1
3	7
4	15
4.5	1
5	10

Marks	*Frequency*
VerwithEx	1
VG	4
NoPr	1
PrNotEnough	1
InclCrit	27
NotInclCrit	10
PoorDef	3
Cl	2
CritLackofForm	11
MForm	1
Compl	1
Conc	1
WellExpl	1
MExpl	1
NoEx	6
AffNotJust	2
A-C:S	1
IncorrectCrit	1
NoComment	4

Table 7.3: Coding table - Students' responses to *Cathy's* answer

Interpretation. Ten students (27%) acknowledge Cathy's correct proof and give this answer full marks. Nevertheless, 23 students (62%) reduce marks to 3 or 4.

Ten students do not explain what they criticize or appreciate about Cathy's proof. The other 27 students give some information on what they value or disapprove in Cathy's answer.

- The most recurrent reason for reducing marks is a lack of formal appearance, criticized by 12 students. Typical comments on Cathy's answer are

  "The proof makes sense but she could have used a more mathematical approach",
  "Good intuitive answer but needs a mathematical proof",
  "Use mathematical notation to show this",
  "Correct answer but show mathematically",
  "The proof should be shown mathematically as well as in words",
  "needs more to be a proof".

  Findings about the students' views on the role of 'mathematical' appearance in a proof are discussed in Section 7.3.2.

- Three of the ten students who give Cathy full marks do so because they appreciate clearness and sufficiency in her answer.
- Two students criticize that Cathy makes an assertion without justification, namely the fact that common factors can be extracted.
- Three students criticize a missing condition in the description of even numbers in Cathy's answer. They propose that even numbers should be defined as 'whole' numbers which can be divided by 2. The students' views about the role of definitions in a proof is discussed in Section 7.3.2.

The view of the representative *Experienced Evaluator* covers several opinions. An evaluator might acknowledge Cathy's wider interpretation of the statement and clear reasoning and regard this proof as best of all of the proposed proofs. In this view Cathy's claim that "the answer will have the same common factor" does not need further justification, and the observation that even numbers are exactly divisible by 2 is the key step. In this context an evaluator might appreciate the fact that Cathy's proof places the statement in a wider mathematical context.

Cathy's answer might also be seen as sufficient, with minor or more significant objections. Two objections to Cathy's answer are described.

- Even numbers are not clearly characterized in Cathy's answer. This can be interpreted as a minor objection, as Cathy's intention is clear, though not fully met by the actual character of this proof.
- Cathy does not justify her statement that "the answer will have the same common factor". In this view "the answer will have the same common factor" is seen as a reformulation of the problem, and a justification for this claim is needed. This objection is more serious in terms of not meeting purpose(s) of proof. According to this view the actual character of Cathy's proof is not appropriately matched to the purpose of providing a reason for the truth of the statement.

The majority of students agree with the opinion that Cathy's answer is good but could be improved. Nevertheless, only a minor group share the criticism of the *Experienced Evaluator*: three students criticize Cathy's definition of even numbers and two the fact that she did not justify all her claims. As seen above, the reason for most students not to approve Cathy's answer fully, in opposition to the *Experienced Evaluator*, is that they criticize its appearance. Appreciation of the fact that Cathy's proof places the statement in a wider mathematical context cannot be found in the students' evaluations of Cathy's answer.

Students' responses to *Dan's* answer.

Marks	*Frequency*
0	1
1	1
2	4
3	12
4	11
4.5	1
5	7

Marks	*Frequency*
VG	2
NoPr	3
PrNotEnough	3
InclCrit	21
NotInclCrit	16
NoDef	2
Cl	2
CritLackofForm	5
NoExpl	2
NoEx	9
NotGen	2
Gen	1
Conc	2
IncorrectCrit	2
NoComment	7

Table 7.4: Coding table - Students' responses to *Dan's* answer

Interpretation. Eight students (21%) seem to be fully satisfied with Dan's answer. They give this approach 4.5 or 5 marks. 23 students (62%) give Dan's answer 3 or 4 marks, which indicates that they value the approach but would prefer some improvements. Six students (16%) give Dan's answer 2 or fewer marks.

Two students assume wrongly that Dan's considerations only hold for positive numbers, and seven students do not comment on Dan's answer.

Ten students do not explain what they criticize or appreciate about Dan's proof. The other 27 students give some information on what they value or disapprove in Dan's answer.

- Five students criticize a deficiency of mathematical formalism, which is indicated by comments such as

  > *"It is not proved mathematically (...)"*,
  > *"It should be proven formally even if it is intuitively true"*,
  > *"An unusual intuitive understanding but not based on any algebraic proof"* or
  > *"Try to come up with formula"*.

  Findings about the students' views on the role of 'mathematical' appearance in a proof are discussed in Section 7.3.2.

- Six students do not accept Dan's answer as a mathematical proof. They see Dan's approach not as a proof at all:

> *"There is no proof of this, it is just an observation"*,
> *"rather than prove it, he's just explained it"* or
> *"Not a proof, just a statement"*

or as not fully accomplished:

> *"Dan's proof here is not complete. He has made a good point but the proof has not quite come full circle"*,
> *"Again, a good answer just not strong enough"* or
> *"Needs further proof but onto something"*.

- Nine students criticize a lack of examples in Dan's approach. The students' views about the relation between examples and proofs are discussed in Section 7.3.2.
- Two students disapprove that Dan in their view did not define an even number. The students' views about the role of definitions in a proof is discussed in Section 7.3.2.
- One student appreciates a general applicability in Dan's answer, whereas two other students disapprove that Dan

  > *"is only using a small range of numbers, which is not sufficient"*.
- Two students criticize that Dan does not explain why his assertions are true.
- Two students appreciate clearness and sufficiency in Dan's answer.

The view of the *Experienced Evaluator* about Dan's answer includes that Dan's characterization of even integers should be justified and that his assertion about the last digit of the sum of two integers should be further explained. Both criticisms can be found only in a small minority of students. The majority of participating students does not seem to recognize a potential in Dan's answer to establish the truth of the statement sufficiently.

Students' responses to *Eve's* answer.

Marks	*Frequency*
0	11
1	7
2	4
3	7
4	4
5	4

Marks	*Frequency*
VG	4
NotGood	2
NoUnd	2
InclCrit	30
NotInclCrit	7
AlgWr	11
NoPr	1
NoDef	2
NotCl	2
Compl	3
NoExpl	5
MExpl	3
NoEx	2
Misc	4
NoComment	3

Table 7.5: Coding table - Students' responses to *Eve's* answer

Interpretation. It is striking that Eve's answer has been evaluated with each of the marks (0–5) by a considerable number of students. Eleven (29%) students give her 0 marks, thirteen students (35%) 1 or 2 marks, another thirteen students (35%) give Eve's answer 3 or 4 and four (10%) students even 5 marks.

Seven students do not explain in their responses what they criticize or appreciate about Eve's proof. The other thirty students give some information on what they value or disapprove in Eve's answer.

- Eleven students (29%) notice that Eve's algebraic manipulations are wrong. Consequently most of them (six students) give her 0 marks, three students give her 1 mark, one 2 and another one 3 marks.

- Thirteen students despite being unable to understand Eve's answer are not able to dismiss it either and seem to consider that it has some merit. Five of them criticize that Eve's answer is *"confusing"*, *"not clear"* or *"unnecessarily complicated"*. Eight of those students claim poor explanation in Eve's answer. They disapprove for example that Eve *"could have made it clearer at the end why it is true"* or that *"she should explain what she's doing"*. One student gives explicit reasons for reducing marks:

  > *"First mark lost because she didn't say why it's true (...)"*.

  The students' views about the role of explanations in a proof are considered in Section 7.3.2.

Nevertheless, despite the fact that these students do not seem to understand Eve's answer, they do appreciate her approach by giving it 2 to 5 marks. These students state what they disapprove in Eve's answer, but they do not explain what they value about this approach. The overall performances of this evaluation task indicate that the participating students highly value 'mathematical looking' appearance in a proof. This leads to the assumption that the fact that Eve's answer includes algebraic formulation and mathematical notation is the reason for those students to give Eve relatively high marks. The role of 'mathematical' appearance in a proof is discussed in Section 7.3.2.

- Two students who approve Eve's answer by giving her 3 or 5 marks criticize that she has not defined even numbers. The students' views about the role of definitions in a proof are discussed in Section 7.3.2.
- Two students suggest that Eve's answer could be improved with an example. The students' views about the relation between examples and proofs are considered in Section 7.3.2.
- One student claims that Eve's answer *"proves nothing"* and another student criticizes that Eve's answer is not related to the statement:

  "There is no mention of x and y being even numbers".

In Eve's case the representative *Experienced Evaluator* offers only one view to evaluate her answer, namely that it is irredeemably incorrect. This view is shared by only eleven students (29%). The vast majority (71%) of the participating students approve Eve's answer somehow by giving her 1 to 5 marks.

Students' responses to *Finn's* answer.

Marks	*Frequency*
0	4
1	6
2	12
3	10
4	1
4.5	1
5	3

Marks	*Frequency*
Int-Ex	11
Int-Pr	16
Int-?	10
VG	3
NoUnd	1
NoPr	3
InclCrit	30
NotInclCrit	7
NoDef	3
NoExpl	9
NoExplWhySt	3
MEx	1
MForm	3
NotGen	16
Misc	1
NoComment	5

Table 7.6: Coding table - Students' responses to *Finn's* answer

Interpretation. Seven students do not explain in their responses what they criticize or appreciate about Finn's proof. The other thirty students give some information on what they value or disapprove in Finn's answer.

The comment of the fictive *Experienced Evaluator* explains that Finn's visual reasoning allows two different interpretations for the evaluator, namely as a pictorial representation of "$12 + 8 = 20$" or as a graphical demonstration of a general proof. Both views can be found in the student evaluators' comments:

- Eleven students (30%) interpret Finn's visual approach as just one example, visualized in a diagram.

  > *"Again Finn's answer only covers 1 solution. He needs to give a general statement"*,
  > *"This proves that it works for* $12+8$. *It doesn't prove for all cases"*,
  > *"Not a proof, just an example"* are a few typical answers.

  They all give Finn only 0 to 2 marks (one 0, five 1 and five 2 marks). The marking is very similar to their marking of Barry's answer, which consists of a list of examples verifying the truth of the statement. The similarity of the responses of the students interpreting Finn's proof intention as comparable to Barry's indicates that they do not appreciate any benefits of this visual approach.

- Sixteen students (43%) show attempts to regard Finn's answer as a visual demonstration of a general proof of the statement. One of them gives Finn full marks, another one 4.5. Ten students give him 3 marks, three others give him 2 marks. One student gives Finn just one mark as in his view Finn fails to prove the statement:

  > *"Although intuitively you can see the idea, there is no definition of an even no, and also thus set only proves it rigorously for values 12* [and] *8"*.

  Two types of criticism are predominant in this group of students.

  - Nine students criticize that the proof idea is not explained, for example

    > *"Good idea, give explanation"*,
    > *"Nice pictures you could have written a line explaining it though"*,
    > *"No explanation. Multiples of 2, therefore even, 2 is an even number"*.

    The students' views about the relation between examples and proofs are considered in Section 7.3.2.

  - Three students miss some 'mathematical' appearance in Finn's visual approach. Though they appreciate the correct idea, they do not acknowledge its visual representation as a mathematical proof:

    > *"Good visual proof but* [doesn't] *use mathematics"*,
    > *"Good visual representation but needs notational explanation"*.

    The role of 'mathematical' appearance in a proof is discussed in Section 7.3.2.

- The responses of ten students (27%) do not clearly indicate how they interpret Finn's visual reasoning.

  - Five students spare any comments. Three of them give high marks (4 or 5)[6], one gives 2 marks, another one 0 marks.

  - One student writes that s/he does not understand Finn's approach.

  - Four students do not accept Finn's reasoning as a proof, one without explanation, one because *"words and numbers are needed"* and two because they consider that it does not provide an explanation of why the statement is true:

    > *"Intuitively correct but needs to explain why the answer means the statement is true"*,
    > *"Finn has just counted dots but has not given an explanation as to why he thinks the statement is true (...)"*.

[6] The fact that those students value Finn's approach so highly might indicate that they interpret it as visual demonstration of a general proof.

7.3.2 Considerations on selected remarkable findings

Some comments about what the students find essential in a proof warrant attention, for example the prevalence of certain expressions like *"proof by example"* or *"not mathematical enough"*. What do students think is a *"mathematical proof"*? Do most of them believe 'mathematical' means *"includes formulas"* or is that a small minority? Four particular themes are discussed below. These are considered worthy of note either because they appeared quite often or because they seemed surprising. The students' comments on the given different answers are analysed in relation to these themes.

"Proof by example": Students' views about the relation between examples and proofs

Almost all of the students did recognize that Barry's answer (see Appendix A) is not a proof and commented on that. Typical remarks on Barry's answer are

> *"He has just given examples, this is not a proof"*,
> *"Not a general solution"*.

This shows that the students are aware of the fact that a mathematical proof has to be *rigorous*, which means that it has to hold true regardless of what test someone may apply to it. They realize that showing a fact on a few examples is not sufficient to prove it. Though this certainly is an encouraging result it should be kept in mind that the students might have shown different findings if they had attempted their own answers first, without getting the chance to compare to other, more 'general' proofs.[7]

Some of the students' comments indicate that "proof by example" might be seen as another type of proof, just not as good as one including a general formula:

> *"It's not bad. But it's only proved by example"*,
> *"Although this does prove the statement, it only does so for a few egs"*.

These students see some purpose of proof as met by Barry's collection of examples. This indicates that they regard the provision of evidence in support of the truth of the statement as a purpose of a proof.

On the other hand, students do give examples a surprisingly high if not essential value within a proof. They often deduct marks on the basis of absence of examples. Some students comment on the absence of examples to explain the answers (*"Give an example to show (...)"*) but most criticize the lack of examples as an essential part of the reasoning. Some students request examples in Aoife's and Cathy's

[7] In the latest test (*Diagnostic-Test08* held in October 2008) new incoming first year honours students were asked to justify the statement themselves before they were shown prepared 'proofs'. Most of the undergraduates did that with one or a few examples. The students may have been aware of the insufficiency of their approach, but they did not indicate this. After being introduced to the various proposed proofs they did recognize Barry's answer as not sufficient.

(essentially correct) answers to *"back up proof"* or to *"finish the proof"*. One student criticizes a correct answer because the statement is *"not proven by numerical example"*. Without any examples students do not seem to be satisfied by an argumentation. A similar observation was made by Fischbein and Kedem (1982). They noticed that students, even after producing a deductive proof of a given proposition, wished to check it in a few cases, an indication that they continued to be uncertain about its truth.

One apparent reason for the students to value examples highly in the context of proofs is to reach an understanding of the statement itself and sometimes also of the argument. Another reason can be to provide evidence in support of the truth of the statement.

However, the fact that examples play such an important role in a proof to the students indicates that they do consider the purposes *verification* and *explanation*, as suggested in Section 3.2, in a proposed proof.

The students' views of the role of examples in the context of proofs require more detailed observations of proof evaluations. This was therefore considered in the design of the interviews 2009 (see Section 8.1.2).

"Not mathematical enough": Students' views about the role of 'mathematical' appearance in a proof

The following are a few selected comments on Cathy's essentially correct answer which is expressed purely in the form of standard text (see Section 7.2 above or Appendix A):

> *"The proof makes sense but she could have used a more mathematical approach"*,
> *"Good intuitive answer but needs a mathematical proof"*,
> *"Correct answer but show mathematically"*,
> *"The proof should be shown mathematically as well as in words"*.

The argument formulated as text was "not mathematical enough" to a third (32%) of the undergraduates. In comparison to Aoife's more algebraic looking approach one student comments:

> *"Although Cathy's answer is true there is too much English and does not mathematically prove it unlike Aoife"*.

These comments raise the question of what the students associate with the term 'mathematical' . Schoenfeld (1985) notes that mathematicians and students seem to have different perceptions of "thinking mathematically":

> "For the students, thinking mathematically involves algebraic tricks and formal language."

Our students' comments on the different answers indicate a similar impression: 'mathematical' to them appears to mean including

- formulas:

  "Try to come up with formula",

- algebraic equations:

  "Give clear equation to support your answer",
  "Would like to see this expanded with a general equation",

- mathematical notation:

  "Use mathematical notation to show this",
  "Cathy's answer is well written and (...) *although she should sum her answer up* (...) *using formal notation"*.

Consequently Eve's answer (see Appendix A), which is fundamentally and irreparably incorrect, but includes algebraic equations and mathematical notation, seems basically correct to more than half of the students. (51% approve Eve's approach by giving her 2 to 5 marks.) Fewer than 30% of the students recognize that this answer is wrong.

A visual approach to prove the statement, such as Finn's answer (see Appendix A), is generally not accepted by the students. 30% (eleven students) interpret the answer as just one example, represented in a diagram. 43% (sixteen students) seem to recognize the idea behind the illustration, but most of them do not acknowledge a *purely* graphical representation of a correct idea as a mathematical proof. Some students explicitly criticize the lack of 'mathematical' appearance:

> *"Good visual proof but* [doesn't] *use mathematics"*,
> *"Good visual representation but needs notational explanation"*.

A proof without numbers and words cannot be sufficient:

> *"There are no words in this proof"*,
> *"Proof is illustrated using graphics rather than numbers"*,
> *"This does not prove anything, words and numbers are needed"*.

Schoenfeld's description seems verified by the impressions arising from the remarks about what the students mean when criticizing proofs as *"not mathematical enough"*. A mathematical proof is not good in the students' view if it does not include some algebraic equations or formulas or any kind of algebraic notation.

The students' responses do not indicate their reasons to value mathematical formalism highly. This question was considered in the design of the interviews (see Section 8.1.2).

"It doesn't explain ...": Students views about the role of explanations in a proof

Hanna (1991) refers to mathematical educators who argue that whether or not an argument is accepted as a proof depends not only on its logical structure, but also on how convincing the argument is. For example Tall (1989) claims that a mathematical proof must be accepted by the mathematical community. Mason, Burton and Stacey (1982) define a proof as an argument that convinces an enemy, Hersh (1993) as a convincing argument, as judged by qualified judges.

To convince a mathematical community, a proof needs to persuade members of the community **that** the mathematical fact is true and also to provide them with understanding of the reasoning **why** it is true. This view plays a role in De Villiers' list of functions of proofs as well. He includes *verification* (concerned with the truth of a statement), *explanation* (providing insight into why it is true) and *communication* (the transmission of mathematical knowledge) (Section 3.2).

These purposes of proofs seem to play a significant role in the students' proof evaluations. Many of the students' responses are concerned with how the reasoning is explained for the reader. A few comments by students indicate the opinion that a valuable proof should convince a reader *that* the statement is true.

Mostly the answers of Eve and Finn get many remarks that they should explain their ideas better. Nine students comment on the absence of explanations of his visual reasoning in Finn's answer even though they recognize his idea, e.g.

> *"Nice pictures, you could have written a line explaining it though"*.

Four students demand more explanations in Eve's answer, for example

> *"Need more explanation"*,
> *"She should explain what she is doing"*.

The positive comments on the highly marked answers often include a note about the good explanations. A few comments on the students' favourite, Aoife's answer, are

> *"Aoife has a very clear and straightforward answer"*,
> *"Well explained answer"*,
> *"Aoife is using clear and simple language to get her answer across"*.

The students struggle more to understand Cathy's answer. Only one student mentions her convincing skills:

> *"Cathy's answer is well written and easy to comprehend (...)"*.

Two students miss an explanation of why the statement is true in Finn's answer, e.g.

> *"Intuitively correct but needs to explain why the answer means the statement is true"*,
> *"Finn has just counted dots but has not given an explanation as to why he thinks the statement is true (...)"*.

All these comments indicate that just having a good idea to prove a statement is not sufficient for most of the students. They regard the convincing nature of the argument and the quality of explanation of ideas as important criteria to value a proof. Thus, *explanation* and *communication* are considered as purposes of proofs.

"No definitions": Students' views about the role of definitions in a proof

Many of the students regard a proof as good if it is written in a certain structure, beginning with a definition. Six students criticize the lack of an explicit definition in Aoife's answer (Appendix A), for example

> *"Define an even number before* [using] *them in the proof"*,
> *"I believe Aoife's answer would be more acceptable had she defined an even number"*,
> *"State definition of even no at start"*.

Similar comments are given to Dan's (by two students), e.g.

> *"True but incomplete. It needs the definition of an even no"*,

Eve's (by two students), e.g.

> *"I believe Eve's answer is lacking as she failed to define an even number"*,

and Finn's answers (by three students), e.g.

> *"*(...) *there is no definition of an even no* (...)*"*.

Three students criticize Cathy's imprecise description of even numbers in the beginning of her proof. On the other hand, two other students explicitly approve Cathy's definitions of even numbers at the beginning of her proof:

> *"She defined even numbers first and made her proof clear"*,
> *"I believe Cathy's answer is completely correct. She has both defined & logically stated mathematical reasoning behind a statement"*.

The appreciation of precise definitions in a proof by a relatively high number of students may be related to the fact that definitions had been a major topic in the workshops a few weeks before the *End-of-the-Year-Test08* was completed. Data from the survey *Diagnostic-Test08* which was held with 103 students at the beginning of their first year at university and included a slight variant of the same task, show different findings. The new incoming students did not give the definitions the same high value within a proof. Possibly because the students participating in *End-of-the-Year-Test08* had seen plenty of definitions in their lectures and they had discussed definitions and their crucial role in mathematics in recent workshop sessions as well as lectures, they had learned to recognize their importance.

7.3.3 Findings from the written evaluations about first year students' criteria to value a proof

Investigation of proof validation and evaluation criteria has been suggested in mathematical education research. For example Selden and Selden (2003, p. 29) propose as one of several future research questions:

> "What are students' perceived and actual criteria for correctness, for example, that they agree the calculations are legitimate, that they can follow the argument line by line, that they can see why the argument begins where it does, and so on?"

Dreyfus (1999, p. 106) suggests developing criteria for teachers to use in their evaluations of students' proofs.

> "Of [...] importance, and [...] open is the development of criteria which can be used by teachers to judge the acceptability of their students' mathematical arguments, and of principles on which the development and examination of such criteria can be based."

This section is about criteria that **students** find important to accept or value a proof, their "perceived and actual criteria for correctness" and also their criteria to consider a proof as **good** in relation to a range of purposes of proofs. The first year students' comments on the proposed 'proofs' from the written exercise (outlined in Section 7.3.1) give some understanding about their criteria when evaluating mathematical arguments. The next section (7.3.4) considers how these criteria indicate which purposes of proofs the undergraduates consider relevant.

Listed below are the criteria which were identified as playing a role in the students' evaluations in this written exercise. The following criteria appear recurrently.

- After their first year in college the students are aware that showing a statement on a few examples is not sufficient to prove the statement. General applicability is mentioned frequently in the students' comments and can be seen as a significant criterion for students to accept a proof.
- On the other hand examples play an important role in mathematical argumentation for students. Even after accepting the correctness of an argument they are not convinced until it is shown with a few examples. Even Barry's proposed list of examples is considered by two students as suitable to convince a *"friend"* of the truth of the statement. These students appreciate attempts to convince the reader in their evaluations.
- Another indicator that convincingness of a proof is important to the students is the fact that the quality of explanations plays a role in their proof evaluations. It seems that if an answer does not show attempts to explain the argument, most of the students reduce marks.
- To the students, a valuable proof should have a certain structure, starting with a definition, followed by some algebraic equations or formulas, and finishing

with one or more examples.

- Structure and formalism seem to be more important to the students than the idea behind a proof. If their expectations are satisfied in this regard most of the students give at least a few marks regardless of the correctness of the particular steps or whether the whole idea makes sense to them or not. It appears that a good idea to prove a statement is not being valued as highly as the structure and formalism of a proof.

- Internal correctness clearly is an important criterion for about a third of the participating students. Eleven students (29%) notice that Eve's approach is incorrect and acknowledge that by giving her zero or one marks. Occasionally correctness of Aoife's or Cathy's answers are approved in the students' comments.

  Questions regarding students' approaches towards internal correctness of a proof, i.e. how students decide whether an argument is correct and how important the internal correctness is for their proof evaluations, are considered in the design of the interviews, see Section 8.1.2.

Listed next are criteria which appeared less often.

- 'Clearness' is mentioned both approvingly or disapprovingly several times in the students' comments. Precise argumentation seems to be important to some of the students.

- Some students appreciate sufficiency without excess in a proof.

- Occasionally students criticize if assertions are stated without justification.

None of the participating students appreciates the fact that Cathy's proof places the statement in a wider mathematical context. Embedding a mathematical fact in a broader context does not seem to be a considerable criterion to value a proof highly for students.

To summarize, most of the students recognize the difference between a demonstration by example and a proof as well as the importance of mathematical definitions, general applicability, internal correctness and convincing mathematical arguments. Precise argumentation, sufficiency without excess and justification of significant assertions in a proof are important to a relatively small proportion of the students. On the negative side the students' formulaic and inflexible picture of valuable proofs have to be noted. Structure and 'mathematical-looking' formalism seem more important to some students than the ideas comprising the argument.

7.3.4 Findings of the written experiment in relation to the schema to interpret student evaluations

The written evaluation task had been analysed before the schema to interpret student evaluations was developed. However, the findings can be connected to the suggested schema. Some of the students' comments indicate how they relate the three facets actual character, intended character and purpose(s) of a proposed proof. Also, the identified evaluation criteria indicate which purposes of proofs the participating students consider relevant. Both issues are discussed in this section.

Do the students relate the three facets of artifacts in their proof evaluations?

Evidence can be found for different attitudes of the students towards the relationships E_{AI}, E_{IP} and E_{AP} among the actual and intended character and the purpose(s) of these proofs.

- Many students show some attempts to decide whether a proposed proof is internally correct or not. This indicates that they do consider the relationship E_{AI} between the actual character of the proof and its assumed intention. These students seem to appreciate internal correctness by giving high marks. Most of those students who identify algebraic incorrectness in Eve's answer acknowledge this by giving low marks.

  Some students seem to be content with considering only E_{AI}. Their comments do not show attempts to consider purpose(s) of proof. For example one student gives Eve's (incorrect) answer most marks and comments:

  > *"Maybe an extra line stating that $2z$ is divisible by 2, therefore it is even would have made the answer better."*

- However, many of the students who show attempts to decide whether a proposed proof is internally correct or not are not satisfied only with their evaluations regarding internal correctness. It is not unusual that students also comment on the quality of explanation or examples or the general appearance of the proposed proof. The following excerpts can be interpreted as examples of evaluations in which the particular proofs are considered sufficient regarding E_{AI}, but not fully accomplished regarding E_{IP} and E_{AP}. Thus, these students appear to consider all three relations E_{AI}, E_{IP} and E_{AP}.

  - Some comments on Aoife's answer:

    > *"This is valid and in proper form"*,
    > *"Very good but would like a practical as well as a conceptual answer"*,
    > *"Aoife has proved the statement is true, I would just advice her to give an example to finish the proof"*.

  - A comment on Dan's answer:

    > *"It should be proven formally even if it is intuitively true"*.

  - A comment on Finn's answer:

> *"Intuitively correct but needs to explain why the answer means the statement is true"*.

- The observation that several students consider internal correctness in a proposed proof is not the only indication that the relation E_{AP} is considered by some of the students. Some of the comments on Finn's answers can be interpreted regarding this relation. For example the comments

  > *"Good idea, give explanation"*,
  > *"Good visual proof but* [doesn't] *use mathematics"*,
  > *"Good visual representation but needs notational explanation"*

  can be interpreted as complaints that Finn's idea, while the students consider that it is good, is not adequately manifested in his proof. Thus the actual character of Finn's proof does not meet its purposes and Finn's answer is not fully satisfying regarding E_{AP}. On the other hand, these students consider Finn's idea as good for proving the statement, and are therefore relating the idea with the purpose *verification* of proof. Thus they are considering Finn's proof as sufficient regarding relation E_{IP}.

- Section 7.3.2 describes the fact that structure and 'mathematical-looking' formalism seem more important to some students than the ideas comprising the argument. These students do focus on the actual character of a proof but in a way that seems to be based on preconceived expectations rather than on relating the actual character to intention or purpose. Thus a proof can be criticized or rejected because its actual character *"looks wrong"* rather than on the basis of relating the actual character to the purposes of proof, or to the intention.

  > *"Correct answer but show mathematically"*,
  > *"Although Cathy's answer is true there is too much English and does not mathematically prove it unlike Aoife"*,
  > *"This proof makes sense but she could have used a more mathematical approach"*

  These students do not seem to attend to the relations E_{AI}, E_{IP} and E_{AP} in their proof evaluations. One student for example gives Eve's incorrect answer full marks because it is in his/her opinion *"well worked out"*. Another student gives Finn's visual approach zero marks, explaining:

  > *"This does not prove anything, words and numbers are needed."*

To summarize, the students' comments provide some evidence that the relations E_{AI}, E_{IP} and E_{AP} are considered in students' proof evaluations. Some students seem content only with relating the actual proof to its intention, i.e. with consideration of E_{AI}. They consider the relation of a proof to an assumed intention more important than its relation to purposes of proof. Other students appear to consider all three relations in their evaluations. Nevertheless, it has to be noted that a significant number of students do not seem to address the relations at all. The comments of these students indicate that they do not relate the actual proposed proof to an assumed intention or to any purpose of proof.

Indications from the written experiment about purposes of proofs in the view of first year students

Investigation of the students' evaluation criteria allows some conclusions regarding the students' views of purpose(s) of proof. Most of the students comment on what they approve or disapprove in a particular proposed proof. Some of these comments focus on the relation between actual and intended character. However, a significant number of the comments suggest that the students do consider some purposes of proof in relation to actual and intended character of the proposed proof when evaluating it. In the following the students' views of purposes of proofs are considered in relation to De Villiers' model of functions of proofs outlined in Section 3.2.

- The students highly value general applicability in a proof to demonstrate the veracity of the proof as well as the statement. This indicates that the purposes of proof named above as *verification* (concerned with the truth of a statement) and *explanation* (providing insight into why it is true) do play a significant role in the students' proof validations and evaluations.
- The students' appreciation of explanation in a proof indicates that *explanation* and *communication* (the transmission of mathematical knowledge) are considered significant purposes of proofs by some of the students.
- De Villiers suggests *systematization* as one purpose of proof, i.e the organization of various results into a deductive system of axioms, major concepts and theorems. *Systematization* does not appear to be considered by the students. None of the students seems to relate the statement to its mathematical background. In particular, none of them indicate appreciation of the fact that Cathy's proof places the statement in a broader mathematical context.
- Other purposes of proofs which are mostly concerned with enhancement of mathematical knowledge such as *discovery* of new results, *intellectual challenge* (the *self-realization/ fulfillment* derived from constructing a proof) or *problem solving competencies* are not considered in the students' evaluations.
- As mentioned in Section 3.2, De Villiers does not claim his list to be exclusive, and a number of expansions have been suggested. The students' evaluations in this written exercises do not indicate that they consider other purposes of proof than those listed by De Villiers.

7.4 Outcome of the written evaluation experiment

The first year students' comments and suggested marks on six proposed 'proofs' of a mathematical statement provide some understanding about their **criteria when**

evaluating mathematical arguments. Most of the students recognize the difference between a demonstration by example and a proof as well as the importance of mathematical definitions, general applicability, internal correctness and convincing mathematical arguments. Precise argumentation, sufficiency without excess and justification of significant assertions in a proof are important to a relatively small proportion of the students. On the negative side the students' formulaic and inflexible picture of valuable proofs must be noted. Structure and 'mathematical-looking' formalism seem more important to some students than the ideas comprising the argument.

Relating these findings to the suggested schema to interpret student evaluations motivates observations regarding some of the research questions as suggested in Figure 4.2.

- **Do the students relate the three facets of the proof, that is, do they consider the relations E_{AI}, E_{IP} and E_{AP}?**
  The students' comments provide some evidence that the relations E_{AI}, E_{IP} and E_{AP} are considered in students' proof evaluations. Some students seem content only with relating the actual proof to its intention E_{AI}, others appear to consider all three relations in their evaluations. Nevertheless, it has to be noted that a significant number of students do not seem to address the relations at all. The comments of these students indicate that they do not relate the actual proposed proof to an assumed intention or to any purpose of proof.
- **What are the purposes of proof for the student?**
  For the students in this written experiment three purposes of mathematical proofs are predominant. Most of the students consider *verification* as a purpose of proof. A significant proportion of the students regard *explanation* and *communication* as purposes of proofs. Other purposes, in particular those described by De Villiers do not appear to be purposes of proof in the views of the participating students.

The written responses of the students provide less information on other research questions suggested in Figure 4.2, for example on how the students read or validate mathematical proofs, or on learning effects during the process of proof evaluation, e.g. whether the purpose(s) of proof *become visible* to the student evaluators.

A wide range of criteria to value a proof have been observed in the students' evaluations. However, most of these criteria appear singularly. While a range of considerations is visible to a reader of the full collection of written evaluations by students, it is unlikely that the full range is considered by many individual students. Also, evaluation habits appear to differ considerably among the students. Both of these issues raise the question of whether practice and discussion of proof evaluation in groups would be particularly useful for enhancing students' ability to evaluate proofs and their knowledge about proof and proof evaluation.

The analysis of the written evaluations leaves some uncertainties and open questions, which influenced the design of the interviews. For example, the students often focus on criticism of the proposed proofs in their written responses to the evaluation exercise. Information about what they approve in a proof is rare. Also it is occasionally difficult to interpret the students' considerations of the proposed proofs, in particular in the case of the one based on visual reasoning. One aim of the interviews was to achieve a clearer understanding of the students' views on mathematical proofs. Other questions arising from the analysis above are concerned with the students' views of the role of examples in the context of proofs, whether students associate mathematical proof with algebraic formulas, or how students assess the internal correctness of a proposed proof, and what overall role this assessment plays in their evaluation of proofs. These questions are considered in the design of the interviews.

The next chapter is concerned with the design, implementation and findings of the interviews held in Spring 2009.

Chapter 8

Interviews

In March 2009 interviews were held with eight students who had attended the written exercise in September as well. Eighteen students were invited to participate in a research project. The students 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.

Section 8.1 describes the chosen methodology and the design of the interviews.

Section 8.2 introduces each of the interview participants individually.

Section 8.3 describes how the interviews have been transcribed, analysed and interpreted. This part also includes the transcripts themselves.

Section 8.4 summarizes the findings of the interviews.

8.1 Design and transcription of the interviews

The interview meetings consisted of two parts, beginning with informal conversations about the students' experience with studying mathematics both in secondary school and university.

The informal interview style was chosen to allow the student to accustom him/herself with the interview situation and to acquire some information about each particular student's mathematical background and study habits. Another aim of this interview part was to create a relaxed atmosphere between interviewee and researcher before the students were introduced to the second interview part which involved some mathematical activities. The informal conversations were followed by more structured dialogues. The students were presented with two mathematical statements and a number of proposed (and partly incorrect) proofs. They were asked to evaluate and rank the proofs.

Section 8.1.1 describes the chosen interview style.

Section 8.1.2 lists a number of themes which were intended to be explored.

Section 8.1.3 presents the two statements and the various proposed proofs.

Section 8.1.4 describes how the interviews were recorded.

Section 8.1.5 describes how the acquired data were transcribed.

8.1.1 Methodology

In a previously conducted pilot project meetings consisted of a number of focus groups with two students and the researcher. The aim was to acquire a picture of the students' views about what they considered important in a mathematical proof. The researcher's role in those meetings was that of an observer who avoids any participation in the discussions. The project partly failed because the participating students did not get involved in discussions about mathematical topics without being prompted by the researcher. When talking about mathematics they addressed the researcher, not their peer students. Therefore I decided for the main study to meet students on a one to one basis to be able to encourage discourse about mathematics. To allow the discourse to be as open as possible I chose to hold *semi-structured interviews*, which are rather flexible in comparison to more formalized *structured interviews*. Semi-structured interviews are guided conversations where broad questions are asked, which do not constrain the conversation, and new questions are allowed to arise as a result of the discussion. This is different from questionnaires and surveys which involve very structured questions from which no deviation is possible. A semi-structured interview is therefore a relatively informal, relaxed discussion based around a predetermined topic. The set of questions is prepared but open, allowing the interviewees to express opinions through discussion. Questions are generally simple, with a logical sequence to help the discussion flow. The interviewer has a framework of prepared themes within which the interview may be guided in the required direction.

8.1.2 Framework of Themes

The list below was written in preparation for the interviews of this study. The various items describe topics to be explored:

- What makes a proof valuable to the student?
- The process of validation:
  - Does the student try to understand the arguments? If so, how does he/she attempt this?
  - Does the student introduce examples in order to verify arguments?
  - Does the student write notes while reading 'proofs' ?
  - Does the student change his/her mind during the meetings?

  - Do the nature and the purposes of proof become *visible* to the student during the process of proof validation?
- How quickly does the student decide whether an argument is correct?
- What does the student consider to be the role of examples in mathematical proof, for instance to enhance understanding or as a (necessary) part of the proof?
- Does the student value formal appearance highly in a proof and if so, why?
- How does the student respond to visual representations of proofs?
  - Does the student accept visual representations as proofs?
  - Could visual representations help to direct the attention of the student to the content of proof, rather than to the syntactic structure of sentences?[1]
- What criteria are used in the ranking or evaluation process?

It should be emphasized at this point, that this framework of themes was used in the preparation of the interviews, but less in their analysis as the research focus shifted. During the processes of transcribing, editing and analysing the research focus and questions changed; some of the topics were neglected while further topics were added as new questions arose in the interpreting process.

8.1.3 Question Design

Preparation of the interviews

Either *interview schedules* or *interview guides* are typically used to prepare for interviews. An *interview schedule* is more formal and mostly used for fully structured interviews, when a project requires uniformity in wording, so that all interviewees hear roughly the same questions. An *interview guide* is an informal grouping of topics and questions that the interviewer can ask in different ways for different participants. Interview guides help researchers to focus an interview on the topics at hand without constraining them to a particular format. This freedom can help interviewers to tailor their questions to the interview context/situation, and to the people they are interviewing. *Interview guides* "allow the interviewer to go down an unexpected conversational path, ask optional questions, or to adjust to the verbal style of the participant" (Lindlof & Taylor 2002, p. 195). In the preparation for my interviews I developed an **interview sketch**, which consists of elements of both, an *interview schedule* and an *interview guide*. The interview sketch can be found in Appendix D.

In the first part of the interviews I planned to ask the students some background questions. The main purposes of the background questions, in addition to getting some contextual data, was to set a relaxed atmosphere for the remainder of the

[1]Hanna and Sidoli (1998, p. 77) raised this question.

interview. This part was intended to be semi-structured, the guide consists of a few suggested questions:

- *When and where did you do your Leaving Cert? Pass or Honours?*
- *Did you like maths in school?*
- *Were you good at maths in school?*
- *Your first year in college is almost finished. How did you like it?*
- *Studying maths at university: what is different from school mathematics?*

In the second and main part of the interviews, the students were presented with several (variously flawed) 'proofs' of two mathematical statements. Interview participants were asked to comment on each of the 'proofs' and to rank them. Each of the students was presented with the statements and 'proofs' in the same order, therefore that part of the interview was planned in form of an interview schedule. Within this schedule the interviews were prepared less formally. Again, a number of questions were suggested, and the discussions about each of the 'proofs' were held in a semi-structural form.

- *What do you think of this answer?*
- *Does the answer prove the statement?*
- *What do you like/ do you not like about this answer?*
- *What advice would you give [proof author's name]? How could he/she improve the answer?*

The students had a few minutes to consider *Statement I* by themselve before they were presented with the first of the prepared 'proofs'. Due to time issues they proceeded to the first of the prepared 'proofs' of *Statement II* immediately after seeing the statement.[2]

Interview Tasks

I will now present the two statements and the various proposed proofs that the interview participants were asked to evaluate. Each of the proposed proofs is followed by a reflection on how a representative *Experienced Evaluator* might evaluate it[3] and

[2] I held the first interview according to plan, and therefore gave the student a few minutes to consider *Statement II* for herself. In addition to the fact that it took significantly more time than anticipated, I realized that the student was not prepared to comprehend the prepared answers as she was still reflecting about her own (wrong) approach to prove the statement. This experience led to the decision to omit the task to consider *Statement II* for themselves in the following interviews.

[3] As in case of the written evaluations tasks the view of a representative *Experienced Evaluator* has been developed in cooperation with Rachel Quinlan. Discussions in a departmental seminar and the survey about professional mathematicians' views of visual reasoning suggest that their evaluation habits and criteria might differ considerably. In the comments of the *Experienced Evaluator* we try to reflect all the different views that were expressed by these experienced mathematicians, even though they were not all fully consistent with each other.

occasionally by considerations on why the particular 'proof' was included in this task. Each of the *Experienced Evaluator*'s comments consists of a direct reaction to the proposed proof, followed by considerations of how this reaction fits into the schema as suggested in Section 4.2. In the interviews the students were presented with each of the proposed proofs seperately, all in the same order. A list of the tasks without any comments is included Appendix E.

Task I.

Consider the following *Statement I*:
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.

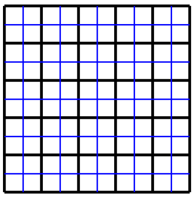
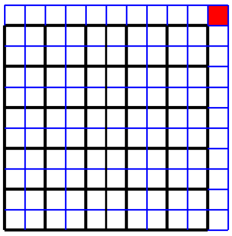
***Experienced Evaluator*'s view of *Anna's* answer.**
Anna's argument centres on her assertion that the last digit of the square of an integer is determined by the last digit of that integer itself. This assertion is correct. It certainly could be argued that the assertion needs some justification. 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 E_{AI}. 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 right explanation. 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 sufficient concerning E_{IP} and E_{AP}.

***Benny*'s answer:**

$2^2 = 4$ even	$3^2 = 9$ odd
$4^2 = 6$ even	$5^2 = 25$ odd
$6^2 = 8$ even	$7^2 = 49$ odd
$20^2 = 400$ even	$23^2 = 529$ odd
$(-16)^2 = 256$ even	$(-19)^2 = 361$ odd

The squares of even numbers are even, and the squares of odd numbers are odd.

***Experienced Evaluator*'s view of *Benny's* answer.**
Benny's answer consists of a collection of examples. While it may be considered by some to be a reasonably convincing demonstration of the fact that squaring preserves parity, the character of this answer does not address the purposes of proving it generally or of explaining why this is true. Therefore Benny's argument is not sufficient regarding E_{IP} and E_{AP}. An experienced evaluator would reject Benny's answer as a proof on these grounds. It might also be noted that Benny's answer is not fully satisfactory in terms of the relation E_{AI} either, because of two errors in his list of examples.
Considerations on the choice of 'proof'.
Benny's answer is similar to Barry's in the written evaluation task (p. 48). Students' responses to Barry's approach indicated that some students consider providing evidence in support of the truth of the statement as a major purpose of proof. Benny's similar approach in form of a collection of examples was included in the interview task to investigate this observation further.

***Ciara*'s answer:**
Square of an even number:

Square of an odd number:


***Experienced Evaluator*'s view of *Ciara's* answer.**
Ciara's answer relies on the idea of representing the square of the natural number n as a $n \times n$ square grid, in which each row and each column consists of n squares. The number of small squares in the whole grid is n^2. Ciara presents two diagrams of this sort, one each for the square of an even number and the square of an odd number. By highlighting alternate horizontal and vertical lines, Ciara enables the reader to gather the small squares in her diagrams into pairs or groups of four. In the example of the square of an even number, these groupings account for all the squares. In the 'odd' example, one small square that is 'left over' is highlighted.

As in the case of Finn's answer in the written exercise (p. 50), very little explanation is provided and it is left to the reader to consider whether Ciara's answer consists only of a pair of examples or whether it can be construed as a general argument. The text 'square of an even number' and 'square of an odd number' suggests that Ciara intends for her diagrams to be interpreted as demonstrating general phenomena as well as specific examples – thus an evaluator would probably consider that Ciara's intention at least takes into account the purpose of providing an explanation of the truth of the statement. However, an evaluator may take the view that Ciara does not provide an argument for moving from her specific examples to the general case, and that the actual character of her proof is not sufficiently developed to meet the purpose(s) of proof. Such an evaluator might conclude that Ciara's proof really only amounts to a pair of examples, but would still be likely to rate Ciara's answer more highly than Benny's in terms of the relation E_{AP}, since Ciara's consists of an attempt to demonstrate **why** 8^2 is even and 9^2 is odd, rather than just observing that $8^2 = 64$ and $9^2 = 81$. However, an evaluator who considers that Ciara's answer only refers to the numbers 8^2 and 9^2 must consider that as a proof her answer is not adequate regarding relations E_{IP} and E_{AP}.

On the other hand an evaluator may consider that Ciara's examples are of a generic quality, and accept that Ciara's diagrams show why the square of an even number will always be a multiple of 2, and why the square of an odd number will always leave a remainder of 1 on division by 2. Such an evaluator may be basically prepared to accept Ciara's answer as a correct proof of the statement, but might still complain that further explanation of the argument would have been helpful, and that the actual character of Ciara's proof is not ideally matched to purposes such as *explanation* and *communication*.

An evaluator might observe that Ciara's diagrams suggest the stronger statement that the square of an even number is always a multiple of 4 and that the square of an odd number always leaves a remainder of 1 on division by 4, although there is no evidence that Ciara has noticed this feature of her own argument. However Ciara's answer could be seen as exemplary of such purposes of proof as *systematization* and *discovery*.

***Darragh*'s answer:**
k is any whole number.
l is any whole number.
$2k$ is an even number.
$2l + 1$ is an odd number.
$(2k)^2 = 4k^2 = 2(2k^2)$ is even.
$(2l+1)^2 = 4l^2+2l+1 = 2(2l^2+l)+1$ is odd.
So the statement is true.

***Experienced Evaluator*'s view of *Darragh's* answer.**
Darragh's answer is an essentially correct proof of the statement. His idea is appropriate for the purpose of explaining why the statement is true; thus his intention is well matched to this purpose of proof. Darragh's answer could be criticized on the grounds that (as Aoife does in the examples from the written exercise, p. 47) he introduces his 'general' even and odd numbers in an awkward and logically oblique manner. Also, his arithmetic manipulation includes an error, which only by chance does not impede his argument. Thus the actual character of Darragh's argument is not a perfect realization of his intention, nor is it perfectly matched to his purposes.

However these deficiencies can be rectified.

***Elaine*'s answer:**
Proof: Let x be an even number and let y be an odd number.
Then $x + y$ and $x - y$ are both odd.

$$(x+y)^2 = x^2 + 2xy + y^2, \qquad (x-y)^2 = x^2 - 2xy + y^2$$

So $(x+y)^2 + (x-y)^2 = 2x^2 + 2y^2 = 2(x^2 + y^2)$.
The number $2(x^2 + y^2)$ is even obviously. Dividing by 2 we find that $x^2 + y^2$ is odd.
Then x^2 and y^2 can't be both even or both odd. So x^2 is even and y^2 is odd.

***Experienced Evaluator*'s view of *Elaine's* answer.**
Elaine's answer is wrong. It begins by appropriately introducing an even and an odd number, then takes the surprising step of focussing on the squares of the sum and difference of these numbers instead of on the squares of the numbers themselves as a reader might expect. Elaine's intention in introducing this sum and difference seems to rest on the incorrect claim that when an even number is divided by 2, the result is odd. This erroneous assumption undermines Elaine's argument, and her intention is not suitable for establishing the truth of the statement. Her conclusion in the last sentence is not justified and an evaluator may suspect that at this point she is in fact assuming the statement that she is intending to prove. Neither Elaine's intention nor the actual character of her proof are connected to any purpose of proof. Because her intention is obscure and seems to be based on an erroneous assumption or belief, it is difficult to judge her answer on the basis of the relation E_{AI}.

***Fintan*'s answer:**
Let $2a$ and $2b$ be any even numbers.
$(2a) * (2b) = 4ab = 2(2ab)$ is even.
Therefore the product of two even numbers is even, and so in particular the square of an even number is even as well.

Let $(2a + 1)$ and $(2b + 1)$ be any odd numbers.
$(2a + 1) * (2b + 1) = 4ab + 2a + 2b + 1 = 2(2ab + a + b) + 1$ is odd.
Therefore the product of two odd numbers is odd, and so in particular the square of an odd number is odd as well.

***Experienced Evaluator*'s view of *Fintan's* answer.**
Fintan distinguishes between the two cases, even and odd numbers. For each he uses a characterizing property, namely the fact that even numbers can be represented as $2k$ and odd numbers as $2k+1$ with k being an integer. He then shows more general facts (*any product of two even numbers is even* and *any product of two odd numbers is odd*) using algebraic manipulations and the formerly noted properties of even or odd numbers. He finally states for both cases that as squares are special cases of products the statement is true.

Fintan's answer is a correct and succinct proof of the fact that the product of two even integers is always even, and the product of two odd numbers is always odd. The intention is clear to the reader and fulfills more than one purpose of proof, not only explaining why the statement is true but also fitting it into a wider mathematical context. Therefore Fintan's argument achieves more than any of the others. The actual proof is correct and in line with the intended proof. Regarding all three relations E_{AI}, E_{IP} and E_{AP} this proof is sufficient.

Ciara's is the only of the other of these arguments that even hints at the extra generality that is explicitly described in Fintan's proof. Students' evaluations of these proofs may allow indications of how they consider the purposes of proof *systematization* and *discovery*.

Task II.

Let f be a quadratic function, $f(x) = ax^2 + bx + c$ with $a, b, c \in \mathbb{R}$ and $a > 0$. Show *Statement II*: f can't have more than two common values with its derivative f'.

***Gerard*'s answer:**
$f(x) = 2x^2 + 3x - 7$
$f'(x) = 4x + 3$

$2x^2 + 3x - 7 = 4x + 3$
$2x^2 - x - 10 = 0$
$(2x - 5)(x + 2) = 0$
$x = \frac{5}{2}$ or $x = -2$ $\longrightarrow$ 2 solutions
So the statement is true.

***Experienced Evaluator*'s view of *Gerard's* answer.**
Gerard demonstrates the truth of the statement for a particular choice of f matching the description in the statement. An experienced evaluator might rate Gerard's answer more highly than Benny's collection of examples (p. 79), because his analysis of his chosen example does contain an argument that suggests a reason for the truth of the statement. One could interpret Gerard's proof as an example of a *generic proof*[4] as suggested by Steiner (1978) and argue further that Gerard's argument has a 'proof idea' whereas Benny's does not. However there is no indication that Gerard intends his example to be considered by the reader as having a generic quality, or that he is really aware himself of the potentially generalizable nature of what he has done. The intended character of Gerard's proof, while well realized in its actual character, is not general and as such it is not well matched to the purpose of proof. Therefore Gerard's answer can be evaluated as sufficient regarding E_{AI} and not sufficient, though better than Benny's, regarding E_{IP} and E_{AP}.

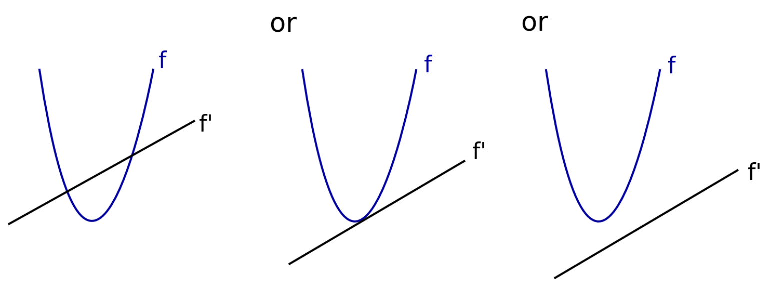
***Helena*'s answer:**
The derivative of a quadratic function is a linear expression. To find common points I solve the equation I get from setting both expressions equal. Setting a quadratic expression equal to a linear expression gives a quadratic equation which can't have more than two solutions.

***Experienced Evaluator*'s view of *Helena's* answer.**
Helena provides a purely verbal explanation of the truth of the statement. Each argument is mathematically correct, the intended proof is successfully implemented in its actual version and regarding E_{AI} the answer is very good. Considering E_{IP}

[4] Alibert and Thomas (1991, p. 216) explain that a *generic proof* works at the example level but is generic in that the examples chosen are typical of the whole class of examples and hence the proof is generalizable. Steiner (1978) suggests the concept of generic proof as students' first acquaintance with proof at university level.

and E_{AP} the proof is sufficient as the truth of the statement is established. However, an evaluator could argue that some claims should be justified – for example the assertion that the derivative of a quadratic expression is linear, or that a quadratic equation cannot have more than two solutions. Depending on the context, an evaluator may feel that some justification should be provided for these assertions and therefore evaluate this answer slightly less favourably.

Ian's answer:



***Experienced Evaluator*'s view of *Ian's* answer.**
Ian's answer represents the quadratic function f and its derivative f' graphically, namely f as parabola and f' as a line. He is using the fact that the derivative of a quadratic function is always a linear function. Ian presents three alternatives possible: they intersect twice, once or not at all - in other words they have at most two common values, which was required to be proved.

On the positive side, there are no mistakes in the implementation of Ian's proof. An evaluator might approve that the truth of the statement is established generally and therefore regard Ian's answer as sufficient concerning relation E_{IP}. On the other side an evaluator might criticize this answer for several reasons. First, the internal arguments are neither articulated nor justified. For example an evaluator might criticize that the key ingredients here are to note explicitly that $a > 0$ and that f' is linear, so the only possibilities reduce to the three cases shown. Second, this approach might not convince a reader without any explanations of the idea behind it. Thus, Ian's answer might be seen as not fully sufficient regarding the relations E_{AI} and E_{AP}.

Note: Acceptability of a diagram as a proof seems to depend on the confidence or expertise of the reader to be able to interpret and articulate the argument implied by the diagram.

***Joan*'s answer:**
Proof: $f(x) = ax^2 + bx + c$, so $f'(x) = 2ax + b$. We want solutions to $f(x) = f'(x)$, i.e.

$$ax^2 + bx + c = 2ax + b.$$

Multiplying by x gives

$$\begin{aligned} ax^2 + bx + c &= 2ax^2 + bx \\ \Longrightarrow ax^2 &= c \\ \Longrightarrow x^2 &= \frac{c}{a} \end{aligned}$$

So the only two possible solutions are given by $x = \pm\sqrt{\dfrac{c}{a}}$.

***Experienced Evaluator*'s view of *Joan's* answer.**
Joan's proof is wrong – when she multiplies by x at line 4 she does so only on one side of her equation, leading her to an incorrect conclusion about the values of x at which $f(x) = f'(x)$ although (somewhat luckily) to the correct conclusion about the number of possible common values. Nevertheless, Joan's strategy is suitable for the purpose of explaining the truth of the statement, she correctly sets up the equation $f(x) = f'(x)$ and successfully conveys to the reader her intention of investigating the solutions of this equation. Her error follows and so her intention is not successfully implemented in the actual character of her proof. However Joan's intention does match her purpose and her proof is satisfactory regarding relation E_{IP}. Because of the error in implementation it is unsatisfactory regarding relations E_{AI} and E_{AP}.

Despite the fact that her error leads Joan ultimately to solve a quadratic equation that has a particularly simple form, an evaluator may feel that, by explicitly writing down the solutions, Joan has more convincingly justified the claim that her equation can have at most two solutions than Helena or Kieran.

***Kieran*'s answer:**
$f'(x) = 2ax + b$
$f(x) = f'(x)$
$\Leftrightarrow ax^2 + bx + c = 2ax + b$
$\Leftrightarrow ax^2 + (b - 2a)x + c - b = 0$
is a quadratic equation which has at most two solutions.

***Experienced Evaluator*'s view of *Kieran's* answer.**
Actual and intended proof agree, the implementation is correct, so related to E_{AI} this answer is good. Considering E_{IP} and E_{AP} the proof is sufficient as the truth of the statement is established.

8.1.4 Recording data

The data were recorded digitally by a small digital voice recorder. Recording the interviews enabled the researcher to capture the interview more or less exactly as it was spoken. Even though the dictaphone was always set in plain view of the participants, they seemed to forget its presence after a while. A video camera might have been more disruptive. The prepared tasks and 'proofs' were presented on plain sheets leaving plenty of space for notes, if needed by the students. Those sheets including notes were kept by the researcher.

8.1.5 Transcribing Interviews

Lindlof and Taylor (2002, p. 205) describe transcribing as "a portal to the process of data analysis", which captures my own experience very well. Even though the whole process is time consuming, listening several times while focussing on various research aspects helped me to identify patterns or interesting views of the students. Some of the students' comments which were interesting for me as a researcher might not have caught the attention of a professional transcriber. Some phrases to which I listened several times would surely have been passed over, had I left the transcribing to an agency.[5] In addition to the fact that professional transcribers would not listen with the researcher's questions in mind, they would need a certain mathematical background to understand some of the students' comments.

In the editing process a major aim was to find the right balance between keeping the transcripts readable and efficient for analysis and the necessity of completeness. The determination of what level of editing is appropriate can only be made by the researcher and has been made in this case during the process of transcribing. Full transcripts carry the danger of being difficult to read: they can be long, repetitive or confusing. Lindlof and Taylor (2002, p. 206) quote DeVault, who claims that "the researcher should include only as much detail as needed to illustrate the analytic points to be made".

To keep the transcripts manageable I structured the interviews, wrote summaries of each interview phase, then gradually transcribed fully those excerpts that I used for the analysis. 'Fillers' like *"Ahm"*, *"like"* and *"you know"* where mostly kept, as they might indicate that a student took a moment to reflect on something or to formulate some comment. On the other hand grammatical errors have been corrected to keep the transcripts readable. Contents were not affected by these changes.

[5] Occasionally phrases were so difficult to understand exactly that I had to listen to them more than ten times.

The following notation set was used in the transcripts of these interviews.

—	A short pause (a few seconds).
[Pause]	The pause lasts more than 30 seconds.
[Long Pause]	The pause lasts a few minutes.
(...)	I left something out.
(...?)	I left something out, because I was unable to identify the words from the audio recording.
[ ]	An explanatory insertion.

8.2 The Participants

This chapter introduces each of the interview participants individually and summarizes the overall impressions and the contents from the first parts of the interviews. To avoid confusion with the names of the fictitious authors of the proofs, the participating students are refered to by code names.

1. ***Student A*** is not originally Irish. She has been living in Ireland for a few years and completed her secondary education in year 2008 in a Secondary School in Ireland. She is studying Honours Mathematics and likes mathematics, as she *"like*[s] *the challenge"*. At University she finds mathematics still challenging, but more difficult, for example in Analysis: *"you have to go through epsilon and proof and all that,* (...) *it's so confusing."* *Student A* works hard, attends lectures and tutorials and visits the maths support center based in the School of Mathematics, Statistics and Applied Mathematics regularly. She describes her study habits as follows: she does not use books, goes to 'Google' or lecture notes and practices exam papers. Even though she struggles in mathematics, *Student A* appears confident and has strong opinions. Generally *Student A* is relatively quick in making decisions.

2. ***Student B*** completed her secondary education in Germany. She is a very confident and capable student. She sees herself at an advantage compared to the Irish students, as the mathematical contents of the first semester was covered in her school mathematics. *"The first semester was pure revision."* (translated from German: *"Das erste Semester war reine Wiederholung."*) [The interview was held in German.]

3. ***Student C*** and ***Student D*** were interviewed together.[6] *Student C* was a lot quicker than *Student D* in finding her own answer and understanding the proposed proofs. She then explained to *Student D*, who, on her own, might have answered differently. Both students completed their secondary education 2008 in Ireland. They are now in the Mathematics and Education Programme at NUI Galway which will qualify them as second level teachers. During the meeting they repeatedly stress their opinion that they will never need or use

[6] *Student C* arrived at the same time as *Student D*, not aware that her own meeting was scheduled a week later. I did not want to reject her and held the meeting with the two students together.

all the mathematics they have to learn at University. They believe that the mathematics up to Leaving Certificate level is all they need to know to proceed with a teaching career.

4. ***Student E*** is a confident student. He has been home schooled and in the informal interview part he engages in explaining reasons and advantages of home schooling. He stresses that he enjoys mathematics and wants to continue studying the subject. *Student E* understands the ideas of the prepared proofs quickly and is able to compare them. He does not struggle with the visual representation, which seems to be unusual within the group.

5. ***Student F*** completed his secondary education 2008 in Ireland, did Honours Mathematics in school and is interested in mathematics. *Student F* is a bit shy and quiet, mathematically not very confident. During the meeting, he appears to be a bit nervous, which might make it difficult for him to concentrate. Several times he apologizes for not being good or professional enough. In comparison to school *Student F* finds university *"harder"*. He describes the difference in his study habits: *"The only way you can really study maths is by doing questions which is the same. Suppose in college, you have to know definitions. That's different. But apart from that* (...)*, I think it's the same."* *Student F* explains why he enjoys maths at university: *"I like the challenge — but it's a good challenge, like, it's hard."*

6. ***Student G*** completed her secondary education in 2008 in Ireland and is now enrolled in the Mathematics and Education Programme at NUI Galway. She does not consider herself as extremely good at mathematics but stresses that she did *"well enough"* in her Leaving Certificate Exam. *Student G* finds honours mathematics at university *"a lot more difficult"* in comparison to school mathematics and does not understand why it is compulsory. To her it would make more sense to do Pass level as *"it's more like Leaving Cert mathematics."* *Student G* worries that she might forget the Leaving Cert mathematics during her years in university. During the meeting she says several times, that she is not good at proving, formulating ideas or *"coming up with ideas"*. When evaluating the proposed proofs she keeps referring to what she has been told in preparation for her Leaving Certificate, which seems to have had an immense effect on her views regarding mathematics. At the end of the interview, *Student G* says that this meeting has changed her mind about the need to do more than Leaving Certificate mathematics.

7. ***Student H*** completed her secondary education 2008 in Ireland, did Honours Maths and *"loved it"* as she liked *"the challenge* (...)*"*: *"You might not get it at first, and then you just get it. — You think you can't get it, and then you get it —."* In general *Student H* appears to be a confident student, who is able to articulate her thoughts. She *"loves the subject"* mathematics, as she stresses several times during the meeting, and *"would love to be a maths teacher"*. However, she considers choosing other subjects as she finds mathematics too time consuming.

8.3 Interpretation of transcripts

This section is concerned with the analysis and interpretation of the interview transcripts.

Section 8.3.1 includes reflections about the interpretation process in the context of the development of the research focus.

Section 8.3.2 presents excerpts from transcripts of the students' evaluations of the proposed proofs, a categorizing analysis of these transcripts and interpretations regarding the research questions of this study.

8.3.1 Methodology

The first considerations after the interviews were to some extent discouraging as each of the participants seemed to have an individual approach to comparing the given 'proofs'. Some of them articulated their thoughts clearly, but the majority struggled in explaining why they evaluated a particular 'proof' in the chosen way. The amount of data was difficult to comprehend. Even after repeated listening to the recordings, a clear thread regarding evaluation skills and habits was difficult to identify and describe. Some doubts began to arise: were the acquired data suitable for addressing the research questions of the prepared framework of topics (Section 8.1.2)?

Problems of this nature are common in the analysis and interpretation phase[7] of a qualitative research project. Lindlof and Taylor (2002, pp. 209/10) describe five major challenges of this phase:

- The amount of data that must be dealt with can be daunting.
- "The findings of any qualitative study can be given multiple plausible interpretations. Alter any aspect of the researcher – personality, value system, culture or theoretical orientation – and the account of observed events will also shift prismatically." Lindlof and Taylor conclude that "one always faces the challenge of determining how to look at data in a way that yields the most interesting and insightful view".
- The research problem may not even be known until the data analysis is well advanced. "Qualitative research is a journey powered by the researcher's own growth of interpretive competence, and concepts and claims usually come into focus only after a long period of study in the field."

[7] Lindlof and Taylor (2002, p. 210) explain Spiggle's (1994) distinction between analysis and interpretation. *Analysis* means the process of labelling and breaking down raw data and breaking them into patterns, themes, concepts and propositions. *Interpretation* is defined as the process of '[making] a construal'. Theory and experience come together in the writing of an interpretive claim.

- The interpretation should be made understandable and acceptable for a wider community. “The research community is always a silent partner at the table where interpretive work is done.”
- Acquiring skills in qualitative analysis is another major challenge.

Each of those challenges were present and influential in the analysis and interpretation of the data.

Two major themes emerged after close, repeated listening to the recordings and reading of the transcripts.

1. In the analysis phase of the interviews it became obvious that not only the validation exercise but also the task of comparing and ranking the given ‘proofs’ encouraged a learning and reflective process. The students’ opinions were considerably influenced by other, more or less good approaches that had already been presented. *Student F* shows his awareness of this effect:

   > *I*: Would you be happy with Gerard’s answer?
   > *Student F*: *“I would now, but I will probably end up changing my mind after seeing the rest.”*

   Some students changed their minds about the presented answers when they were asked to rank them. The observed process was as much evaluation (including decisions about what is valuable in a proof) as validation of proofs. The notions of *proof evaluation* and *proof validation* are clarified in Section 1.2.

   The shift in focus from proof validation to proof evaluation corresponds to the underlying theoretical framework, namely considering proofs as artifacts in mathematical practice. Evaluation of artifacts is a crucial activity in any community of practice, therefore examination of evaluation of the artifacts proofs might be fruitful. Investigations of performances, habits and criteria associated with proof evaluations carry a potential to give insights into knowledge and attitudes towards proofs, see Section 4.1. To interpret students’ proof evaluations a schema was developed, which has been introduced in Chapter 4. One major theme of the interview interpretation is how students validate and evaluate mathematical proofs. How do they try to understand the proof? Do they consider actual features of the proof and its purposes and the author’s intention and relate them to each other? What do they accept as a mathematical proof? What do the students regard as purposes of the proof?

2. Section 2.3 emphasized the significance of *transparency of artifacts* in the learning process. According to Hemmi (2006) the fact that the nature of proof is often not *visible* to incoming students constitutes a major obstacle to their participation in the community of mathematical practice. One theme of the interview interpretation is whether the nature of a proof and its purposes become more *visible* to students through the practice of proof validation and evaluation. Do evaluation tasks such as this help to make the nature and purposes of proof *visible* to the students?

The proposed interpretations of students' validation and evaluation habits are led by my interest in their views and knowledge about mathematical proof. The choice of investigation themes is appropriate to attain insights regarding some of the questions of the prepared framework of topics (Section 8.1.2), for example what makes a proof valuable to the students, what criteria are used in the ranking or evaluation processes, whether the students change their minds during the meetings or whether features and purposes of proof become *visible* to students during the process of proof validation. Other questions suggested in this framework did not get much attention in the interpretation of the interviews, for example the questions about students' validation processes and the role of examples in proofs. Reasons for excluding these questions include the fact that the interviews did not provide as much information regarding those topics as anticipated, and the time demand of the process of analysing and interpreting interviews. Another researcher might have focussed on different aspects of the interview transcripts and illuminated the students' performances from another view. I do not claim to interpret all aspects of the interviews, but some which elucidate first year students' views of mathematical proofs and their purposes in mathematical practice in a plausible and useful way.[8]

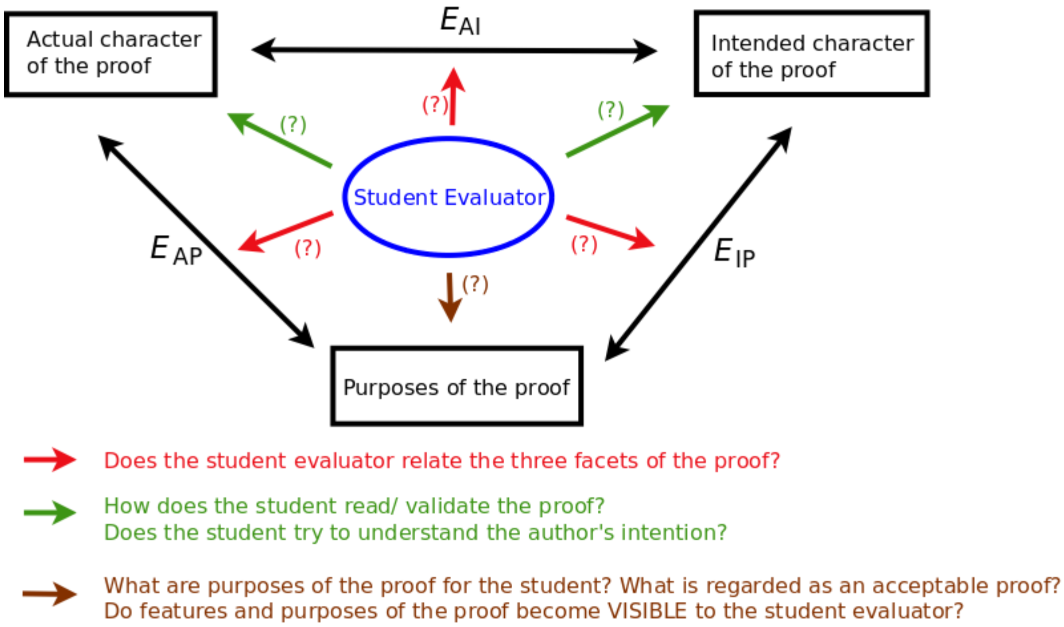
In this study *exemplars* in the form of brief interview segments are used to advance interpretive claims. An *exemplar* is a part of a project's data that is shaped and used to advance an argument. Lindlof and Taylor (2002, p. 234) describe: "When the researcher is ready to develop an interpretation of the results of the data analysis, these incidents – or at least the most persuasive ones – are then shaped into exemplars."

8.3.2 Student Evaluations: Transcript-Excerpts and Interpretations

This section presents excerpts from transcripts of the students' evaluations of the proposed proofs. It also suggests an analysis of these transcripts and interpretations regarding the research questions as outlined in Figure 4.2, which is reproduced on page 93 below. For each of the proposed proofs, the transcript excerpts are listed, followed by a table including codings of the evaluation contents. These coding tables provide overviews about the participants' evaluations of the particular proofs. The heading lines refer to the codes for the participating students. The students' proof evaluations were coded regarding various categories, for example whether a student regards a particular proof as sufficient, what he/she approves or disapproves about the proof or whether and how he/she compares the proof to others. The coding scheme in Appendix F includes explanations of the categories. The codes in the bottom lines of the tables refer to how the students placed the proof in

[8]Lindlof and Taylor (2002, p. 240) explain that "qualitative researchers do seek to produce and demonstrate credible data. They want to inspire confidence in readers (and themselves) that they have achieved a right interpretation. Notice that we did *not* say, *the* right interpretation. An indefinite number of interpretations could be constructed from any research experience, but usually the ones that researchers choose to develop are those that they find most plausible, insightful, and/or useful."

the ranking of all proposed proofs of the associated statement. Following the coding tables are interpretations of the students' proof evaluations, occassionally using short transcript exemplars to advance certain claims. In these interpretations the contents of the coding tables are interpreted and the corresponding views of the *"Experienced Evaluator"* are related to the students' evaluations. The interpretations also address the research questions listed in Figure 4.2:



After presenting interpretations of the students' proof evaluations for each task, their rankings of the proposed proofs are analysed and the findings are summarized.

Student Evaluations of ‘proofs’ of *Statement I*

I.1. Student evaluations of *Anna’s* answer.

I.1.1. Transcript excerpts.

***Student A*:**

> *“I think it’s correct, because she gives examples, and then she explains what she does. So her answer is right.”*

***Student B*:**

> *“I’m missing the proof, that you —, this general — She took the numbers from 1 to 9, but what about all the other numbers? — I’m missing the generalization. — Nice example, but that’s about it.”*

[Translated from German: *“Bei ihr fehlt mir der Beweis, daß man, —, also dieses generelle — Sie hat die Zahlen von 1 bis 9 genommen, aber was ist mit den ganzen anderen Zahlen? — Die Generalisierung fehlt mir. — Schönes Beispiel, aber das ist es dann auch.”*]

***Students C/D*:**

> *Student D*: *“I do not know if that can be taken as a valid answer but it does work.”*
> *Student C*: *“Most marks”*
> *Student D*: *“Ya”*
> *Student D*: *“It seems very simple or something.”*
> *I*: Is it a proof?
> both: *“It’s not a proof, but it works”* (...)
> *Student C*: *“It’s not a general case. (...) I would give it most marks.”*

***Student E*:**

> *Student E*: *“Anna’s is not general. — She is only giving examples.”*
> *Student E*: *“She doesn’t prove that ‘When you square, the answer will end in 0,4,6 ...’. If she’d proved that, it would be o.k.”*

***Student F*:**

> *Student F*: *“Only for certain examples, it is not general.”*
> *Student F*: *“She has only used up to ten numbers.”*

***Student G*:**

> *Student G*: [Pause] *"Very cool.* [laughs] *It's different, totally different from — in school. Ahm* (...) *I could never think of anything like that* (...) *the way she writes it down* (...)*"*
> *I*: Do you think, Anna proves the statement?
> *Student G*: *"Not really, — but, I don't know — not really, she's just saying they are odd and that's it."* [Long pause]
> *"It's wrong* (...) *You know the way there is only certain odd numbers* (...)*. It's stating the obvious — I don't know."*

The interviewer tries to find out what advice *Student G* would give Anna.

> *Student G*: *"I'd always do numbers as well, but we've always be advised to try letters."*
> *I*: Why are letters better than numbers?
> *Student G*: *"Because like ahm, letters* [stand for (?)] *various things —"*
> *I*: Do you like this answer?
> *Student G*: *"It's different."*

***Student H*:**

> *Student H*: [Pause] *"It's a good answer. She is just using examples, like. There is no kind of proof, it's just —. But it does make sense. And, I'd say you could use it as/ to actually answer it."*
> *I*: Why? If it doesn't prove it?
> *Student H*: *"Well, she does show that whatever what she'd multiply or what she'd square, an even number will come up at the end. The same for the odd."*
> *I*: But if she shows that, doesn't that mean she proves it?
> *Student H*: *"Ya, she just proves it by example."*

A	***B***	***C/D***	***E***	***F***	***G***	***H***
C Pr?	NotC NoGen	C NoPr NoGen	NotC NoGen	NotC NoGen	C Pr?	C NoPr
1^{st}	4^{th}	3^{rd}	4^{th} jointly	4^{th} jointly	5^{th} jointly	2^{nd}

Table 8.1: Coding table - Evaluations of *Anna's* answer

I.1.2. Interpretation. The coding table[9] refers to three groups of comments.

1. **C/NoPr**: The student is happy with Anna's answer and likes the fact that she 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.
2. **C/Pr?**: 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 students regard Anna's answer as a sufficient proof of the statement.
3. **NotC/NoGen**: The student does not accept the answer as a proof of the statement because it *"is not general"*.

The opinions of five students belong to (C/NoPr) or (C/Pr?):

- Two of them (*Student A*, *Student G*) do not seem to focus on relations E_{IP} or E_{AP} as there is no evidence to suggest that they are considering the purposes of mathematical proof.
- Some of those students (*Student C*, *Student D* and *Student H*) 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:

  "It's not a general case." (*Student C*)

  However, 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 E_{AI} may be more important to them than the relations E_{IP} or E_{AP}:

  "It's not a proof, but it works"(C/D),
  "There is no kind of proof, it's just —. But it does make sense."(*H*).

  Internal correctness seems to be considered more important by this group of students than the purpose of establishing the general truth of the statement, at least to judge from their considerations of this example.

[9]Occasionally codes have been named after students' phrasing. Some of these codes might cause confusion for a reader. For example, code 'C' signifies that the student regards the answer as sufficient. Often students express this using the term 'correct' and therefore this opinion is labelled 'C' not 'S'. Explanations for each code can be found in Appendix F.

The three (NotC/NoGen)-students do relate the actual proof not only with the author's intention but with its purposes and therefore do consider relations E_{IP} and E_{AP} as well as E_{AI}. 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 give attention to all three relations E_{AI} – E_{AP} in their evaluations of Anna's answer.

The transcripts of those students who do consider relations E_{IP} and E_{AP} in their evaluations allow inferences about which purposes they acknowledge. 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 (NotC/NoGen)-students (*Student E*) 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 o.k."*

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 F* for example seems to see it as basically the same as Benny's proof which consists of a collection of examples, just *"up to ten numbers"*. *Student B* like *Student F* 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."*

Student G'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 becomes 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 G*'s comments do not show clearly what she considers as a valid or valuable proof, but she certainly thinks about it.

I.2. Student evaluations of *Benny's* answer.

I.2.1. Transcript excerpts.

***Student A*:**

> *Student A*: *"I think it's also true, — similar to Anna's work. It is true, ya, it's correct."*
> *I*: Does it prove the statement?
> *Student A*: *"Yes, it proves the statement, ya."*

***Student B*:**

> *"More example than proof than Anna's was. What about all the other numbers?"*

[Translated from German:

> *"Mehr Beispiel als Beweis als Anna hatte. Was ist mit den ganzen anderen Nummern?"*]

About the last sentence:

> *"She could have just copied this."*

[Translated from German:

> *"Das hätte sie ja einfach abschreiben können."*]

***Students C/D*:**

> *Student C*: *"The other one was more like, it dealt with more numbers, but this just takes a couple of numbers."*
> *Student C*: *"Anna's* [is] *better than Benny's."*
> *Student D*: *"Benny's is true but he doesn't prove it."*
> *I*: But you have said, Anna's answer doesn't prove it either.
> *Student D*: *"She has this — with the endings."*
> *Student C*: (...?)

***Student E*:**

> *Student E*: *"He's proved it for those values, but he hasn't proved it for every value."*
> *Student E*: *"This is the most basic you can do. Everyone can do that. He hasn't really thought about it at all."*
> *Student E*: *"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. Really it's the same."*

***Student F*:**

> *Student F*: *"Pretty much the same really. A wider range."*
> *I*: Which would you prefer?
> *Student F*: *"I wouldn't really prefer either to be honest."*
> *I*: Is it different to what you would have done?
> *Student F*: *"I probably would have done pretty much the same thing."* (...?)
> *I*: So, that's not a proof for you?
> *Student F*: *"No, that's just a statement. Anna's is more of a proof. — She says why the squares are odd, because they end in that. He just presumes that they are odd numbers, I suppose."*

***Student G*:** *"That's examples* (...?)*"*

***Student H*:**

> *Student H*: [Pause] *"I don't think the* 6^2 *equals* 8*, I don't know if that's right."*
> *I*: Oh, that's a typo. That should be 36, of course.
> *Student H*: *"And, that's shown, that you multiply any even number at all. It's shown a bigger range, really, of even numbers or odd numbers, that shows that you're still ending with an odd number. He's still proving it. I think it's not as clear as Anna's answer.* (...?) *But he shows a bigger range, shows that it can be done for any number.* (...) *Rather than just showing how it ends, he kind of does an overall, shows it for bigger numbers as well."*

A	***B***	***C/D***	***E***	***F***	***G***	***H***
C Pr	Ex NoPr NoGen	NoGen	NoGen	St NoPr	Ex	Pr Gen/T
A-B:S	A:MPr(B)	A:MPr(B) A-B:S	A:MPr(B) A-B:S	B:WR(A) A:MPr(B)		B:WR(A)
2^{nd}	6^{th}	4^{th}	4^{th} jointly	4^{th} jointly	4^{th}	5^{th}

Table 8.2: Coding table - Evaluations of *Benny's* answer

I.2.2. Interpretation. Benny's answer consists of a collection of examples. In the view of the representative *Experienced Evaluator*, Benny's argument is not sufficient regarding E_{IP} and E_{AP} as the character of this answer does not address the purposes of proving the statement generally or of explaining why it is true. Because of the typos this answer is not considered as fully successful regarding E_{AI} either. The view of the *Experienced Evaluator* does acknowledge that some evaluators might consider Benny's answer to be a reasonably convincing demonstration of the fact that squaring preserves parity. All these views can be identified in the students' evaluations, however individually.

- Two students are happy with Benny's answer, but for different reasons.
  - *Student A* appreciates its internal correctness (she does not notice the typos). *Student A* also values examples to verify the statement very highly in a proof and consequently considers Anna's and Benny's approaches as best of all proposed ones. She seems to believe that a major purpose of proof is to explain the contents of the statement.
  - *Student H* interprets Benny's answer as a representative choice of examples, covering positive and negative integers. Even though she considers E_{AI} carefully and is the only student who recognizes the typos, *Student H* approves what she sees as the general aspect of Benny's answer.

    But he (...) *shows that it can be done for* <u>*any*</u> *number."*

    The awareness of general applicity, although perhaps underdeveloped, indicates that *Student H* relates Benny's answer at least to the purpose of proof *verification*, which leads her to the opinion that Benny's answer, rather than Anna's, is a valid proof:

    "He's still proving it."

- Five students recognize that Benny's answer does not prove the statement, mostly because it's not generic. (Only *Student F*'s comment does not indicate clearly why he does not accept Benny's answer as valid proof.)

- The evaluation by *Student G* does not indicate whether she regards Benny's answer as sufficient.

Interestingly seven of the eight students comment in the interviews on how they compare Benny's answer to Anna's, even though they have not been 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"* (*Students C/D*),
  > *"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"* (*Student E*).

- 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"* (*Student F*).

Considering Benny's answer in comparison to Anna's, some of the students who have interpreted Anna's proof as a list of randomly chosen examples when discussing Anna's answer now identify some potential in Anna's answer to provide a general proof:

- *Student F* 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 B* 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.

The students' comments on Benny's approach provide us with some information about what purposes or aspects of proofs are considered important to the students. To the majority (five students), generality is a substantial feature of a valuable mathematical proof. Three students regard the identification of general patterns and one the inclusion of explanations as valuable in their evaluations of Benny's answer.

I.3. Student evaluations of *Ciara's* answer.

I.3.1. Transcript excerpts.

***Student A*:**

> *Student A* [laughs]
> *I*: Does this prove the statement?
> *Student A*: *"It proves the statement, but it doesn't explain anything and there are no examples. It does prove the statement* (...) *the squares and everything, but* (...) *just two boxes and* (...) *— It is right"*
> *I*: Why is it right?
> *Student A*: *"Because* – [long pause] – *I don't understand this one, but I understand this here, by counting the squares"* (...?)
> *I*: Can you see the even number? The first diagram, it's the square of which even number?

[The interviewer tries to explain the connection between Ciara's diagrams and the statement. It is not clear whether *Student A* understands this explanation.]

> *Student A*: *"It answers the question, but it doesn't explain."*
> *I*: Would she write a line to explain it, would this be a proof?
> *Student A*: *"She should write a line and also explain it in numbers, to explain, and then it would be a proof"*

***Student B*:**

> *Student B*: *"As a proof not enough."*
> *I*: How does the diagram explain the statement?

Student B explains her understanding of Ciara's diagram. Her explanation demonstrates that she interprets the diagram as a visual representation of a general proof of the statement. *Student B* explains Ciara's assumed intention correctly.
[The transcript of *Student B*'s explanation of her view of Ciara's diagram is omitted.]

> *Student B*: *"This is a good approach, but you have to think a lot, what she is doing there. — You have to explain everything you are doing. Maths is as much writing as calculating, actually more writing than calculating. And here nothing is written. — This might be a good answer, but when you are sitting in front of it, you have to think a lot, what had she thought when doing it. — I'm missing generalizations, formula, example and explanation."*
> *I*: An example is part of a good proof?
> *Student B*: *"To answer the question. — What I'm missing is this x, to know that it's* [applicable] *for all."*

[Translated from German:

> *Student B*: *"Als Beweis nicht genug"*
> *I*: Wie erklärt das Bild die Aussage?

Student B erklärt ihre Interpretation von Ciaras Diagramm.

> *Student B*: *"Das ist ein guter Ansatz, aber man muß sich sehr überlegen, was sie da macht. — Man muß alles, was man macht, erklären. Mathe ist genausoviel schreiben wie rechnen, eigentlich mehr schreiben als rechnen. Und hier ist nichts geschrieben. — Das ist vielleicht 'ne gute Antwort, aber wenn man davor sitzt, muß man sich sehr überlegen, was hat die sich dabei gedacht. — Mir fehlen Generalisierungen, Formeln, Beispiel und Erklärung."*
> *I*: Ein Beispiel gehört zum guten Beweis?
> *Student B*: *"Zur Beantwortung der Frage. — Was mir fehlt ist dieses x, daß man weiß, daß es für alle geht."*]

***Students C/D*:** [Their first reaction is] *"Wow"*.

> *Student C*: *"I don't get it at all."*
> *Student D*: *"No."*
> *I*: Can you try to figure it out?
> [Silence]
> *I*: Would you be the teacher, and you'd get this answer, what would you do?
> *Student C*: *"You don't want to discourage them. I think I'd ask her 'Could you write something to explain it to me', to give her something."*
> *Student D*: *"I don't know."*
> *Student C*: *"I think I would ask, what they did, what was the method."*

***Student E*:**

> *Student E*: *"It's quite good. — I like it. I'm not sure if it's mathematically right. Because, she is going by the fact that this is going to be even, this bit (probably) — and that's even and that, and this makes it odd, this one here."*
> *I:* And the first one?
> *Student E*: *"Well, that's the numbers along the side multiplied by this side gives an even number."*
> *I:* How do you know they are even?
> *Student E*: *"Well — you can see they are even by the squares, there is two at each side (...)."*
> *Student E*: *"I think, it's a lot better than the other ones. There is more thought in it. I'm still not sure if it is —."*
> *I:* Would you be the teacher, would you give Ciara some advice?
> *Student E*: *"I'd try to get her to prove it mathematically."*
> *I:* What does that mean, mathematically?
> *Student E*: *"With — Something you can actually read. Not drawings.*

— Hm. That's my opinion, I suppose it's fine."
I: Would you accept it like this?
Student E: *"Yes. I think it's fine."*
I: But it's not mathematical enough?
Student E: *"No, hm. Ya, I think it is mathematical. Sorry, it's mathematical, but, I suppose, it's just a different way. I would recommend that, if you write down a line like (...) try to prove it with a k (...)."*
I: That would be a different approach. Suppose leaving it and explaining this approach, would that be as good as an algebraic approach?
Student E: *"Probably, yes. Because, I think it's the same, having* 1 *at the end extra. I wouldn't really have noticed that, until you pointed it out. She leaves a lot for the person to, actually, to deduce."*

***Student F*:**

Student F: [laughs] *"Is that all she has written?"*
I: Yes.
Student F: *"Ah, ok."*
Student F: *"I can't even tell, I'm too lazy to count, are that even and odd numbers? — I suppose, she should have more written. Because it's just a bunch of squares and you are supposed to be seeing that she has even numbers there and odd numbers there. They are even and odd and then squares."*
I: Would she explain her ideas, what she is thinking, would the answer be fine?
[The interviewer explains the connection between Ciara's diagrams and the statement.]
Student F: *"Ya. I think so, ya."*
I: But without explanation?
Student F: *"No."*

Student F's advice to Ciara: *"Explain it further!"*

***Student G*:**

Student G: [Pause] *"It's good, isn't it? — Very good."*
I: What do you like about it?
Student G: *"Visual, like it's, ahm, graphic and the whole lot. It's more interesting, it would grab your attention quicker than the rest, the rest is all like theory, getting answers. This is you can see it and it's visual. Ya, it's good."*
I: Where do you see the even and the odd numbers?
Student G: *"Hah?"*
The interviewer tries to explain her question.
Student G: *"Ahm, she's got (...?) Odd by odd gives an even – oh no."*
[Pause] *"5 boxes by 5 boxes gives 25. And here is 10 by 10 gives 100 — What do you mean?"*
I: If I had my doubts that the statement is true, how can I be convinced

> by this picture that it's true? That's what the student is asked to do, isn't she?
> *Student G*: *"I suppose she just has a few numbers as well, and that's it. I mean she applied — to make the five odd — Ya, I get what you mean."*
> *I*: I agree with you, I think this answer is very good. I am interested in **why** you think it's good and how does it prove the statement.
> *Student G*: *"Hm, there is no theory behind this. It's just another examples like Benny's. No theory or proof behind it."*
> *Student G* comments about how students would react to this.

When asked to rank the proposed answers, *Student G* comments again on Ciara's approach:

> *"I know there is no theory behind it, but I like that".*

***Student H*:**

> *Student H*: [Pause] *"Are you supposed to count the blue boxes rather than the —* [Pause] *The first one is like, she is showing that it'll always be inside one of the boxes, like. That if you, if you multiply it by something, it'll always be kind of —, it will always be inside the box. In the second one there will always be one or more outside,* (...) *there will always be one extra type of thing."*
> *I*: Is that a good answer?
> *Student H*: *"I think it's a good answer. I think it's a kind of logic, it's like, it would make you think more. It shows that you know —. The only thing is I don't really know how it's proved as such. It's just kind of showing that there will always be an even number* (...?) *and that there'll always be an odd number so done that it's out. I don't really understand how it exactly proves it."*

When asked to rank the proposed answers *Student H* comments again on Ciara's approach:

> *"I'm not sure about Ciara's answer, how it proves, that's the only one that I wouldn't be sure of. That proves something about —. It proves that an even number can be inside the box, and an odd number would have to have a natural number outside the box. But it doesn't necessarily prove the squaring —."*

A	***B***	***C/D***	***E***	***F***	***G***	***H***
NoUnd	Und	NoUnd	Und	Und-I(?)		Und?
			VG		VG	VG
C			C		C-Ex	
Pr	PrNotEnough		Pr		NoPr	Pr?
NoExpl	NoExpl	NoExpl	NoExpl	NoExpl		
	NoForm					
	NoEx				ApprDid	
	NoGen					
			C:MTh(AB)	MI		
6^{th}	5^{th}	6^{th}	3^{rd}	6^{th}	3^{rd}	6^{th}

Table 8.3: Coding table - Evaluations of *Ciara's* answer

I.3.2. Interpretation. A remarkable variety of opinions on Ciara's approach is observed, the comments differ substantially. These comments provide some information about what the participants value or disapprove about Ciara's visual approach.

- From the beginning *Student G* seems to be very impressed by Ciara's approach:

  > *"It's good, isn't it? — Very good."*

  When asked for the reasons she explains that she likes Ciara's approach as it is visual:

  > *"It's more interesting, it would grab your attention quicker than the rest".*

  However it is not clear whether *Student G* relates Ciara's answer to the statement. When the interviewer tries to find out how *Student G* connects the diagram to the statement, i.e. to even or odd numbers, she reacts with surprise. Only after the interviewer tries to explain a connection between the diagram and the statement, *Student G* seems to become aware of such a connection. Nevertheless, *Student G* does not consider that the text 'square of an even number' and 'square of an odd number' suggests that Ciara intends for her diagrams to be interpreted as demonstrating general phenomena as well as specific examples. Instead she interprets it as visualisation of one particular example:

  > *"I suppose she just has a few numbers as well, and that's it* (...) *There is no theory behind this. It's just another examples like Benny's. No theory or proof behind it."*

  Even after considering all the other approaches to this task *Student G* does not change her opinion about Ciara's answer:

  > *"I know there is no theory behind it, but I like that".*

  Student G does not seem to relate Ciara's approach to the statement. Also she does not seem to be concerned about the internal correctness of this approach or about other relations to an assumed intention. Thus *Student G* does not seem to consider relation E_{AI} in her evaluation of Ciara's answer. Regarding relation E_{IP}, *Student G* does not appear to be concerned whether Ciara's answer establishes the truth of the statement. She does not seem to consider the purposes *verification* or *explanation* at all. However, *Student G* appreciates a didactic value in Ciara's answer. One advantage of a pictorial representation is its ability to stimulate the reader's attention and interest. *Student G*'s appreciation of didactic values indicates that she considers the purposes of proof *communication* and maybe *discovery*.

- *Student F* shows attempts to view Ciara's answer as an approach to prove the statement generally:

  > *"(...) you are supposed to be seeing that she has even numbers there and odd numbers there. They are even and odd and then squares."*

  However, he does not engage in an interpretation of the diagram:

  > *"I'm too lazy to count, are that even and odd numbers?"*

  Consequently, *Student F* criticizes the lack of explanation:

  > *"I suppose, she should have more written."*

  During this phase of the interview *Student F* does not yet arrive at his view to interpret the diagram. Only when ranking all of the proposed proofs of *Statement I*, *Student F* arrives at his opinion:

  > *"Ciara's is worst, because it's only one example anyway, really."*

  Student F does not try to understand the intention behind Ciara's diagram and therefore does not seem to consider relation E_{AI}. He rejects Ciara's approach as he finds it difficult to comprehend. Thus - the purpose *communication* seems to play a role in his evaluation of Ciara's answer. In the end he interprets Ciara's approach as visualisation of one example which he does not regard as sufficient to establish the truth of the statement. General applicability and therefore the purpose *verification* seems to be considered in his evaluation of Ciara's proof.

Three students (*Student B*, *Student E* and *Student H*) do consider that Ciara's examples are of a generic quality. The comments of *Student B* and *Student E* indicate that they understand how Ciara's approach carries potential to prove the statement generally, that is they understand Ciara's assumed proof idea. *Student H* also interprets Ciara's answer as an approach to prove the statement generally. However, *Student H* does not recognize the assumed idea behind Ciara's proof.

- *Student B* and *Student E* understand the assumed idea behind Ciara's proof. Nevertheless they evaluate this approach differently.

  - *Student B* regards it as *"not enough"* to be a proof. She misses *"generalisations, formula, example and explanation"*.

  - *Student E*'s view is similar to that suggested in the view of the *Experienced Evaluator*. He accepts Ciara's answer as a correct proof of the statement, but criticizes the deficit of explanation:

    > *"She leaves a lot for the person to, actually, to deduce."*

    Student E is the only interview participant who recognizes and mentions the similarity to an algebraic approach.

    > *"I think it's the same, having* 1 *at the end extra."*

- *Student H* seems to understand Ciara's diagram partially. *Student H* recognizes that the square of an even number can be represented as *"inside one of the boxes"* and that in the representation of the square of an odd number *"there will always be one extra type of thing"*. She does not seem to understand how a diagram like the first one could represent any product of even numbers or a diagram like the second one could represent any product of odd numbers, or in particular how the diagrams represent the squares of even or odd numbers: *"But it doesn't necessarily prove the squaring —."*

These students seem to consider all three relations E_{AI} – E_{AP} in their evaluations of Ciara's answer.

The evaluations of *Student A*, *Student C* and *Student D* do not indicate clearly how they interpret Ciara's approach.

- *Student A* does not understand Ciara's approach, but is not aware of her lack of understanding. Initially she regards this answer as a correct and sufficient proof:

  > *"It proves the statement* (...) *the squares and everything* (...) *— It is right"*

  When the interviewer explains the connection between Ciara's diagram and the statement, *Student A* reconsiders her approval and criticizes Ciara's poor explanation:

  > *"It answers the question, but it doesn't explain.* (...) *She should write a line and also explain it in numbers, to explain, and then it would be a proof."*

- *Student C* and *Student D* just criticize the lack of explanation. It is not clear how they would have evaluated this answer if they had more understanding.

As these students fail to see an intention in Ciara's approach, no suggestions can be made on how they consider the relations E_{AI} or E_{IP} in their evaluations of Ciara's answer. They all criticize Ciara's lack of explanation and therefore consider *communication*. Relations to other purposes of proofs do not seem to be relevant in these evaluations.

Only two of the students compare Ciara's answer to those they had seen before (Anna's and Benny's):

- *Student E* prefers Ciara's because *"there is more thought in it"*. This may indicate an appreciation of attempts to find generalizations and reasons why something is true in a proof. Thus, *Student E* considers the purposes *explanation* and maybe *systematization* in his evaluation of Ciara's proof.
- *Student G* favours Ciara's answer because it is unusual.

The fact that six of the students explicitly criticize the deficiency of explanation in Ciara's proof suggest that they consider the purpose *communication* in their evaluations.

None of the students observes that Ciara's diagrams suggest the stronger statement that the square of an even number is always a multiple of 4 and that the square of an odd number always leaves a remainder of 1 on division by 4. The purposes *systematization* and *discovery* seem to play at most an insignificant role in these students' evaluations of Ciara's approach.

I.4. Student evaluations of *Darragh's* answer.

I.4.1. Transcript excerpts.

***Student A*:**

> *Student A*: [Pause] *"The answer is also true, but no presentations of the letters and no explanations, it's not enough. He didn't give examples like* (...?). *He just uses letters, didn't do anything about it."*
> *I*: What is missing?
> *Student A*: *"Examples* (...?)*"*
> *I*: Why examples, to understand it better or to complete the proof?
> *Student A*: *"To complete the proof, to explain more what he does, because the person wants to understand everything."*

Student A ranks Darragh's answer fifth (out of six), because she finds it not explained enough. In comparison to Elaine's approach, which she ranks higher than Darragh's:

> *"Because Elaine's answer is a lot more comprehensible. Because of the way she explains it. Darragh says k is any whole number and l is any whole number. He wasn't so specific."*

***Student B*:**

> *Darragh*'s answer: *"Better than the other ones — because you see the mathematical way, why those even and those odd. — Would you see the l as an x — It is good alright, but — I'm still missing something. — Why did he choose k and l instead of just one letter? — It looks a bit short. I would have expected more, more thoughts. — More words are lacking here, like 'if' und 'then', probably an example to make it more perceptible. The formula are perfectly fine, but I would say, because this — ahm, let $2k$ be x and — $2x$ is always even. — The last step is lacking — and examples."*

[Translated from German: *"Besser als die anderen — weil man den mathematischen Weg sieht, warum die gerade und die ungerade. — Wenn man das l als x sehen würde. — Es ist schon gut, aber — es fehlt mir immer noch was. — Warum hat er k und l genommen anstatt nur einen Buchstaben? — Es sieht ein bißchen wenig aus. Ich hätte mehr erwartet, mehr Gedankengänge. — Hier fehlen mehr Worte, wie 'if' und 'then', wahrscheinlich ein Beispiel, um das anschaulicher zu machen. Die Formeln sind 'perfectly fine', aber ich würde dann sagen: 'Weil das — ähm, let $2k$ be x and — $2x$ is always even. Der letzte Schritt fehlt — und Beispiele."*]

***Students C/D*:**
[Immediate reaction:] *"That's good."*
[*Student C* and *Student D* check the argument line by line, and detect a typo.]

> *I*: What do you like about it?
> *Student C*: *"He is actually taking the general case."*
> *I*: Is this answer perfect, or do you miss something?
> *Student C*, *Student D*: *"I think it's ok."*

***Student E*:**

> *Student E*: *"It's good but why does he need to introduce l?"*
> *Student E*'s advice: to use just one variable.
> *Student E*: *"The rest is fine."*

***Student F*:**

> *Student F*: *"That's better in my opinion anyway. — I like that answer."*
> *I*: What do you like about it?
> *Student F*: *"The fact that it counts for all numbers. And — ah — you can't say that it's not true, really, I don't think, unless there is something that I'm missing. Ya, because he shows that two times any number is an even number, and then you add 1 to that, and obviously it's going to be odd. That's a good answer, I think."*

[When ranking the proposed proofs of *Statement I*, *Student F* appreciates that Darragh's answer is sufficient without excess:]

> *Student F*: *"Darragh is more straightforward, Fintan's is very dragged out, in my opinion."*
> *I*: What does that mean, 'dragged out' ?
> *Student F*: *"That's a — Darragh just says exactly what, you know, it's short and still proves it. That one you have to read a lot more."*

***Student G*:**

> *I*: We have letters now.
> [Pause]
> *Student G*: *"That's good. I like that.* [laughs] *I wouldn't even think of something like that."*
> *I*: Why is this a good answer?
> *Student G*: *"I don't know, to me it just kind of proves —"*
> [She reads the argument again.]
> *"There is more kind of like proof behind it."*
> *I*: Why?
> *Student G*: *"Ahm."* [*Student G* explains Darragh's answer.] *"That's good."*

***Student H*:**

> *Student H*: [Pause] *"Where does he take the 2 from?"*
> *Student H* finds the typo. She then explains Darragh's answer.
> *Student H*: *"I think it's more thought — It's more of a — what do you call it — ahm — like, ah, using like functions or using, rather than examples, it's using more of a (...?). Can't think of the word. It's a good answer, though. It makes sense."*

A	***B***	***C/D***	***E***	***F***	***G***	***H***
C PrNotEnough NoExpl NoEx	VG MExpl NoEx	VG Gen/T	VG	VG Gen	VG	VG T
	Why2Var		Why2Var			
	D:ExplWhy (ABC)				D:MPr (ABC)	D:MTh (ABC)
R-NotExpl				R-ApprSuff		
5^{th}	2^{nd}	2^{nd}	2^{nd}	1^{st}	1^{st}	4^{th}

Table 8.4: Coding table - Evaluations of *Darragh's* answer

I.4.2. Interpretation. Only one of the students (*Student A*) does not evaluate Darragh's approach as good, though she considers it *"true"*, meaning it does not include logical mistakes. In terms of the schema, *Student A* regards Darragh's proof as sufficient regarding relation E_{AI}. However, in the view of this student, Darragh's answer fails regarding E_{IP} (and therefore also E_{AP} as she sees actual character and intention matched) and is not sufficient to be a proof of the statement because the reasoning is not explained enough to the reader of the proof. *Student A* suggests either an inclusion of examples or a written explanation

> *"to complete the proof, to explain more what he does, because the person wants to understand everything"*.

It is not clear whether *Student A* understands Darragh's answer herself. Some of *Student A*'s comments in the ranking process show that her intention, when valuing explanations in a proof highly, is not only to provide the reader with a better understanding of the proved statement, but also to improve the comprehensibility of the proof itself.

> *"Because Elaine's answer is a lot more comprehensible. Because of the way she explains it. Darragh says k is any whole number and l is any whole number. He wasn't so specific."*

Student A's comments show, that she considers sufficiency of a proof as less important than a didactic aspect, which is how difficult or simple both statement and proof are introduced to a reader. *Student A* seems to be happy with a 'proof' if she can understand what it is saying and if it helps her to understand what the statement is saying, regardless of whether it provides any explanation for the **truth** of the statement. To this student the main purposes of a proof seem to be to explain the contents of a statement and to say something *"comprehensible"*. She relates the actual character of the proposed proof to these purposes and therefore considers relation E_{IP}. *Student A* does not seem to consider any of the purposes as suggested by De Villiers, except maybe *communication*.

All the other students regard Darragh's answer as a good approach, for various reasons:

- Three students (*Student C*, *Student D* and *Student F*) appreciate that Darragh's approach proves the statement in general. This indicates that they consider the purpose *verification* in their evaluations of Darragh's answer.
- When ranking, *Student F* compares Darragh's to Fintan's answer and values its sufficiency without excess:

  > *"Darragh just says exactly what, you know, it's short and still proves it."*

  Avoiding excess in a proof is important to *Student E* and *Student B* as well. That is indicated when they criticize that Darragh uses two variables, which they find unnecessary.
- Two students (*Student B* and *Student H*) describe some reasons why they value Darragh's answer more highly than the three previous ones. *Student B* explicitly says that in her opinion this answer provides the reader with insights into **why** the proved mathematical fact is true:

  > *"Better than the other ones — because you see the mathematical way, why those even and those odd."*

  This shows that she considers the purpose *explanation* (providing insight into why it is true) in her evaluation of Darragh's answer.
- *Student H* struggles to describe why she prefers Darragh's answer:

  > *"I think it's more thought — It's more of a — what do you call it — ahm — like, ah, using like functions or using, rather than examples, it's using more of a (...?), can't think of the word. It's a good answer, though. It makes sense."*

  Like a similar comment of *Student E* regarding Ciara's answer, this comment seems to indicate an appreciation of attempts to find general patterns and mathematical context in a proof.

Two students (*Student E* and *Student G*) do not refer to any reasons why they approve this answer. *Student G* mentions that in comparison to the answers she had seen before *"there is more kind of like proof behind it"*, but she does not clarify this comment.

Only two of these students explicitly mention reasons to slightly disapprove Darragh's answer:

- *Student B* criticizes, like *Student A*, Darragh's poor (in her opinion) attempts to explain statement and reasoning to the reader:

  > *"More words are lacking here, like 'if' und 'then', probably an example to make it more perceptible. The formula are perfectly fine, but I would say, because this — ahm, let $2k$ be x and — $2x$ is always even."*

- Three students (*Student H*, *Student C* and *Student D*) identify the error in Darragh's arithmetic manipulation, but this does not have a remarkable impact in their evaluations. It may be that these students see that Darragh's idea is appropriate and that the error is not fatal. The other students, seeing that the idea is good, may not even read Darragh's manipulations. This indicates that to these students, like most of the other participants, the relation between intended proof and purposes of proofs E_{IP} is more significant than the actual implementation (relations E_{AI} and E_{AP}). This raises the question whether some students (unconsciously) decide not to be concerned about E_{AI} and E_{AP} if they are completely satisfied by E_{IP} and by the author's intention. This issue will be discussed in Section 8.4.2.

In the view of the *Experienced Evaluator* Darragh's style of introducing his general even and odd numbers can be criticized. This objection does not appear in the transcripts.

All participating students seem to consider some purposes of proofs in Darragh's approach. The transcripts indicate that they consider *verification*, in particular generality, *communication*, *explanation* (i.e. the understanding of **why** the proved mathematical statement is true) and also sufficiency without excess in their evaluations of Darragh's proof.

A significant range of features and purposes of proofs have been recognized by the participants. However, only one or at most two of them are mentioned by any individual student. The various reasons for approving Darragh's approach spread evenly in the table above. This might indicate that knowledge about the nature of mathematical proofs is only *visible* in parts to individual students and could be improved through evaluation practice in groups.

I.5. Student evaluations of *Elaine's* answer.

I.5.1. Transcript excerpts.

Student A thinks Elaine's answer is similar to Darragh's. However, she prefers Elaine's because

> *"everything she did is right, but she should give examples to complete the proof".*

Student A checks Elaine's answer line by line, and concludes: *"that's right"*. [*Student A* speaks very fast, her comments are difficult to transcribe.]

***Student B*:**

> *Student B*: *"Well, ahm, I think, would you put Darragh's and Elaine's together, it would be quite good. But about Elaine's I miss, what Darragh has, this x, well I mean, this includes that you can odd + odd, even + even, well that you know, this here is for 'plus', but as well for 'times'. I'm missing the multiplication as well as in all other answers, that when you multiply something out, it is, well odd and odd, odd times odd is odd, even times even is even. Because that is actually, what the statement is saying. And, ah, I mean, Darragh has answered this quite well, but Elaine's is, ahm — well I don't know, somehow I'm missing something. I think it's all very, how do you say it, confusing, that ahm —.*
> [*Student B* reads quietly: *'Number* 2 *is even —'*]
> *For example if she is saying here 'dividing by* 2 *we find that $x^2 + y^2$ is odd', ahm, but you don't see it. Well you don't see, that it is odd. Because, what means odd for me is this $2x + 1$. And here you don't see anything about* 1*, well here it is so, ahm, I mean she says, ok if we divide this by* 2*, ok it is odd, but she does not say, why. She does not have a <u>proof</u> for this, well it is so, I mean, you see that this here is even, this number* 2 *and then $x^2 + y^2$, that this is even, you can see. But then she states an assumption, and you don't see, why."*
> *I*: Would you otherwise say, the answer is correct?
> *Student B*: *"Well, I think the approach is a bit strange, but —, well I can't really say I think it's correct or I think it's incorrect, because I did a completely different approach. But, ahm, well I would say she is, well she is good, but she hasn't thought it through until the end/ hasn't completed thinking it through."*
> *I*: Would you miss examples again?
> *Student B*: *"Well, most of all I would miss an explanation."*

[Translated from German:

> *Student B*: *"Also, ähm, ich find das hier, ich mein, wenn man Darragh's und Elaine's zusammen packen würde, wäre das wahrscheinlich ziemlich gut. Aber bei Elaine fehlt mir das, was Darragh hat, also dieses x, also ich mein, hier ist drin, daß man halt odd + odd, even + even, also daß man weiß, hier ist das für 'plus', aber genauso für 'mal'. Die Multiplikation fehlt mir genau wie in allen anderen Antworten, daß wenn man was multipliziert, ist es, also odd und odd, odd times odd is odd, even times even is even. Weil das ist ja eigentlich, was das Statement aussagt. Und, äh, ich mein, Darragh hat das wohl so gut beantwortet, aber bei Elaine ist das, ähm — ja, ich weiß nicht, irgendwie, ähm, fehlt mir da was. Ich find das alles sehr, wie heißt das, verwirrend, daß ähm —.*
> [*Student B* liest leise: *'Number* 2 *is even —'*]
> *Ja, zum Beispiel wenn sie hier sagt, 'dividing by* 2 *we find that* $x^2 + y^2$ *is odd', ähm, aber das sieht man nicht. Also man sieht nicht, daß es ungerade ist. Weil, was für mich dieses ungerade ausdrückt, ist dieses* $2x + 1$. *Und hier sieht man nichts von* 1, *also hier ist das so, ähm, ich mein sie sagt, ok, wenn wir das durch* 2 *teilen, ok, ist es ungerade, aber die sagt nicht, warum. Die hat dafür keinen* <u>Beweis</u>, *also das ist so, ich mein, ok, man sieht, daß das gerade ist hier, also diese Nummer* 2 *und dann* $x^2 + y^2$, *daß das gerade ist, sieht man. Aber dann stellt sie eine Behauptung auf, und man sieht gar nicht, warum."*
> *I*: Würden Sie sonst sagen, die Antwort ist richtig?
> *Student B*: *"Ja, ich find den Ansatz ein bißchen komisch, aber — ja, es ist komisch zu sagen, ich find's richtig oder ich find's falsch, weil ich halt einen komplett anderen Ansatz gemacht habe. Aber, ähm, also ich würde sagen, sie ist, also sie ist wohl gut, aber sie ist nicht bis zum Ende ausgedacht."*
> *I*: Würden sie wieder Beispiele vermissen?
> *Student B*: *" Ja, ich würde vor allem eine Erklärung vermissen."*]

***Students C/D*:**

> *Student C*: [Long pause (2 minutes)] *"I think, it's alright."*
> *Student D*: *"I don't know."*
> [The students check the argument carefully.]
> *Student D*: *"How does she know, this* $(x^2 + y^2)$ *is odd?"*
> *Student C*: *"Because* $x + y$ *is odd, and then she squares, and it's still odd."*
> *Student D*: *"*$x^2 + y^2$ *could still be even."*
> *I*: Is the whole proof wrong?
> *Student D*: *"She is only assuming, that this is odd"*
> *I*: What advice would you give Elaine?
> *Student C*: *"She can't just assume that* $(x^2 + y^2)$ *is odd."*
> *I*: Is everything wrong?
> *Student C*: *"It's right up to there"* [refers to 'obviously'].

***Student E*:**

Student E: [Pause] *"It's fine up to here"* [refers to 'obviously'].
Student E: *"And then here, what does she mean here* (...)*? Why is that* [refers to '$x^2 + y^2$'] *odd?"*
Student E: *"And here she is saying 'x^2 and y^2 can't both be odd.' 'So' she says 'So x^2 is even and y^2 is odd'. But why is x^2 even? Couldn't y^2 be even and x^2 odd or vice versa?"*
Student E: *"I don't really know if that's what she means, maybe I can't see it. She has the volume by* 2*, ok, we have $x^2 + y^2$. How does she get that? —"* [laughs] *"I don't know. Ahm. I don't think it is very good as a proof."*
I: Do you think it's correct?
Student E: *"Hmmm. No. I don't think it's correct."*
Student E: *"Here like, just, even there, she just took x^2 to be the even one, but couldn't y^2 be even? So, it's — it's not, no, I don't think, it's correct."*

***Student F*:**

Student F: [Long pause] *"I'm really sorry."*
Student F: *"I suppose it's grand, there is nothing really wrong with it."*
I: What do you like about this answer?
Student F: *"It shows in another way that even numbers squared are even and odd numbers squared are odd."*

***Student G*:**

Student G: [Pause] *"'Yeah, it's good, too." "It's more kind of — a little bit more kind of confusing. Because she's got two odd numbers here and is given any even number."* (...) [Pause]
Student G: *"I don't think this is too clear here.* (...) *$x^2 + y^2$ might not be odd. This could be 3 and 1 —"* (...?)
I: What advice would you give Elaine? Is she completely wrong?
Student G: *"She took a good approach to it. I think towards the end she's kind of tripping up a little bit."*
I: Did she lose track?
Student G: *"Well, no, it kind of — she took a good approach"* (...?)

Student G's advice to Elaine: *"She needs to be more careful or what."*

***Student H*:**

> *Student H*: [Pause] *"It's a lot more complicated I think than the others. Ahm — It does show that the* $2x$ *or* 2 *one* $x^2 + y^2$ *are even, cause it's originally, anything squares I think is —, but then — Because it's multiplied by* 2 *it's going to be even. But when it's not multiplied by* 2 *it can be odd, because* $x + y$ *is squared, that makes sense. And then she just decides that if, one of them has to be odd, and one of them has to be even in order for it to be an odd number. So, then she just says that* x^2 *is even and* y *is odd. It's a good answer, it's a long one. I think it's a lot more complicated than the other ones. It makes sense."*

A	***B***	***C/D***	***E***	***F***	***G***	***H***
C Pr NoEx	C? Conf/GAppr Pr? No Expl AffNotJust	C→PC AffNotJust	PC→NotC NoPr AffNotJust	C Pr	PC−MMEnd VG	 VG
D-E:S E(D)					MConf (ABCD)	MCompl (ABCD)
3^{rd}	3^{rd}	5^{th}	6^{th}	3^{rd}	2^{nd}	3^{rd}

Table 8.5: Coding table - Evaluations of *Elaine's* answer

I.5.2. Interpretation. It is striking that only one student (*Student E*) shares the view of the representative *Experienced Evaluator* and decides that Elaine's answer is wrong and definitely not a sufficient proof of the statement. Consequently he ranks it last out of the six proposed answers.
Four students (*Student B*, *Student C*, *Student D* and *Student G*) regard Elaine's approach as valuable, but not fully sufficient.

- *Student B* describes her difficulty in deciding whether Elaine's approach is correct or not, considering that she would have taken a different approach. In the end she arrives at the decision to call the approach 'good' and ranks it relatively highly (3rd). *Student B* criticizes the deficit of explanations.
- *Student C* and *Student D* do not find any logical mistake in the first few lines of Elaine's answer and therefore value that part of the proof as correct: *"It's right up to there"*. Like *Student B* and *Student E* they criticize that Elaine makes assertions without justifying them. *Student C* and *Student D* regard this as a major error and rank Elaine's answer second last (5th).
- *Student G* on the other hand, even though she recognizes one of the logical mistakes and finds the answer more *"confusing"* than the others, appreciates the approach and ranks Elaine's answer relatively highly (2nd). Unfortunately she does not explain her reasons for her positive evaluation. She interprets the mistake as minor:

  > *"I think towards the end she kind of tripping up a little bit."*

The remaining three students (*Student A*, *Student F* and *Student H*) regard Elaine's answer as a correct and sufficient proof. However, two of them criticize it slightly:

- *Student A* criticizes a lack of examples.
- *Student H* disapproves that the argument is in comparison to the others *"more complicated"*.

Student F does not find any logical mistakes which is the reason why he regards this answer as correct. All three of these students rank Elaine's answer third.

None of the students seem to question Elaine's proof method or idea, in particular the curious move of introducing the numbers $x + y$ and $x - y$ and looking at their squares. They all read the answer line by line. Those who find one of the logical mistakes weigh this differently in their evaluations. All students, except *Student B*, seem to presuppose that the intended proof meets its purposes without verifying this, they do not consider E_{IP}. (*Student B* says *"the approach is a bit strange"* and seems a bit sceptical about it. This may be interpreted as a weak approach to consider E_{IP}. However, her scepticism is not strong enough to prompt her to investigate E_{AP} expecting to find errors. She only goes so far as to criticize a lack of explanation and is not prepared to say the proof is wrong.) The other students seem content with relating the actual character of Elaine's answer with her assumed intention and just evaluate E_{AI}, however differently. (*Student E*'s rejection of Elaine's answer seems to be mostly on the basis of E_{AP} and maybe E_{AI}, he does not seem to consider whether there might be a valuable intention somewhere.)

Four of the students criticize that Elaine makes assertions without justifications. Of these students, some ask for reasoning of correct assertions and some for reasoning of incorrect assertions. Sufficient reasoning within a proof seems to be important to those four students.

I.6. Student evaluations of *Fintan's* answer.

I.6.1. Transcript excerpts.

Student A*: *"Fintan and Elaine are almost similar, but he has a different method"
[*Student A*'s description of these differences are impossible to understand. Therefore this part of the interview has not been transcribed.]
Student A summarizes: no examples, but it is correct.

***Student B*:**

> *"It's the best. — I'd like to see formula instead of the last few sentences. He proves the statement for products."*

[Translated from German:

> *"Ist die beste. — Ich hätte gerne eine Formel statt den letzten Sätzen. Er beweist die Aussage für Produkte."*]

Again *Student B* criticizes the deficiency of examples in Fintan's answer.

***Students C/D*:**

> *Student D*: *"It's kind of like Darragh's."*
> *Student D*: *"It works."*
> *Student C*: *"Ya"*
> *Student C*: *"Fintan's is better."*
> *I*: Why is Fintan's better?
> *Student D*: *"He has explanations with it."*
> *Student C*: *"Backing it up."*
> *Student D*: *"Ya, he is backing it up."*
> *Student C*: *"He is explaining what's happening."*
> *Student C*: *"He's shown that he's understanding."*
> *Student D*: *"Darragh is very to the point, he has not written explanations with it."*

The interviewer asks, whether Darragh and Fintan do "exactly the same thing and write it down differently". In their answer *Student C* and *Student D* express the opinion that Darragh and Fintan are using the same method, but different variables, and Fintan explains more.
[This last part was very difficult to transcribe.]

***Student E*:**

> *Student E*: [Pause] *"Ya, he is good."*
> *I*: Advice?
> *Student E*: *"I think it's good. It's fine."*

Commenting on his ranking ***Student E*** explains why he prefers Fintan's to Darragh's answer:

> *Student E*: *"Ahm — They are very close. Darragh should put the variables together. Just because it, ahm, they kind of join the odd and the even together, not distinct them. The relationship between odd and even. Fintan has the relationship —*
> *I*: Do Darragh and Fintan use the same method?
> *Student E*: *"Ya, they use the same method."*
> *I*: And prove the same fact?
> *Student E*: *"Ahm"* - [meaning 'yes']

***Student F*:**

> *Student F*: [Pause] *"I think that's ok."*

***Student G*:**
[Interruption – Pause]

> *Student G*: *"He's not really squaring, is he? He's just multiplying it. I wouldn't really go that way. (...)"*
> *I*: Do you think, Fintan's approach would answer our question?
> *Student G*: *"No."*

***Student H*:**

> *Student H*: [Pause] *"He kind of does it, he shows that the product of <u>any</u> two even numbers can be even, and the product of any two odd numbers can be the same, so, ya, if you are squaring an odd number or squaring an even number,—"*
> *"It's a good answer, it's a different take than the rest of them, like."*
> *I*: What's different about it?
> *Student H*: *"Ah, he doesn't try to prove the square of the number as such, he tries to prove that you can take any two even numbers and multiply them, and you'll end up with an even number, the answer has to be even. So that means when you square an even number you will get an even number, and the same for the odd. — He just does it a different way."*
> *I*: Does it prove the statement?
> *Student H*: *"Oh ya, because he has proven that if you square it could be even."*

A	***B***	***C/D***	***E***	***F***	***G***	***H***
C NoEx	C MPr NoEx	C	VG	VG	NC	VG MPr
F-E:S (almost)	F(all)	F-D:S F(D) F:ExplWhy(D)	F-D:S F(D) F:ExplWhy(D)	F-D:S D(F) D:MSucc(F)	F(all)	
4^{th}	1^{st}	1^{st}	1^{st}	2^{nd}	5^{th} jointly	1^{st}

Table 8.6: Coding table - Students' evaluations of *Fintan's* answer

I.6.2. Interpretation. Only one of the students (*Student G*) regards Fintan's proof as not correct. In *Student G*'s opinion this proof fails E_{IP} as it does not meet the purpose *verification.* (She does not see any *"squaring"* in Fintan's answer and concludes that it must be wrong.) Consequently she ranks Fintan's answer last, together with Anna's.

The remaining seven students agree that Fintan's answer is sufficient. However, one of these students (*Student A*) does not seem to understand Fintan's answer. This is indicated by the fact that *Student A* thinks Fintan's approach is *"almost similar"* to Elaine's. This student seems to be content with her considerations on E_{AP}, with the purpose *communication* in mind. As with most of the previous approaches, she criticizes the deficit of examples in this answer. *Student A* ranks Fintan's answer fourth, after Anna's, Benny's and Elaine's.

The other six students seem to understand Fintan's answer. The fact that they take some time to read it carefully suggests that they consider E_{AI} and E_{AP} as well as E_{IP}. They all regard Fintan's answer as sufficient. Five of them decide that Fintan's is the best of the proposed answers, and one student (*Student F*) ranks Fintan's answer second.

- Two students (*Student B* and *Student H*) observe approvingly, as suggested in the view of the *Experienced Evaluator*, that Fintan has proved a more general statement:

  > *"He proves the statement for products."* (*Student B*)
  > *"He tries to prove that you can take any two even numbers and multiply them, and you'll end up with an even number, the answer has to be even. So that means when you square an even number you will get an even number, and the same for the odd."* (*Student H*)

  This indicates that they, unlike their peer students in this instance, give explicit attention to the proving method or idea in addition to other features of the answers. *Student E* and *Student H* appreciate that Fintan's proof shows some attempts to relate the statement to a wider mathematical context. They appear to consider De Villiers' suggested function *systematization* as one purpose of proof.

- Four of the students regard Fintan's and Darragh's answers as quite similar and compare them as follows.

- *Student F* prefers Darragh's because he finds it more concise:

  > *"Darragh is more straightforward, Fintan's is very dragged out, in my opinion.* (...) *Darragh just says exactly what, you know, it's short and still proves it. That one you have to read a lot more."*

- The three other students (*Student C*, *Student D* and *Student E*) prefer Fintan's answer as they appreciate its clarifying and explaining value, though in different ways.

  * *Student C* and *Student D* like Darragh's answer because he *"has explanations with it"* (*Student D*) and *"is explaining what's happening"* (*Student C*).

  * *Student C* adds that *"he's shown that he's understanding"*.

  * *Student E* acknowledges that Fintan's approach illustrates the *"relationship"* between even and odd numbers.

    > *"They are very close. Darragh should put the variables together. Just because* (...) *they kind of join the odd and the even together, not distinct them. The relationship between odd and even. Fintan has the relationship.*

    Student E considers how the proof is related to the mathematical contents of the statement. This quote shows that he considers the purpose *communication* as suggested by De Villiers.

• *Student B* criticizes that Fintan's answer does not include any examples. Like *Student A* she seems to consider a verification, *that* the statement is true (rather than why) as a purpose of proof.

Again a variety of acknowledged features and purposes of mathematical proofs can be identified in the students' opinions on Fintan's answer, namely clearness of the proving method, sufficiency without excess, explanation of <u>why</u> a statement is true, explanation of <u>that</u> a statement is true and how clearly the contents of the proof are connected to the statement. In De Villier's terms, these evaluations indicate that the participating students consider *verification*, *explanation*, *communication* and *systematization* as purposes of proofs. The students' choices of appreciated features and purposes of proof as evaluation criteria differ remarkably and are quite evenly spread between the eight students.

Students' Ranking of answers on Task I

	A	***B***	***C, D***	***E***	***F***	***G***	***H***
1.	Anna	Fintan	Fintan	Fintan	Darragh	Darragh	Fintan
2.	Benny	Darragh	Darragh	Darragh	Fintan	Elaine	Anna
3.	Elaine	Elaine	Anna	Ciara	Elaine	Ciara	Elaine
4.	Fintan	Anna	Benny	Benny Anna	Anna Benny	Benny	Darragh
5.	Darragh	Ciara	Elaine			Fintan Anna	Benny
6.	Ciara	Benny	Ciara	Elaine	Ciara		Ciara

Table 8.7: Students' rankings of the six proposed 'proofs' of *Statement I*

The following bulletpoints outline some of the students' considerations when they were ranking the proposed proofs of *Statement I*.

- ***Student B*** places
  - Elaine 3^{rd}: *"Maybe ok for evens, but not for odds. Not finished."* [Translated from German: *"Vielleicht für gerade ok, aber nicht für ungerade. Nicht fertig."*]
  - Anna 4^{th}: *"not a proof"* [Translated from German: *"kein Beweis"*]
  - Ciara 5^{th}: *"graphical, but that's about it, no explanation, no nothing"* [Translated from German: *"anschaulich, aber nichts dabei, keine Erklärung, kein nichts"*]
  - Benny 6^{th}: *"especially not a proof, but no answer either"* [Translated from German: *"ist vor allem kein Beweis, aber auch keine Antwort"*]

- ***Student E*** prefers Fintan's to Darragh's answer:

  I: Why is Fintan's better than Darragh's?
  Student B: *"Ahm — They are very closed. Darragh should put the variables together. Just because it, ahm, they kind of join the odd and the even together, not distinct them. The relationship between odd and even. Fintan has the relationship — "*
  I: Do Darragh and Fintan use the same method?
  Student B: *"Ya, they use the same method."*
  I: And prove the same fact?
  Student B: *"Ahm"* - [meaning 'yes']

- ***Student G*** comments on her ranking of answers on **Task I**:
  - *"Darragh's is the best."*
  - *"Elaine did good. Just the end that took her up."*
  - [about Ciara's answer:] *"I know there is no theory behind it, but I like that."*
  - [about Fintan's answer:] *"I don't see that very much at all."*
  - *"I give Benny, then Fintan. And Anna's."*
- ***Student F*** comments on his ranking of answers on **Task I**:
  - *"I like these answers* [Darragh's, Fintan's and Elaine's]*, because they are general and are for any numbers. The first ones are only about a certain amount of numbers, if someone is going to be picky about it."*
  - *"I like Anna's and Benny's equally, I suppose."*
  - *"Ciara's is worst, because it's only one example anyway, really."*
  - *"Darragh I like most - probably, then Fintan, then Elaine."*
  - *Student F*: *"Darragh is more straightforward, Fintan's is very dragged out, in my opinion."*
    I: What does that mean, 'dragged out'?
    Student F: *"That's a — Darragh just says exactly what, you know, it's short and still proves it. That one you have to read a lot more."*
    I: Do they use the same method?
    Student F: *"No, I don't think so. Wait, no — It is, it's just, ah. They are kind of similar, but they wouldn't be the exact same. It's like, they have the same sort of generalisation for odd numbers and even numbers."*
  - *" Elaine's answer I find confusing myself, that's just personally."*

To determine an **average ranking of the eight students** points were assigned in relation to each ranking place:

5 points – first place in students' ranking,

4 points – second place in students' ranking,

3 points – third place in students' ranking,

2 points – fourth place in students' ranking,

1 point – fifth place in students' ranking,

0 points – sixth place in students' ranking.

The distribution of points in the students' rankings is listed in Table 8.8[10] below.

Students:	$5pts$	$4pts$	$3pts$	$2pts$	$1pt$	$0pts$
A	Anna	Benny	Elaine	Fintan	Darragh	Ciara
B	Fintan	Darragh	Elaine	Anna	Ciara	Benny
C/D	Fintan	Darragh	Anna	Benny	Elaine	Ciara
E	Fintan	Darragh	Ciara	(Benny	Anna)	Elaine
F	Darragh	Fintan	Elaine	(Anna	Benny)	Ciara
G	Darragh	Elaine	Ciara	Benny	(Fintan	Anna)
H	Fintan	Anna	Elaine	Darragh	Benny	Ciara

Table 8.8: Distribution of students' rankings of the 'proofs' of *Statement I*

To determine an average ranking value the numbers of points for each answer are added:
Anna's answer: $1 \times 5pts + 1 \times 4pts + 1 \times 3pts + 1 \times 2pts + 2 \times 1.5pts + 1 \times 0.5pts = 17.5pts$
Benny's answer: $1 \times 4pts + 2 \times 2pts + 2 \times 1.5pts + 1 \times 1pt + 1 \times 0pts = 12pts$
Ciara's answer: $2 \times 3pts + 1 \times 1pt + 4 \times 0pts = 7pts$
Darragh's answer: $2 \times 5pts + 3 \times 4pts + 1 \times 2pts + 1 \times 1pt = 25pts$
Elaine's answer: $1 \times 4pts + 4 \times 3pts + 1 \times 1pt + 1 \times 0pts = 17pts$
Fintan's answer: $4 \times 5pts + 1 \times 4pts + 1 \times 2pts + 1 \times 0.5pt = 26.5pts$

Overall the eight students ranked the proposed answers as follows.
They valued Fintan's ($26.5pts$) and Darragh's ($25pts$) answers best, followed by Anna's ($17.5pts$) and Elaine's ($17pts$), then Benny's ($12pts$), and Ciara's answer worst ($7pts$).

[10] Two answers ranked equally are labelled in brackets and get the average of the shared points.

Interpretation. It is significant, that no two of the students rank the six proposed proofs in the same way. An experienced evaluator would probably rank the proposed proofs as follows:

1. Fintan, 2. Darragh, 3. Ciara, 4. Anna, 5. Benny, 6. Elaine.

Only one of the students (*Student E*) ranks the answers similarly. The diversity of rankings encourages the suggestion of a wide diversity of criteria and purposes of proofs indicated by the students' evaluations. The evaluation transcripts above provide some insight into which criteria and purposes of proofs the students consider relevant in their proof evaluations. The ranking outcomes provide a concise overview and prompt the following additional observations.

- In *Student A*'s evaluation it appears most relevant **how the content of the statement is explained** by the proposed answer. Consequently she favours answers consisting of examples such as Anna's (in her view) and Benny's. Ciara's answer does not reinforce *Student A*'s understanding of the statement and is therefore ranked last by this student.
  Comprehensibility also plays a (minor) role in *Student B*'s evaluations. She comments on Darragh's approach:

  > *"More words are lacking here, like 'if' und 'then', probably an example to make it more perceptible."*

  The reasoning as well as the statement itself should be made *"perceptible"*, which is indicated by *Student B*'s asking for more examples.

- **Generality** of a proof is an important evaluation criterion to most of the students. Just two students (*Student A* and *Student G*) do not mention generality in their comments to the proposed answers. The others all rank Darragh's, Fintan's and Elaine's answers relatively highly, considering these answers to be more general than the others. To *Student H* generality plays a role in her evaluation, but is less important than to the others. She ranks Anna's answer second, even though she considers that this answer does not prove the statement in general. Benny's answer which she regards as more general than Anna's ranks second last. Clearness and explanation of mathematical contents seem to matter more in a proof to *Student H* than generality.

- Five students (*Student B*, *Student C*, *Student D*, *Student E* and *Student F*) mention at some point during the interviews that they appreciate **considerations and explanations about why the statement is true**, in certain proposed answers. Attempts to enhance the reader's understanding in a proof are reasons why *Student C*, *Student D* and *Student E* rank Fintan's answer first. *Student F* prefers Anna's answer to Benny's for the reason that

  > *"she says why the squares are odd, because they end in that. He just presumes that they are odd numbers."*

  Student B prefers Darragh's answer to those answers she had seen before (Anna's, Benny's and Ciara's) as she finds this answer

> *"better than the other ones (...) because you see the mathematical way, why those even and those odd."*

Student E and *Student B* prefer Fintan's to Darragh's for the reason that Fintan's answer makes it easier for the reader to understand why the statement is true. This argument is presented by both as a reason why they criticize Darragh's use of two variables.

> *Student B*: *"Why did he choose k and l instead of just one letter? It looks a bit short. I would have expected more, more thoughts."*

> *Student E*: *"Darragh should put the variables together. Just because (...) they (...) join the odd and the even together, not distinct them. The relationship between odd and even. Fintan has the relationship."*

- Some of the students take into account whether the proposed proof **emphasizes some mathematical contents or general patterns**. Three students mention these aspects as reasons for preferring Anna's to Benny's answer.

  > *"She has this — with the endings."*(*Students C/D*).

  > *"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."* (*Student E*).

  Student E prefers Ciara's answer to those he had seen before because *"there is more thought in it"*, which also indicates an appreciation of attempts to find generalisations and mathematical context in a proof. *Student H* struggles to describe similar reasons why she likes Darragh's answer.

  > *"I think it's more thought — It's more of a — what do you call it — ahm — like, ah, using like functions or using, rather than examples, it's using more of a (...?). Can't think of the word. It's a good answer, though. It makes sense."*

- Some students appreciate a **didactical value** in a proof, which includes how well a reader's interest is caught or how well both statement and proof are being explained to the reader. This didactical value may also include how a slightly unusual approach or viewpoint might be valued as in Ciara's answer – an alternative way of representing mathematical information may be appreciated. This aspect seems to play a major role to *Student G*, who for that reason, ranks Ciara's answer third although she interprets this answer as just one example.

  > *"I know there is no theory behind it, but I like that."*

  The requiring of explanations and verifications of the statement by *Student A* and *Student E* might indicate appreciation of didactical values in a proof as well.

The **proof idea or method** did not seem to play a significant role in the students' proof evaluations, which is indicated by three observed phenomena. Firstly, Ciara's answer was liked least considering the ranking of all eight students together, with only 7 points. The intrinsic idea behind this approach seemed unimportant for the students. Secondly, although Fintan's argument uses the same basic method as Darragh's, the fact that Fintan recognized that the method applies more generally and proved the statement as a special case of a more general fact, is noticed by only two of the students. The third surprising fact indicating poor appreciation of proof ideas or methods is the relatively high ranking score of Elaine's answer (17 points, similar to Anna's, favoured over Benny's and favoured by far over Ciara's approach). None of the students questioned the method of this proof.

Student Evaluations of 'proofs' of *Statement II*

II.1. Student evaluations of Gerard's answer.

II.1.1. Transcript excerpts.

***Student A*:**

> [very quickly] *"Yes, he proves the statement, but doesn't explain anything. There is no explanation."*
> *"It does prove the statement in equation, but doesn't finish in words."*

***Student B*:**

> *"I'd have preferred to see a graph, as — as A.* [a lecturer at NUI Galway] *likes to see a graph. — (...) just an example, it's not a proof."*

[Translated from German:

> *"Ich hätte lieber 'nen Graph gesehen, weil — weil A. immer gern 'nen Graphen sieht. — (...) nur 'n Beispiel, es ist kein Beweis."*]

***Students C/D*:**

> *Student C*: *"He is not like doing the general case, he is probably trying to get his head around it. He needs to develop it more."*
> *Student C*'s advice to Gerard: *"Generalize"*.

***Student E*:**

> *Student E*: *"This is just picking one, one quadratic equation, proving it for one quadratic equation. It's a bit like Benny's from the previous example. — I wouldn't give it a very high mark."*

***Student F*:**

> *Student F*: *"Ah ya – ok. — It's a better answer than what I would have given. I'm not very good at proving things to be honest."*
> *I*: Would you, being a teacher, be happy with Gerard's answer?
> *Student F*: *"I would now, but I will probably end up changing my mind after seeing the rest."*

***Student G*:**

> *"Ya, ahm — It's good, isn't it?"*

***Student H*:**

Student H: [Pause – sighs] *"Ya, it makes sense. It shows it can only have two solutions, can't have any more."*
I: So would you be happy with that answer? Would you give Gerard some advice?
Student H: *"Maybe, he just kind of did it by example. Maybe have some wording saying where he is getting stuff or (...?)"*

A	***B***	***C/D***	***E***	***F***	***G***	***H***
				VG	VG	VG
Pr NoExpl	NoPr Ex MGr	NoPr NoGen	NoPr Ex	(VG)→(NoGen)		 Ex
			G-B:S			
				R-H:MGen(G)	R-J(G)	
2^{nd}	5^{th} jointly	4^{th}	5^{th}	4^{th}	2^{nd}	$3rd$

Table 8.9: Coding table - Students' evaluations of *Gerard's* answer

II.1.2. Interpretation. Seven out of the eight students are at some point aware that Gerard shows the truth of the statement for just one particular function.

- Four of them (*Student B*, *Student C*, *Student D* and *Student E*) decide immediately that Gerard's answer does not prove the statement in general. They all regard the answer as unsatisfactory and rank it last or second last.
- Two students (*Student F* and *Student G*) change their opinion about Gerard's approach during the interview meeting. Both like this answer when they see it first, only the engagement with other attempts makes them aware of its deficiency. However, they interpret this differently:
  - *Student F* disapproves the lack of generality and ranks Gerard's answer second last.
  - *Student G* ranks this answer second, even after realising that

    "he just subbed in values, subbed in for a and b".
- *Student H* approves Gerard's approach and ranks Gerard's answer third out of five:

  "It makes sense. It shows it can only have two solutions, can't have any more."

  Student H appreciates, as suggested in the view of the *Experienced Evaluator*, that Gerard's analysis of his chosen example does contain an argument that offers a reason for the truth of the statement.

One student (*Student A*) regards Gerard's answer as a sufficient proof of the statement. She only criticizes that Gerard *"doesn't explain anything"*. The fact that *Student A* ranks Gerard's approach highly (second), intensifies the impression that explanation of the content of the statement seems to be the most considerable purpose of proof to this student.

Most of the students see Gerard's answer as just one example. In their view it fails on E_{IP} and E_{AP} on the basis of a lack of generality. Gerard's intented proof consists of one example, such as Benny's on *Statement I* and as such it fails E_{IP} as it is not sufficient to establish the truth of the statement. However, other than Benny's answer, Gerard's answer does contain a good explanation and is therefore better regarding E_{AP} than E_{IP}. *Student H* is the only student who gives some attention to this fact. The majority of students do not seem to consider a distinction between E_{IP} and E_{AP} in their evaluations of Gerard's answer.

II.2. Student evaluations of Helena's answer.

II.2.1. Transcript excerpts.

***Student A*:**

> *"Helena's answer is right but — it's only words. There is no mathematical equation in it or examples or anything. She explains everything in words, but then you need examples to back up the words."*

[It is not clear whether *Student A* understands Helena's approach.]

***Student B*:**

> *"Ahm, yes, a good response line. But, ahm, what you are asked to do is 'show this'. What Helena didn't do, in my opinion. Therefore, I mean, it is well explained, ahm —, I mean, she explains, what she would do, but she doesn't do it. And that's, I don't know."*

[Translated from German:

> *"Ähm, ja, guter Antwortsatz. Aber, ähm, das, was man machen soll, heißt ja, 'show', also 'zeige das'. Was Helena nicht gemacht hat, finde ich. Also, ich meine, es ist gut erklärt, ähm —, also ich meine, sie erklärt, was sie machen würde, aber sie macht's nicht. Und das ist, weiß nicht."*]

***Students C/D*:**

> *Student D*: *"She'd have to give an example."*
> [The students discuss that Helena's answer plus example would be fine, e.g. one like Gerard's.]
> *Student C*: *"Even that, ya, to show, that she understands it."*
> *I*: Without example you wouldn't be happy?
> *Student D*: *"Just to back it up better."*
> *I*: What do you mean with 'backing it up'?
> *Student D*: *"To make the proof stronger."*
> *I*: Which means, to make it more understandable?
> *Student C*: *"To show that she herself understands it."*
> *Student D*: *"That's what we always had to do."*

***Student E*:**

> *Student E*: [Pause] *"Ahm — I'm happy with that."*

Later *Student E* ranks Kieran and Helena first, noting their answers would be the *"same thing really"*.

***Student F*:**

> *Student F*: *"It's very much the same as Gerard's, just written down instead. There is nothing wrong with it. — I'd be grand with both. Had to read over that once or twice before I got it, but it's grand."*

During the ranking *Student F* compares Helena's answer with some of the other approaches. He prefers Helena's approach to Gerard's:

> *"Even though it's* [Helena's] *just written down, it's more general"*.

Student F approves Joan's answer in comparison to Helena's:

> *I*: Why do you prefer Joan's to Helena's?
> *Student F*: [Joan's] *"is shown, she is doing the equations herself, not just writing them out."*

Student F also explains why he regards Kieran's approach as better than Helena's:

> *Student F*: *"I prefer doing out general equations."*
> *I*: Why?
> *Student F*: *"That's just the way I do my maths. It's easier to understand if it's written down like that. That would like confuse me."*

***Student G*:**

> *Student G*: *"Good, too. But — I've be kind of — How would a Leaving Cert, do an example, compete in a way like Gerard did? (...)"*
> *Student G*: *"I would be thinking what to expect of a Leaving Cert student —"*
> *I*: Try to forget the Leaving Cert. Think of first year University students.
> *Student G*: *"It's kind of stating what you do. And he"* [Gerard] *"is showing you. — I don't know."*

After *Student G* has ranked the answers the interviewer asked why she ranked Helena's approach second last.

> *Student G*: *"That's the kind of thing what you'd put at the end. She's saying what she is doing but there is nothing to kind of show what she's doing."*

***Student H*:**

> *Student H*: *"Ya, it makes sense, that she says you know that if you have a linear expression put to a quadratic one it will come out as a quadratic equation which can only have two answers. So, whenever you do the derivative of a quadratic function, you always get a linear expression. So, it's true."*
> *I*: What advice would you give Helena?
> *Student H*: *"Maybe give an example after that, to just show it, rather than just writing it."*

In a little conversation about proofs at the end of the meeting *Student H* resumed about Helena's approach again.

> *I*: What about Helena's answer, just words, no formula? Is that not a proof?
> *Student H*: *"That, in a way, that is a proof. But I would have liked to see an example. Although it proves it, it's written, and I think it's kind of, you have to kind of grasp a bit more, before you —. It does prove, but I can see that the other ones probably do prove it, but when you are not used to, even though it's written, you are not used to see it written, you have to think about everything that was written rather than just kind of feel 'oh ya, that makes sense'. You have to think a lot harder."*

A	***B***	***C/D***	***E***	***F***	***G***	***H***
Und? C			VG	C	VG	C
NoForm NoEx	PrNotImpl Expl	Pr NoEx	Pr		NoEx	Pr NoEx
			H-K:S	G-H:S $\downarrow$ R-H:Gen(G) R-(J,K):MImpl(H)	G:MImpl(H)	
3^{rd}	3^{rd}	2^{nd}	1^{st} jointly	3^{rd}	4^{th}	2^{nd}

Table 8.10: Coding table - Students' evaluations of *Helena's* answer

II.2.2. Interpretation. It is striking that the majority of students in this group regard Helena's answer as not fully sufficient. *Student E* is the only student whose evaluation of Helena's approach is similar to the evaluation described in the view of a representative *Experienced Evaluator*. *Student E*'s comparison of Helena's with Kieran's approach (*"same thing really"*) and his ranking (both 1^{st}) indicate that he recognizes and appreciates the proof idea of Helena's answer. *Student E* considers Helena's proof sufficient regarding all three relations E_{AI} – E_{AP}.

The other students treat Helena's written approach as not fully accomplished. Arguments of two kinds are apparent:

- Three of the students (*Student C*, *Student D* and *Student H*) regard Helena's answer as a correct proof, but not sufficiently explained and verified with examples. Abstract justification is not enough to this group of students.
  - *Student C* and *Student D* require examples so that the proof writer can convince the reader of his/her own understanding of the proof. It seems that they think primarily in terms of school exams. The teacher had to be convinced that an answer was not just memorized. *Student C* and

Student D do not seem to imagine that a student could compose a proof in the form of a passage of a text.

- *Student H* on the other hand would like to see an example to make it easier for the reader to understand Helena's proof. In her opinion the reader has *"to think a lot harder"* if a proof is presented in the form of a passage of a text.

All three rank Helena's answer second after Kieran's. In their opinion this proof is not fully sufficient regarding E_{AP} as it does not meet their requirement of comprehensibility.

- *Student B*, *Student F* and *Student G* interpret Helena's approach as description of a proof rather than a proof.
  - *Student B* claims that Helena *"explains, what she would do, but she doesn't do it"*.
  - *Student G* criticizes that Helena is *"saying what she is doing but there is nothing to* (...) *show what she's doing"*.
  - *Student F* favours Joan's and Kieran's answers because in comparison to Helena they are *"doing out general equations"*:

    > [Joan's] *"is shown, she is doing the equations herself, not just writing them out"*.

  This group of students rank Helena's answer third or fourth out of five.

These students seem to interpret Helena's text as a description of an intention that is somehow (in their view) not carried out. They seem to be happy with E_{IP} but not with E_{AI}. In the view of these students, inclusion of some algebraic equations, or formulas, or of algebraic notation of some kind, seem to be essential features of the intended mathematical proof. In comparison to the evaluations of Helena's answer, the same students were happy with Darragh's answer regarding E_{IP}. However, they were also happy with the algebraic appearance of Darragh's proof and approved the proof without considering E_{AI} further. Darragh's typo was nearly invisible to the students. This difference among the students' attitudes towards two proofs which they consider satisfying regarding E_{IP}, and also some of the quotations (i.e. *Student B*'s), support the conjecture that some students might associate mathematical proof with algebraic formalism. The responses on the written evaluation exercises already indicated that students value mathematical formalism highly (see Section 7.3.2). Findings from the interviews regarding this issue are discussed in Section 8.4.4.

In their comments on Elaine's answer, *Student F* and *Student B* (among others) criticize that assertions were made but not justified. Comments on Helena's answer like *"she doesn't do it"* may be interpreted as further criticism of inadequate justifications.

Student A is missing examples and equations in Helena's answer as well. Her evaluation would fit into both of objections described above.

The majority of student evaluators do not seem to accept explanations consisting of general descriptions written entirely in standard text as sufficient proofs. However, their comparison of Helena's and Gerard's answers indicate an appreciation of the generality of Helena's approach. For example this is mentioned explicitly by *Student F* when he compares Helena's with Gerard's answer:

> *"Even though it's* [Helena's] *just written down, it's more general"*.

The majority (all participants except *Student A* and *Student G*) rank Helena's answer more favourably than Gerard's.

II.3. Student evaluations of Ian's answer.

II.3.1. Transcript excerpts.

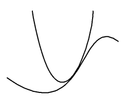
***Student A*:**

> *Student A*: *"It does prove the statement, but there is no explanation here of what he is doing. He just put the graphics and — he has to explain what he did."*
> *I*: Do you know what he is doing, could you explain it to me?
> *Student A*: *"Ya — ya."*

Student A's explanation shows that she does not understand why f' is linear. [The explanation was very difficult to understand and even more to transcribe. Therefore a transcript of this explanation has been omitted.]

> *"I think it depends on the equation she is using."*

Student A interprets Ian's three graphs as some examples. She does not realize that f is always parabolic and f' linear. She regards the three graphs as graphs of three examples and she does not identify the functions. When asked, whether the graphs of f and f' could look like



so that they were more than two common values, *Student A* could not explain why this was not possible.
Student A would like to see some explanations. Her advice to Ian:

> *"that he needs to explain what he is doing because we don't know the functions of the graph and try to explain his concept, because no one knows what he is doing. — He wouldn't get full marks."*

***Student B*:**

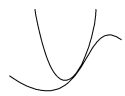
> *"Nice graph, probably also correct, yes, it is correct, but nothing is explained. I mean, it shows the functions, the derivatives, but if I'd get to see this, ahm, well shown, but where are the explanations? Well ok, the question is 'show that ...', but nevertheless it lacks an explanation."*

[Translated from German:

> *"Schöner Graph, wahrscheinlich auch richtig, ja, es ist richtig, aber es ist nichts erklärt. Ich meine, es zeigt die Funktionen, die Ableitungen, aber wenn ich das sehen würde, ähm, ja, gut gezeigt, aber, wo sind die Erklärungen? Also, ok, die Frage heißt ja, 'Zeige, daß ...', aber es fehlt trotzdem die Erklärung."*]

***Students C/D*:**

[First reaction of both students:] *"Oh god"* [They laugh.]
Student D: *"Diagrams aren't enough at all."*
Student D: *"He'd have to explain what he is doing. You can see it, but it's not enough on its own. — Just to back it up."*
I: Do you understand what he is doing?
Student C [laughs]: *"I don't really understand what he is doing."*
I. sketches a graph:



Why doesn't f' look like this?

Student D: *"We know f' is linear."*
Student C: *"He should give a line just to say what he is thinking, —"*
Student C suggests what Ian could say in that line.
[Difficult to transcribe.]
I: With a line like that, would you be happy?
Both: *"Ya, just to back it up."*

When ranking the five proposed answers, *Student D* comments about Ian's answer:

"It's good when you know what he means."

***Student E*:**

Student E: [Pause, laughs] *"Ya, I think it's fine. Well, I think he should mark, write like 'This is the quadratic function', though I suppose he says f there and f' there. — He should explain what he means by this, think, like, see that line, explain that there are two points, or one point, there. He should say, a line can only —, the graph of a quadratic is a parabola, so the line can cross that parabola only twice, at most."*
I: How does he know that f' is a line?
Student E: *"Well, it's supposed to be linear. — It always has to be linear."*
I: Should he say that?
Student E: *"Ya, he should actually. He should write, the graph of a quadratic will always be a parabola, the graph of a linear will always be a line, and a line can only intersect at maximal two points."*
I: And you'd want that to be said?
Student E: *"Yes, that should be said."*

***Student F*:**

> *Student F*: *"Ok — I prefer to see more writing in it to just graphs* (...?)*"*
> *I*: What advice would you give Ian?
> *Student F*: *"I suppose, give an example for each or something to show they are actually true. Because he's expecting you to believe him that this is the only ways it can be done, if you know what I mean. — With examples he'd show better."*

When ranking the answers *Student F* criticizes Ian's approach:

> *"He doesn't do anything for me to believe that this is true."*

***Student G*:**

> *Student G*: *"No, I'd more inclined to go for the other two."*
> *I*: Is the answer wrong?
> [Pause]
> *Student G*: *"I don't think he shows too much."*

***Student H*:**

> *Student H*: [Pause] *"I think it's a bit harder to understand the graphs. I was never very good at graphs. I get that, you know, that f is the quadratic one, and f' would be the linear. Ah — and that the placement of the linear could be different.* [Pause] *Well, I'm not as sure, how it shows that it's only two answers. Because it shows that there is three ways that it can be put, but I don't understand, where it says that there is only two."*

A	***B***	***C/D***	***E***	***F***	***G***	***H***
NoUnd/ I-Ex C NoExpl	 C NoExpl	NoUnd/Und NoExpl	VG C NoExpl	Und? NoEx MExpl	Und? C?	NoUnd C?
		R-Good			GH(I)	
5^{th}	5^{th} jointly	3^{rd}	3^{rd}	5^{th}	5^{th}	5^{th}

Table 8.11: Coding table - Students' evaluations of *Ian's* answer

II.3.2. Interpretation. Three out of the eight students (namely *Student B*, *Student D* and *Student E*) seem to understand Ian's approach. They appear to be happy with this proof regarding E_{AI} and E_{IP}. However, on relation E_{AP} these students regard Ians' answer as not fully sufficient as they disapprove its poor explanation. This criticism weighs differently in these students' evaluations:

- Even though *Student D* and *Student E* criticize the poor attempts to explain this answer, they appreciate the proof idea and rank it third, after Kieran's and Helena's approaches.

- To *Student B* on the other hand, the quality of explanations seems of more importance in a good proof than a correct idea behind it. She ranks Ian's answer last, together with Gerard's.

Student C understands Ian's approach after it is explained by her peer *Student D*. She then agrees with *Student D*'s opinion.

The remaining four students (*Student A*, *Student F*, *Student G* and *Student H*) do not appear to understand Ian's idea for proving the statement.

- *Student A* interprets the graphs as those of three particular functions. She does not recognize the generality of this approach at all. *Student A* is happy with E_{IP}, but not with E_{AP} because in her view the main purpose of proof, comprehensibility, is not met by Ian's proof. This student struggles in her considerations on E_{AI} as she has difficulties to understand Ian's intention.

- *Student F* and *Student G* do not engage in an interpretation of Ian's visual reasoning. They regard Ian's proof as not sufficient on E_{AP} as it is not comprehensible in their views. These students do not show any attempts to understand Ian's intention and consequently do not consider relations E_{AI} and E_{IP}.

  - This is not surprising in the case of *Student F* as he shows similar reactions to Ciara's attempt to prove *Statement I* by means of a visual argument (p. 104). *Student F* seems to expect a proof to be written in a form that makes the details of the argument fully explicit.

    > *Student F*: *"He doesn't do anything for me to believe that this is true."*

  - However, *Student G*'s reaction to Ian's answer is more surprising as she reacted very positively to Ciara's approach (p. 104). The fact that she appreciates Ciara's but not Ian's visual attempt suggests that it is the unusual character rather the visual nature which she approves. (Graphs are visual but familiar to students in their mathematical work.) The interpretation that *Student G* appreciates the didactic value of Ciara's answer seems to be supported. An advantage of an unusual representation is its ability to stimulate a learner's attention and interest, in *Student G*'s opinion.

- *Student H* makes some effort to understand Ian's intention and considers relation E_{AP}. She recognizes the generality of this approach:

  > *"it shows that there is three ways that it can be put"*,

  but does not realize that the intersections of the graphs correspond to the common values of the functions:

  > *"I don't understand, where it says that there is only two"*.

Student H fails to understand Ian's intention and therefore does not consider relations E_{AI} and E_{IP}.

These four students all rank Ian's approach last of the five proposed proofs.

Two students do appreciate the idea of Ian's proof and its potential to prove the statement sufficiently. However, the majority, whether they understand the idea or not, do not value it highly in their evaluations. As in the case of the evaluations of *Statement I*, the proof idea seems to be less important to the students than how it is explained or communicated to a reader. Some students seem to regard a proof as good if it requires minimal engagement of the reader. Proofs that use a style of argument and presentation that is likely to be familiar to these students are more likely to be evaluated favourably. A proof that requires the reader to interpret a manner of representation or take a viewpoint different from the obvious or conventional one is less likely to be ranked favourably, even by students who understand it and believe it to be correct. There also seems to be a reluctance to engage with arguments that look a bit different. This may be understandable and just due to familiarity and experience.
These observations suggest that the purposes *explanation* and *communication* are considered more than *verification* or *intellectual challenge*.

II.4. Student evaluations of Joan's answer.

II.4.1. Transcript excerpts.

Student A likes Joan's answer, though in *Student A*'s opinion it should be explained more, with examples:

> *"Ya, Joan's proof is correct, but she didn't give an example."*

[The rest of the dialogue was difficult to transcribe.]

***Student B*:**

> *"Well, I think, to compare both, I think Joan's answer is pretty good, because above all it is generalized, you can use it for all. And, ahm, well ok, she has done it with this abc- rather the pq-formula, but it shows — that it has just two possible answers. What it doesn't show is, that it can have just one or less. I mean, if the roots are negative, no, if —, well, I think it's actually pretty good."*

[Translated from German:

> *"Also ich finde, um die beiden zu vergleichen, ich finde Joan's answer ziemlich gut, weil das vor allem generalisiert ist, daß man das für alle nehmen kann. Und, ähm, ja ok, sie hat das jetzt wieder mit diesem abc- anstatt von der pq-Formel gemacht, aber es zeigt, — daß es nur zwei mögliche Antworten gibt. Was es allerdings nicht zeigt, ist, daß es auch eine oder weniger geben kann. Ich meine, wenn die Wurzeln negativ werden, nein, wenn —, ja also, ich finde es eigentlich ziemlich gut."*]

***Students C/D*:**

> *Student C*: *"That's what I did."*
> *Student D*: *"works"*
> *Student C*: *"She multiplied by x on the one side, not by the other side."*
> *Student D*: *"Oh yes."*

***Student E*:**

> *Student E*: [Pause] *"Why does she multiply by x?"* [laughs]
> *Student E*: *"She should have left this and — try still to get the roots."*
> *I*: Is Joan on a wrong track or completely wrong?
> *Student E*: *"It's completely wrong."*

In the ranking ***Student E*** places Joan second last, ahead of Gerard's, explaining he would give her

> *"marks for the beginning, because she has a right track"*.

***Student F*:**

> *Student F*: *"That's a good answer. It's general again. I knew I'd change my mind about the first fella. — That's far better than the rest I've seen."*

Student F initially ranks Joan's first and Kieran's second. He comments about Joan's and Kieran's answers:

> *"Between them I'm not sure which would be first."*

With some reflections, *Student F* changes his mind:

> *Student F*: *"I like Joan's and Kieran's the same, because she has done out equations — but actually, ah sorry don't mind it's grand. Ah no, why is she multiplying just one side of the equation? — Ach - I never want to be a lecturer. — I probably prefer his."* [Kieran's]
> [*Student F* swaps Kieran's and Joan's answers.]
> *Student F*: *"I don't really know, why she is multiplying just the one side, no I know, why."*
> *I*: Why?
> *Student F*: *"So that'll leave the equation."* [laughs]
> *I*: Is it ok to do that?
> *Student F*: *"I'm not sure. That's why I'm thinking. I didn't think he could before, but I think I heard some discussion in the workshop, that you could multiply only one side. I don't know what I'm thinking. — I suppose she could say that only that part is multiplied by x."*
> *I*: So you think, the maths is right, but it's not described properly?
> *Student F*: *"I don't think you can just multiply the right side by x. I'm not really sure about that."*

***Student G*:**

> *Student G*: *"Ya, I like that. — I think that's like, very clear."*

When ranking the proposed answers, *Student G* mentions that in comparison to Gerard's answer she regards Joan's as

> *"(...) kind of more theory, he just subbed in values, subbed in for a and b, hers is more like — more kind of theory."*

***Student H*:**

> *Student H*: [Pause] *"I'm not quite sure, how she just multiplies by x, but she only multiplies one side, rather than, if she is multiplying by x, she should be multiplying **everything** by x. — Ahm. So I don't really know how true it is. But if that part is —/ does work out, then ya, it does prove that the answer is + or $-\sqrt{\frac{c}{a}}$. I just wouldn't be quite sure, if she is multiplying that x (...?)"*
> *I*: So you are not sure about this bit, and up to here, the equation, it's fine?

Student H: *"That's fine."*
I: When you say, you are not sure, does that mean you are not sure, or are you saying, this is wrong?
Student H: *"Ah, I would think, it would be wrong, because she did multiply across by x."*

A	***B***	***C/D***	***E***	***F***	***G***	***H***
VG	VG		VG	VG	VG	
C NoEx/ MExpl	Gen	C → NotC Err	NotC Err	Gen R-Err	Cl	NotC GAppr/ Err
				J(GHI)		
			R-GAppr	R-C→NotC	R-J:MTh(G)	
1^{st}	1^{st}	5^{th}	4^{th}	$1^{st} \rightarrow 2^{nd}$	1^{st}	4^{th}

Table 8.12: Coding table - Students' evaluations of *Joan's* answer

II.4.2. Interpretation. It is striking that Joan's answer is regarded either as very good (by half of the students) or very poor (by the other half), the students do not see it in the average range, for example as 'ok' or 'just good enough'. Consequently it is mostly ranked best or worst answer, at most second best or second worst.

- Three of the four students who rank the answer highly do not recognize its error. (*Student F* detects the error in the ranking process, but even then he changes his opinion just slightly.) These students appreciate, that the proof
  - is general:

    "I think Joan's answer is pretty good, because above all it is generalized, you can use it for all." (*Student B*)
    "That's a good answer. It's general again." (*Student F*)

  - or well structured:

    "I think that's (...) very clear." (*Student G*).

  - Three of them further acknowledge that every assertion in her argument seems justified. Evidence is found for example when *Student F* compares Joan's answer to others in the ranking process. *Student F* explains why he prefers Joan's answer to Helena's answer:

    [Joan's] *"is shown, she is doing the equations herself, not just writing them out"*

    or compares it to Kieran's argument:

    "I like Joan's and Kieran's the same, because she has done out equations".

  - In their comments about Kieran's approach *Student A* and *Student B* appreciate that Joan attempts to accomplish her proof in detail:

> *"Joan's answer is more specific than his answer, because she does more equations (...)"* (*Student A*).

Student B thinks that Kieran *"finishes where Joan had begun"*. She regards Kieran's answer as less good than Joan's because she prefers

> *"the quadratic equation to be multiplied out"* .

These students appear to consider relation E_{IP}, that is they relate Joan's assumed intention to several purposes of proof, namely its potential to verify the truth of the statement in general (*verification*), how each step of the reasoning is explained to the reader and therefore provides insight into why the statement is true (*explanation*), and also how the proof is presented to the reader (*communication*). On the other hand, relations E_{AI} and E_{AP}, concerned with how the intention and purposes are implemented in the actual proof, appear to play at most a minor role in the evaluations of Joan's proof for this group of students.

As in the case of Darraghs' answer, it seems that some of the students, once satisfied with the proof idea or intended proof, do not even consider its implementation. Only if they struggle in identifying a proof idea or if they are somehow sceptical about it, they read a proposed proof carefully line by line. This issue is discussed in Section 8.4.2.

- The four students who evaluate Joan's answer as poor, recognize its error. Nevertheless, they all appreciate its inception.

  - *Student C* and *Student D* identify *Student C*'s own (correct) idea to prove the statement:

    > *Student C*: *"That's what I did."*
    > *Student D*: *"works"*

  - *Student E* places Joan's answer second last, before Gerard's, explaining he would give her

    > *"marks for the beginning, because she has a right track"*.

  - *Student H* acknowledges the attempt as well. She would accept it as a valid proof if its implementation were correct:

    > *"I don't really know how true it is. But if that part is —/ does work out, then ya, it does prove that the answer is + or* $-\sqrt{\frac{c}{a}}$*."*

    Student H later arrives at the conclusion, that the implementation is incorrect, and ranks Joan's answer second last, before Ian's.
    It is interesting that even though *Student H* is in doubt about the multiplication step, it does not seem to occur to her to test this doubt by checking if $\pm\sqrt{\frac{c}{a}}$ really are solutions of Joan's original equation. *Student H* eventually decides that Joan's multiplication step is wrong, but not by using her own (correct) logic.

> These students are not immediately convinced by Joan's intention and investigate the proof more carefully. They consider all three relations E_{AI} – E_{AP}. While appreciating the promising proving method in considering E_{IP}, they realize that Joan's approach fails in the implementation, that is regarding E_{AI} and E_{AP}. The evaluation habits as well as the judgments observed in the students' comments are comparable to those described in the view of an *Experienced Evaluator*.

In comparison to Darragh's error, Joan's is more serious. Both of them reach the correct conclusion in terms of the statement, but Joan's $\pm\sqrt{\frac{c}{a}}$ are not the correct solutions of the quadratic equation that she (correctly) starts with. More students notice Joan's error than Darragh's – this may be because the simple nature of Joan's final quadratic and the surprisingly simple form of the roots raises a suspicion that something could be wrong. This suggests that some students even if they are happy with E_{IP}, still look at E_{AP}, because something about the final lines looks somewhat strange and they are a bit sceptical about Joan's proof idea. Another explanation may be that these students at the end of this interview section, initiated by practice of proof evaluation, have learned to consider the idea of a proof in their evaluations.

II.5. Student evaluations of Kieran's answer.

At this stage of the meetings the students seem to get a bit tired of this kind of exercise. Most of them discuss Kieran's answer in less detail than the former suggested proofs. *Student C* does not comment on Kieran's approach at all.

II.5.1. Transcript excerpts.

***Student A*:**

> *"Kieran's answer is similar to Joan's answer. Joan's answer is more specific than his answer, because she does more equations, he didn't explain anything. (...) He didn't do anything about the derivative of f (...)"*

[*Student A* decribes what Kieran is doing.]

> *"At least he should prove with examples to show and explain what he just did. He didn't explain anything."*

***Student B*:**

> *"Well, he has a correct approach, by any means. What you have to think about, that you set them equal, ahm. Well, he finishes where Joan had begun. I think, he has the right approach, but he would have to continue — well as answer to the question I would have somehow, I think, that's not enough. Well, I mean, I would have liked the quadratic equation to be multiplied out. Because 'show' means to convince someone else, that it truly is a fact."*

[Translated from German:

> *"Also, er hat nen richtigen Ansatz, auf jeden Fall. Was man halt überlegen muß, daß man sie halt gleich setzt, ähm. Also er endet da, wo Joan angefangen hat. Ich finde, er hat den richtigen Ansatz, aber er hätte das noch weiterführen — also als Antwort auf die Frage würde ich das irgendwie, ich finde, das ist nicht genug. Also, ich meine, ich hätte dann gerne die quadratische Gleichung ausmultipliziert. Weil 'show' heißt, daß man den anderen überzeugt, daß es wirklich so ist."*]

***Students C/D*:**

> *Student D*: *"Yes that's a good idea."*
> *Student D*: *"That would get full marks."*
> *Student C* does not say anything.

***Student E*:**

Student E: *"I think that's fine."*

***Student F*:**

Student F: *"Ok — ahm — I'd probably show what $f(x)$ was first because the ax^2 comes out randomly, and probably finish the equation there instead of just leaving it like that. Oh, I don't know, I wouldn't be able to do that, but — just kind of lack there —. And he just expects us to believe that at most two solutions, even though, you kind of know that, because it's x^2 actually."*

[*Student F* realizes that he does know a quadratic equation cannot have more than two solutions.]

Student F: *"No, it's grand, just ahm show what $f(x)$ was first."*

When ranking the answers, *Student F* explains why he prefers Kieran's answer to Helena's:

Student F: *"I prefer doing out general equations."*
I: Why?
Student F: *"That's just the way I do my maths. It's easier to understand if it's written down like that. That would like confuse me."*

***Student G*:**

"Yeah, it's good, too. It's kind of –. It's very short."

***Student H*:**

Student H: [Pause] *"It's short and it makes sense. He proves it very quickly, so —"*
I: Would you be perfectly happy with that, would you be the teacher?
Student H: *"Ya, he does exactly what you'd want him to do. He shows what the derivative is, can only have the two answers, by showing that the quadratic equation can only have two answers. And he doesn't just write the ends, he actually says that the quadratic equation which has at most two solutions."*

A	***B***	***C/D***	***E***	***F***	***G***	***H***
		VG	VG		VG	VG
NoEx NoExpl	GAppr NotConvEnough			NoDef Form	(short)	Conc
			H-K:S	J-K:S $\rightarrow$ K(J) $\rightarrow$ K(H)		
4^{th}	2^{nd}	1^{st}	1^{st} jointly	1^{st}	3^{rd}	1^{st}

Table 8.13: Coding table - Students' evaluations of *Kieran's* answer

II.5.2. Interpretation. *Student A* compares the appearance of Kieran's proof to that of Joan's and considers E_{IP}. She relates Kieran's proof to those purposes which (in her view) are most relevant, namely comprehensibility and *communication*. *Student A* criticizes poor explanation and a lack of examples. She neither considers Kieran's intention nor the relations E_{AI} or E_{AP}. *Student A* ranks Kieran's answer second last, only followed by Ian's approach. She is the only student who evaluates Kieran's answer as not sufficient.

The remaining seven students seem to be happy with Kieran's intention, on both relations E_{IP} and E_{AI}.

- Five of them like Kieran's answer best of the proposed answers. They regard Kieran's proof as sufficient on all three relations E_{AI} – E_{AP}.
  - Three students do not explain why they evaluate Kieran's approach so highly.
  - *Student F* appreciates its algebraic appearance:

    > *"It's easier to understand if it's written down like that."*

  - *Student H* approves the fact that Kieran's proof is sufficient without excess:

    > *"It's short and it makes sense.* (...) *he does exactly what you'd want him to do."*

- *Student G* does not seem to be fully satisfied in respect of E_{AP}. She also mentions the shortness of Kieran's answer:

  > "(...) *It's very short."*

  The fact that this student ranks Kieran's approach third and therefore less favourably than the other students indicates that in her case the comment on the shortness of this proof can be interpreted as negative criticism of the comprehensibility of Kieran's answer. Thus, *Student G* considers E_{AP} mainly with the purpose *communication* in mind.

- While *Student B* is also happy in respect of E_{IP} and E_{AI} and appreciates Kieran's approach, she criticizes it on the basis of E_{AP} considering that Kieran (in her view) has not explicitly proved the fact that a quadratic equation has at most two solutions:

> *"I would have liked the quadratic equation to be multiplied out."*

Student B criticizes that Kieran's approach is not as convincing as it should be. Consequently she ranks Kieran's answer second after Joan's, which in her reckoning proves every step explicitly.

Students' Ranking of answers on Task II

	A	***B***	***C, D***	***E***	***F***	***G***	***H***
1^{st}	Joan	Joan	Kieran	Kieran Helena	Kieran	Joan	Kieran
2^{nd}	Gerard	Kieran	Helena		Joan	Gerard	Helena
3^{rd}	Helena	Helena	Ian	Ian	Helena	Kieran	Gerard
4^{th}	Kieran	Ian Gerard	Gerard	Joan	Gerard	Helena	Joan
5^{th}	Ian		Joan	Gerard	Ian	Ian	Ian

Table 8.14: Students' rankings of the five proposed 'proofs' of *Statement II*

The following bulletpoints outline some of the students' considerations when they were ranking the proposed proofs of *Statement II.*

- ***Student E*** placed
  - Kieran and Helena 1^{st}: *"same thing really"*,
  - Joan 4^{th}: *"marks for the beginning, because she has a right track"*,
  - Gerard last: *"doesn't give you any information.* (...) *That is not a proof. I don't think it is relevant for the question."*
- ***Student F*** first ranks the proposed proofs in the following order:

  1. Joan, 2. Kieran, 3. Gerard, 4. Helena, 5. Ian.

  He comments about Joan's and Kieran's answers:

  "Between them I'm not sure which would be first".

  Student F then changes his mind and swaps Gerard's and Helena's answer:

  1. Joan, 2. Kieran, 3. Helena, 4. Gerard, 5. Ian.

  "Even though it's [Helena's] *just written down, it's more general".*

  Student F also compares Joan's answer to Helena's:

  I: Why do you prefer Joan's to Helena's?
  Student F: [Joan's] *"is shown, she is doing the equations herself, not just writing them out."*

  Finally *Student F* considers Joans' and Kieran's answers again:

  "I like Joan's and Kieran's the same, because she has done out equations ... but actually, ah sorry don't mind it's grand. Ah no, why is she multiplying just one side of the equation?"
  "Ach - I never want to be a lecturer"
  "I probably prefer his" [Kieran's].

Consequently *Student F* changes his ranking again:

1. Kieran, 2. Joan, 3. Helena, 4. Gerard, 5. Ian.

> *Student F*: *"I don't really know, why she is multiplying just the one side, no I know, why."*
> *I*: Why?
> *Student F*: *"So that'll leave the equation."* [laughs]
> *I*: Is it ok to do that?
> *Student F*: *"I'm not sure. That's why I'm thinking. I didn't think he could before, but I think I heard some discussion in the workshop, that you could multiply only one side. I don't know what I'm thinking.* (...) *I suppose she could say that only that part is multiplied by x."*
> *I*: So you think, the maths is right, but it's not described properly?
> *Student F*: *"I don't think you can just multiply the right side by x. I'm not really sure about that."*
> *I*: Do you think, all the other answers are correct?
> *Student F*: "(...) *He* [Ian] *doesn't do anything for me to believe that this is true.* (...) *I prefer doing out general equations."*
> *I*: Why?
> *Student F*: *"That's just the way I do my maths. It's easier to understand if it's written down like that. That would like confuse me."*

Student G comments on her ranking of answers to *Task II*:

> *I*: Why Joan's more than Gerard's?
> *Student G*: "(...) *kind of more theory, he just subbed in values, subbed in for a and b, hers is more like* (...) *more kind of theory."*
> *I*: Why Kieran next?
> *Student G*: *"Kieran's is very short."*
> *I*: Why not Helena's?
> *Student G*: *"That's the kind of thing what you'd put at the end. She's saying what she is doing but there is nothing to kind of show what she's doing."*

Similar to the ranking of *Task I* an **average ranking of the eight students** was determined. To do so, points were assigned in relation to each ranking place:

- first place in students' ranking $\sim$ 4 points,
- second place in students' ranking $\sim$ 3 points,
- third place in students' ranking $\sim$ 2 points,
- fourth place in students' ranking $\sim$ 1 points,
- fifth place in students' ranking $\sim$ 0 points.

The distribution of points in the students' rankings is listed in Table 8.15 below.

Student:	$4pts$	$3pts$	$2pts$	$1pt$	$0pts$
A	Joan	Gerard	Helena	Kieran	Ian
B	Joan	Kieran	Helena	(Ian	Gerard)
C/D	Kieran	Helena	Ian	Gerard	Joan
E	(Kieran	Helena)	Ian	Joan	Gerard
F	Kieran	Joan	Helena	Gerard	Ian
G	Joan	Gerard	Kieran	Helena	Ian
H	Kieran	Helena	Gerard	Joan	Ian

Table 8.15: Distribution of students' rankings of the 'proofs' of *Statement II*

To determine an average ranking value the numbers of points for each answer are added:
Gerard's answer: $2 \times 3pts + 1 \times 2pts + 3 \times 1pt + 1 \times 0.5pts + 1 \times 0pts = 11.5pts$
Helena's answer: $1 \times 3.5pts + 3 \times 3pts + 3 \times 2pts + 1 \times 1pt = 19.5pts$
Ian's answer: $3 \times 2pts + 1 \times 0.5pts + 4 \times 0pts = 6.5pts$
Joan's answer: $3 \times 4pts + 1 \times 3pts + 2 \times 1pt + 2 \times 0pts = 17pts$
Kieran's answer: $4 \times 4pts + 1 \times 3.5pts + 1 \times 3pts + 1 \times 2pts + 1 \times 1pt = 25.5pts$

Overall the eight students ranked the proposed answers as follows.

> They valued Kieran's ($25.5pts$) best, followed by Helena's ($19.5pts$) and Joan's ($17pts$), then Gerard's ($11.5pts$), and Ian's answer worst ($6.5pts$).

Interpretation. An *experienced evaluator* might rank the proposed proofs as follows:

Helena's and Kieran's answer first, followed by Ian's.
Gerard's and Joan's answers would probably be ranked last.

The students' rankings do not differ as much from the assumed ranking of experienced evaluators as in the ranking of *Task I*. Nevertheless, some major differences are apparent.

- Firstly, Kieran's answer is evaluated significantly better than Helena's by the students, whereas most experienced evaluators would probably consider both answers as equally good. Only one of the students share that opinion, the majority of them regard Helena's written approach as not fully complete, or as less comprehensible than Kieran's which is expressed in algebraic terms.
- Further notable is the fact that Joan's answer is evaluated surprisingly highly by the students, very close to Helena's, even though it is wrong.

  This might be an effect of different evaluation habits. Where a professional evaluator by and large would consider all three relations $E_{AI} - E_{AP}$ between actual and intended character of a proof and its purpose(s),[11] four of the

[11] A professional mathematician may omit the reading of all the details of a proof when reading a research article on his/her own specialism. However, these habits do not seem to be common when mathematicians validate proofs in a topic they are less familiar with. See Section 8.4.2 for a

students, once satisfied with the proof idea or intended proof in addition to its appearance, do not seem to consider its implementation. Thus, they appear to be content with considerations regarding E_{IP}. In this case, these four students agree that Joan's proof inception is promising regarding the purpose of proof to establish the truth of the statement. Being happy with its appearance as well, they forego a careful investigation of its implementation, that is considerations regarding E_{AI} and/or E_{AP}.

Another reason for some of the students' positive ranking of Joan's approach is their high appreciation of the fact that Joan's answer may be interpreted as including an attempt to prove explicitly that a quadratic equation has at most two solutions. This attempt was noted in the views of the *Experienced Evaluator*, but did not have a significant impact on the suggested evaluations.

- Another interesting difference is the diverse evaluation of Ian's graphical approach. This is regarded relatively unfavourably by most of the participating students, but appreciated much more by mathematicians.

The students' evaluations and ranking of the proposed answers to the second task indicate which criteria and purposes of proofs they consider relevant.

- Again **generality** of a proof is important to most of the students. More than half of them recognize that Gerard's approach is not general and therefore rank it 4^{th} or 5^{th}. Comments on Helena's answer indicate some of the students' appreciation of generality of a proof as well, see page 137.

- Two other students (*Student G* and *Student H*) recognize the lack of generality in Gerard's answer as well, but they appreciate its explanatory value more highly. To them it is important that a proof is **comprehensible to the reader**. The importance of this aspect of a proof in the students' evaluations is also indicated by the fact that Helena's proof written in text was liked noticeably less than Kieran's algebraically written proof.

  Student H: *"You have to think a lot harder."*

  Other indications that some students consider the comprehensibility of a proof to be an important evaluation criterion are their comments on Ian's approach. Two of the students do not show any attempts to try to understand Ian's answer.

- Comparison of the ranking results of the two graphical approaches indicate slightly more importance of **the idea of the proof** in the comments about Ian's answer than in those about Ciara's. Two of the three students who acknowledge Ian's proof idea rank this answer directly after Kieran's and Helena's. On the other hand *Student B*, who understands Ian's idea as well, ranks this answer last - equal to Gerard's whose lack of generality she had identified. In these cases the idea of a proof seems to be of less significance in *Student B*'s evaluation than the quality of explanation.

discussion on this issue.

- The surprisingly high score of Joan's answer implies the importance of **justification of each stated assertion** to the participating students.
- Three students interpret Helena's answer as a description of a proof instead of a proof itself. This intensifies the impression from the written exercises that to the students the **formal appearance** is crucial in a mathematical proof.
- *Student H*'s comments about Kieran's answer show that **sufficiency without excess** in a proof plays a role in her evaluation:

  > *"It's short and it makes sense.* (...) *He does exactly what you'd want him to do."*

  She ranks Kieran's answer first.

8.4 Outcome of the interviews

This section decribes the findings from the interview experiment. The interpretation outcomes are summarized and related to the findings from the written evaluation task and to the schema as suggested in Section 4.2.

Section 8.4.1 outlines **criteria** employed by the students for valuing a proof in their oral proof evaluations. The identified evaluation criteria indicate which **purposes of proofs** the participating students consider relevant. These criteria and purposes are related to those identified in students' written responses to an evaluation task.

Section 8.4.2 is concerned with how the students relate the three facets of artifacts in their proof evaluations. Those observations motivate some hypotheses about the students' **proof validation habits**.

Section 8.4.3 considers observed **learning effects** in the students' proof evaluations.

Section 8.4.4 recalls a number of themes which arose in the process of interpretating the written evaluations. These themes had been considered in the design of the interviews.

8.4.1 Findings from the interviews about first year students' views of a valuable proof

Students' criteria for valuing a proof

Section 7.3.3 was concerned with findings from the students' written evaluations about their criteria for accepting or valuing a proof. These findings include that most of the students recognize the importance of mathematical definitions, general applicability, internal correctness and convincing mathematical arguments. A relatively small proportion of the students considered precise argumentation, sufficiency without excess and justification of significant assertions in their written responses on proposed ‘proofs’. Comparable and some additional criteria for accepting a proof are identified in the interview evaluations.

In total, the students considered a relatively wide range of criteria for accepting a proof. However, in many instances only one or at most two of them are mentioned by any individual student. (This observation is discussed in Section 8.4.3.)

Listed below are the criteria which played a role in the students' evaluations in the interviews.

- **General applicability** is mentioned frequently in the interview transcripts and can be seen as a significant criterion for students to accept a proof.
- Some students consider as most relevant in their evaluations how **comprehensible** the proposed proof is and how it explains the content of the statement.
- Some students appreciate considerations and **explanations** about **why** the statement is true.
- 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 explained to the reader. This didactical value may also include how a slightly unusual approach or viewpoint might be valued – an alternative way of representing mathematical information may be appreciated.
- Some students see **justification of each stated assertion** as an important criterion for valuing a proof.
- Some students consider **formal appearance** as crucial in a mathematical proof.

The students' written evaluations suggested that some students consider the proving method or idea as less important in a proof than its appearance. This observation is confirmed by the interview evaluations on *Task I*. In these evaluations the proof idea or method did not seem to play a significant role. However, the evaluations on *Task II* suggest that the proof idea has slightly more importance for some of the participating students. The importance of the proving method or idea in the students' proof evaluations seems to depend on the comprehensibility of the particular proof. This issue is considered in Section 8.4.2.

Purposes of proofs in the view of first year students

The identified evaluation criteria indicate which purposes of proofs the participating students consider relevant in their proof evaluations.

- The students' high appreciation of general applicability in a proof to demonstrate the veracity of the proof as well as the statement indicate that the purposes of proof *verification* (concerned with the truth of a statement) and *explanation* (providing insight into why it is true) do play a significant role in the students' proof validations and evaluations.

  However, as in the case of the written evaluations, the interview findings show that to some students the most pertinent consideration is how the statement itself is verified and explained by the proposed proof. To those students the purpose of proof seems to be merely to support the truth of the statement, i.e. to convince the reader *that* the statement is true, in De Villiers' term *verification*. Other purposes, in particular *explanation* of *why* the statement is true, are considered but seem to be less important to this group of students.

- The students' appreciation of comprehensibility of a proof indicates that *explanation* and *communication* (the transmission of mathematical knowledge) are considered significant purposes of proofs by some of the students.

- De Villiers suggests *systematization* as one purpose of proof, i.e the organization of various results into a deductive system of axioms, major concepts and theorems. The interpretation of the written evaluations suggested that *systematization* was not considered by the students. This is different from the interpretations of interview evaluations, in which a few attempts of students to consider *systematization* could be identified.

- Another purpose of proof suggested by De Villiers is *discovery* of new results. This purpose also was not identified in the written evaluations of students. The oral evaluations showed an appreciation of some sort of didactic value in a proof, by some students. This appreciation may be understood by a consideration of the purpose *discovery*.

- Other purposes of proofs which are mostly concerned with enhancement of mathematical knowledge such as *intellectual challenge* (the *self-realization/ fulfillment* derived from constructing a proof) or *problem solving competencies* were not considered in the students' written evaluations. This is similar in the interview evaluations.

- As mentioned in Section 3.2, De Villiers does not claim his list to be exclusive, and a number of expansions have been suggested. As with the students' written evaluations, the oral evaluations do not indicate that the students consider other purposes of proof than those listed by De Villiers.

Several of the students' evaluations indicate that some students, whether they understand the proof idea or not, do not value it highly in their evaluations. These students seem to regard a proof as good if it requires minimal engagement of the

reader. Proofs that use a style of argument and presentation that is likely to be familiar to these students are more likely to be evaluated favourably. A proof that requires the reader to interpret a manner of representation or take a viewpoint different from the obvious or conventional one is less likely to be ranked favourably, even by students who understand it and believe it to be correct. There also seems to be a reluctance to engage with arguments that look unusual. This may be understandable and just due to (un)familiarity, experience and expectation.
These observations suggest that in some cases the purposes *explanation* and *communication* are considered more than *verification* or *intellectual challenge*.

8.4.2 Students' proof validation and evaluation habits

Do the students relate the three facets of artifacts in their proof evaluations?

As in the case of the written evaluations, the students' oral proof evaluations provide evidence of various attitudes towards the relationships E_{AI}, E_{IP} and E_{AP} among the actual and intended character and the purpose(s) of these proofs. The interviews also allow observations of how some of the students' evaluation habits progressed. The following attitudes could be identified in the interview transcripts.

- *Student A* seems to be content with her considerations on E_{IP}. Doing so she has mostly one purpose of proof in mind, which in her conception seems to be to explain what the statement is saying and maybe to verify it with an example or two. This is not on De Villiers' list and would not generally be considered a purpose of proof. Without any prompting she does not seem to consider E_{AI} and E_{AP}, that is how intention and purposes are implemented.
- It seems that if the students consider reflections on E_{IP} as relatively straightforward, they may omit further investigations of the implementation of the proof, that is on relations E_{AI} and E_{AP}.
  - In the case of Benny's collection of examples in *Task I*, the majority (six) of the students do not investigate E_{AI} after considering E_{IP}. None of them notes the error in Benny's answer.
  - Some students see the relations E_{AI} and E_{AP} as relatively unimportant and any errors as inconsequential in their proof evaluations if the proof meets both of the following conditions:
    1. They regard the idea as suitable, that is they regard the proof as sufficient in respect of E_{IP}.
    2. The proof is presented in the form of algebraic formalism.

  The evaluations of Darragh's and Joan's answers suggest that some of the students, once satisfied with what they acknowledge as the proof idea or intended proof, do not even consider its implementation. In terms of the schema, these students are not concerned about E_{AI} and E_{AP} if they are completely satisfied by E_{IP} and by the author's intention. Only if they struggle in identifying a proof idea (as in the case of Elaine's approach), they read a proposed proof carefully line by line and therefore consider the relations E_{AI} and E_{AP} (see hypothesis below).

Helena's proof, written purely in text, is considered sufficient in respect of E_{IP}. However, as this proof does not include any algebraic notation, the students do not approve this answer without any further considerations.

Similar behaviour to that described in the hypothesis may appear in mathematical practice as well. For example the mathematician Thurston (2006, p. 43) describes how, if he is satisfied that the authors are "putting in enough rigamarole" to carry the idea, he might not read all the details of a proof or check for inconsequential errors. However, to recognize whether an author puts in "enough rigamarole" requires plentiful and substantial experience with mathematical proof as well as possibly profound mathematical knowledge. Also, mathematician's validation habits depend on the circumstances. Thurston describes how he reads a mathematical paper in a field in which he is "conversant". In other fields and in the context of evaluations, he might not omit a careful investigation of the implementation, that is of considerations of E_{AI} and E_{AP}. Alcock and Weber (2004) consider how mathematicians determine whether an argument is a valid proof. The participants in their study (eight mathematicians) were aware that the study was concerned with their validation habits. It cannot be guaranteed that they showed the same habits that they would employ when reading a mathematical paper without being observed. Also, even though the article does not describe each participants mathematical expertise, it is likely that their specialisms are in various mathematical fields. Thus, these mathematicians validate the proposed proofs under different circumstances than those decribed by Thurston. These circumstances are more likely to be comparable to those of the students in this study. Alcock and Weber describe that their participants first examined the proof methods employed. If those were judged to be reasonable the participants would proceed to the second stage of line-by-line verifications. In terms of the schema, these mathematicians considered E_{AI} and E_{AP} if satisfied by E_{IP}.

The high appreciation of Joan's (wrong) approach by half of the participating students can be explained by their positive judgement of Joan's initial proof idea resulting in the omission of considerations regarding E_{AI} and E_{AP} in their evaluations of Joan's answer. This leads to some questions:

* Do the students tend to overestimate their knowledge about mathematical proofs?

* Do they inappropriately feel 'conversant' in some fields of mathemat-

ics?

As we consider the newcomers' learning and behaviour in a community of mathematical practice, this issue seems interesting and worthy of further investigation.

These observations indicate that to most of the students the relation between intended proof and purposes of proofs E_{IP} is often more significant than the actual implementation, that is the relations E_{AI} and E_{AP}.

- In case that they are somehow sceptical about the proof idea, some of the students do consider E_{AI} as well as E_{IP}.

  This is suggested by the differences of the students' evaluations of Darragh's and Joan's answers. In comparison to Darragh's error, Joan's is more serious. Both of them reach the correct conclusion in terms of the statement, but Joan's $\pm\sqrt{\frac{c}{a}}$ are not the correct solutions of the quadratic equation that she (correctly) starts with. More students notice Joan's error than Darragh's – this may be because the simple nature of Joan's final quadratic and the surprisingly simple form of the roots raises a suspicion that something could be wrong. This suggests that some students if sceptical in respect of E_{IP}, they do look at E_{AI}, because something about the final lines looks a bit strange.

- If the students struggle in identifying a proving method or idea, they tend to avoid further investigations into this and therefore do not consider relations E_{IP} and E_{AP}. They are rather content with considerations on E_{AI} in such cases.

  These habits are observed in the students' evaluations of Elaine's (wrong) 'proof'. Most of the students do not question Elaine's method at all, only one mentions some scepticism but that does not prompt her to investigate E_{AP} to find some errors. The other students seem content with relating the actual character of Elaine's answer with her assumed intention and just evaluate on the basis of E_{AI}.

- If the students fail (or do not even attempt) to understand the author's intention, they might not consider E_{AI} and E_{AP}. At most they may hint on E_{IP} in respect to the purpose *communication* and criticize that the proof is difficult to comprehend.

8.4.3 Learning processes indicated by the students' proof evaluations

Do the nature or purposes of mathematical proof become *visible* to the student evaluator?

Section 2.3 was concerned with the *condition of transparency* of artifacts. Finding the right balance in the condition of transparency is regarded as crucial in teaching and learning processes. Consequently, transparency of the artifact proof was considered as important for the learning of mathematics. The *condition of transparency* of proofs in mathematical practice is the balance between *visibility* and *invisibility* of proofs. Devising or studying a proof might involve a focus on objects and properties particular to the mathematical topic. The nature of proof may not be in the forefront of the student's attention, i.e. may be *invisible*. At the same time, proving and its techniques can require explicit attention and therefore proof needs to be *visible*. Learners need to understand the significance of mathematical proof. One question of this thesis is whether the nature of proof and its purposes become *visible* to students through the practice of proof validation and evaluation.

The transcripts do not provide a lot of explicit evidence that the nature of proof and its purposes became *visible* to the students. However, some excerpts suggest that students engaged in reflections about the nature of proof. Frequent pauses during the evaluation processes may indicate that the students stopped to think about what they value or disapprove about a particular proof and how they can substantiate their evaluation. Occasionally the interviewer's prompting encouraged reflections about the nature of proof. For example *Student G*'s reaction to Ciara's answer (p. 104) indicate that learning processes are initiated by the task. Her first reaction indicates that she admires the unusual approach. After careful prompting by the interviewer a reflection process is initiated and the student becomes more and more unsure, until at some point she almost decides that this is not a proof.

In particular comparison of different answers seems to encourage reflections of these kind to make some features and purposes *visible* to the students. Some of the students seem to be aware of this effect, at least when they have arrived at *Task II*. For example *Student F* says explicitly that he would accept Gerard's answer *"now,"* but would *"probably end up changing* [his] *mind after seeing the rest."* Examples are the changes in some of the students' opinions about Anna's proof. These students identified an attempt at an explanatory argument in Anna's answer only after seeing Benny's, having previously dismissed Anna's as *"just examples"*. This is an example at least of the purpose of explaining **why** the statement is true becoming *visible*.

Some differences in the students' performances in the two tasks may also indicate learning effects for some of the students caused by their reflections on their evaluations. These differences may be attributable to aspects of the two tasks, but they may indicate the development of a process of reflective engagement.

- Some of the students' proof evaluations suggest that the proof idea might have played a more significant role in these students' evaluations of answers on *Task II* than in those on *Task I*, see considerations on p. 155.

- The students' rankings of *Task II* do not differ as much from the assumed ranking of experienced evaluators as in the ranking of *Task I*.

Can proof evaluation practice in groups encourage learning about the nature of proof?

A significant range of features and purposes of proofs have been recognized in the participants' proof evaluations. However, often only one or at most two of them are mentioned by any individual student. This might indicate that knowledge about the nature of mathematical proofs is only *visible* in parts to individual students and could be improved through evaluation practice in groups.

8.4.4 Findings of the interviews in relation to issues arising from the students' written evaluations

Some issues arising from written exercises were considered in the interview design. Whether the interviews provided deeper insights into these matters is considered below.

The students' views about the role of examples in a proof

The students' responses to the written evaluation task suggested that in their views examples play an important role in the context of proofs. Without any examples students did not seem to be satisfied by an argumentation. In Section 7.3.2 two possible reasons to value examples highly in a proof were considered: one is to reach an understanding of the statement itself and sometimes also of the argument. Another reason might be to provide evidence in support of the truth of the statement. The use of examples for both of these purposes is common in the mathematical research literature, though there the examples would not be incorporated into proofs.

Some of these views can also be found in the interview transcripts. In particular *Student A* insists on inclusion of examples in a proof. In her evaluations it appears most relevant how the contents of both, the statement and its justification, is explained by the proposed answer. She regards examples in a proof as crucial to provide comprehensibility. *Student A*'s positive evaluation of Benny's answer on *Task I* suggests that for her, examples can provide adequate justification.

The remaining students did not seem to value examples in a proof as highly as suggested by the written and *Student A*'s oral evaluations. However, examples are occasionally sought by these students as well. Different reasons for this can be found.

- *Student B* demands inclusion of examples in Ciara's, Darragh's and Fintan's answers. She distinguishes between *proof* and *answer* and sees examples as important in a valuable answer:

  > *Student B*: *"I'm missing generalizations, formula, example and explanation."*
  > *I*: An example is part of a good proof?
  > *Student B*: *"To answer the question."* (...)

- *Student C* and *Student D* miss examples in Helena's answer. They see themselves in the role of a teacher and in their views examples would show them that Helena understands her own argument. These students may want to ascertain that Helena did not just memorize the argument without understanding it.

  Student D: *"She'd have to give an example."*
  (...)
  Student C: *"Even that (...) to show, that she understands it."*

- *Student G* and *Student H* also miss examples in Helena's answer. In their views examples would make Helena's proof written in text more comprehensible:

  Student H: *"That, in a way, that is a proof. But I would have liked to see an example. Although it proves it, it's written (...) you have to kind of grasp a bit more (...) You have to think a lot harder."*

  Student F requires examples in Ian's answer, also to make this answer more understandable for him:

  Student F: *"(...) give an example for each or something to show they are actually true. Because he's expecting you to believe him that this is the only ways it can be done (...) With examples he'd show better."*

The written evaluations suggested that students consider examples as essential in a mathematical proof. This impression is not confirmed by the interview evaluations. Only one out of the eight students shares this view. Some of the other students also request examples not as an essential part of the proof but rather as an addition, mostly to enhance the reader's understanding of the argument.

The students' views about the role of mathematical formalism in a proof

The students' formulaic and inflexible picture of valuable proofs was noted in the interpretation of the written evaluations in Section 7.3.2. Structure and 'mathematical-looking' formalism seemed more important to some students than the ideas comprising the argument. A visual approach to prove the statement was generally not accepted by the students and an argument formulated as text was "not mathematical enough" to a third of them. In general, a good idea to prove a statement was not valued as highly as the structure and formalism of a proof. One explanation suggested was that some students associate mathematical proof with algebraic formulas. This theme was regarded worthy of further attention.

These findings are comparable to those of the interpretations of the interview evaluations. The students' attitudes towards visual arguments, arguments written entirely in standard text, and incorrect arguments (possibly involving algebraic formalism), were similar to those attitudes of students in their written evaluations. In both tasks, the visual approaches were ranked least favourably. On the other hand, both incorrect arguments, which include algebraic equations and mathematical notation, scored relatively high in the students' rankings. It seemed that, mostly because the incorrect arguments appeared 'mathematical' to some of students, they presupposed that the proof intention was sufficient. These students were content with considerations on each step of the arguments, that is with considerations on E_{AI}. However, it can be noted that these attitudes were shown by seven out of the eight students in their evaluations o answers on *Task I*, but significantly fewer (four students) in their evaluations of answers on *Task II*. This may be because Elaine's wrong answer on *Task I* is more seriously erroneous than Joan's wrong answer on *Task II*. Another possible explanation is that, considering the fact that Joan's answer is one of the last proposed proofs, the previous exercises have begun to initiate a learning process, prompting some students to investigate the method or the intention of the proposed proof.

Some students presuppose in their oral evaluations that a proof meets its purposes if it includes algebraic equations and mathematical notation. Algebraic appearance of a proof has a significant impact in their evaluation habits. This reinforces the observation, as suggested in the interpretations of the written evaluations, that some students may associate mathematical proof with algebraic formulas.

The students' views about the role of explanation in a proof

Section 7.3.2 describes how some of the students' responses to the written evaluation task indicated that just having a good idea to prove a statement is not sufficient for most of the students. They seemed to regard the convincing nature of the argument and the quality of explanation of ideas as important criteria to value a proof.

Some students valued comprehensibility highly in their oral evaluations. These students may require more explanations (beside examples) in the proposed proofs. Explanations may be required to enhance the reader's understanding of the statement or of its proof. This is similar to the findings of the interpretations of the written evaluations. Inclusion or the quality of explanations seemed to play a more significant role in the students' written evaluations than in those performed in the interviews. An explanation for this may be the fact that the interviewer intervened with prompting if the students struggled to understand the arguments. This was not the case in the written evaluation task.

Chapter 9

Summary of the Practical Part

The Practical Part of this thesis included descriptions and outcomes of the written experiment *End-of-the-Year-Test08* and the interviews conducted in March 2009.

In Chapter 6 practical aspects and the evolution of the research setting were described.

Chapter 7 was concerned with the written evaluation task. This chapter included a description of this task and reflections on how an experienced evaluator might consider the proposed proofs. The methodology used to manage the acquired data was described. An overview of the acquired data was given, as well as an interpretation of the students' responses to the evaluation task in relation to their proof evaluation criteria. These findings were related to the proposed schema to interpret proof evaluations.

Chapter 8 was concerned with the interviews held in March 2009. This chapter included considerations on the interview design as well as a description of the tasks that were used and of the interview transcripts. It also included interpretations of the transcripts in light of the schema suggested in Chapter 4. These findings were related to those of the written evaluations.

The observations of both experiments in the light of the schema suggested some interesting findings regarding the students' criteria for valuing a proof, their views on purpose(s) of proof, their validation habits, and learning effects initiated and/or supported by practice of proof evaluation.

Part III

Concluding discussion

Chapter 10

Some closing remarks

This final part discusses some advantages and shortcomings of the schema that has been developed and used for the interpretation of evaluations of mathematical proofs. It further suggests inclusion of proof evaluation activities in small groups into curricular learning activities for novice students. Finally potential for further research is discussed, arising from the theoretical framework developed in Part I of this thesis and the empirical study described in Part II.

Advantages and shortcomings of the schema to interpret proof evaluations

As discussed in Section 2.2.2, artifacts can be evaluated only on the basis of fitness-for-purpose. Different users may have different purposes in mind for a particular artifact, or may disagree on the merits of a particular artifact for the same purpose. They may evaluate the same artifact differently, according to their individual conceptions of its purpose and according to their individual experiences of the usefulness of its features. The semi–structured interviews conducted in 2008 as part of the study yielded a substantial quantity of data. Data from each interview included the student's oral evaluation of each of the eleven proposed proofs from *Task I* and *Task II*, but otherwise the overall data set was quite unstructured. The transcripts and audio recordings of the interviews might have prompted a reader or listener to make a number of interesting observations about individual students' responses to particular tasks, but an analytic tool was needed for a more focussed analysis of this complex data system. One of the benefits of the schema was that it provided terminology, language and context for the consideration of students' evaluations of the proposed proofs. Most students' comments on the proposed proofs could be interpreted as relating to at least one of the relations E_{AI}, E_{AP} and E_{IP}. In some cases it was not obvious which of these relations was evoked by a particular comment; for example it was not always clear whether a student's (or experienced evaluator's) response indicated attention to E_{AP} or E_{AI} or both. This occasional ambivalence was not problematic and was not considered to be a weakness of the schema. The schema was very valuable in the interpretation of the interview data, when considered as a flexible and general framework in which to consider patterns of behaviour in evaluation of mathematical proofs. The three facets of proofs that it involves are not mutually exclusive and are closely interrelated. The schema is not a tool for classifying evaluations or for classifying the comments of evaluators; rather it is a mental model for considering what we do when we evaluate a proof.

Some student evaluations seemed difficult to interpret using the schema, at least initially. Relations between the facets could not easily be identified in these cases. However, considerations of these comments in the light of the schema revealed valuable information about what purposes and features of proofs are considered important by evaluators. As documented in Chapter 8, some of the student participants in this study did not always seem to give due attention to what might be considered conventional purposes of proof. Some evaluations seemed to indicate a focus on features of proofs that were not obviously connected to these purposes, such as the presence or absence of examples or of algebraic formalism. Evaluations of this nature can be interpreted in the light of the schema, if the notion of 'purpose' in E_{AP} and E_{IP} is understood to entail any purposes that the evaluator may have in mind. In cases where such individual purposes apparently clash with more conventional ones, it may be suggested that the schema highlights incomplete or inadequate conceptions of the nature and purposes of proof in a usefully precise way. Whether explicit attention to the relations described in the schema could be valuable in helping students to both expand and focus their understanding of the nature of mathematical proof may be a question worth investigating.

Proof evaluation as learning activity for students?

Section 2.1 explained that in order for novice students to take advantage of the collective context at university to support their learning, it is important that a sense of belonging to the new learning community is developed. This applies for both the community of students as well as the community of mathematical practice. Considering this I suggest practice of proof evaluation activities in small groups, in particular for novice students.

Data from the written exercise and interviews that comprised the practical part of this study indicate an encouragingly wide range of features, purposes and criteria that are considered by students in their validations and evaluations of mathematical proofs. This may suggest that in a collective sense, the knowledge of the first year class about the nature of mathematical proof is quite extensive. However, specific items from the wide range of criteria and purposes mentioned by students generally appeared in isolation; it was unusual for more than one or two of them to appear in the comments of any individual student on a proposed proof. Individual students potentially have much to learn from and teach to each other about the role and purposes of proof in mathematical practice. In particular I suggest that interaction and discussion can contribute in an essential way to the development of the skills involved in proof validation, proof evaluation, and ultimately proof construction.

Proof evaluation is an essential activity in mathematical practice that can be performed from elementary to advanced levels of mathematical learning and in all subdisciplines, and seems to naturally provoke discussion. Thus the practice and discussion of proof evaluation in groups may be a particularly valuable activity both to develop a student community of mathematical practice and to enhance individual students' knowledge about proof.

These suggestions can be interpreted as consistent with the socio-cultural perspective, which views the learning process as the product of a cultural and social meaning shared in a learning community (Lave & Wenger 1991).

Ideas for further research

Some of the data acquired in the course of this study (in particular the *Diagnostic-Test08* and a preliminary survey on mathematicians' views of visual reasonings) has not yet been analysed in detail, due to time limitations and other factors. It is envisaged that the analysis of these data will be completed in the light of the schema. In particular, findings on the participating mathematicians' validation and evaluation practices may be associated with existing research literature and with the findings in this study about students' views of visual reasonings. This analysis may suggest future research directions involving comparison of the evaluation habits of novice students and experienced practitioners, and exploration of the opportunities afforded by the practice of proof evaluation for access to mathematical practice.

A longitudinal project involving sustained practice of proof evaluation, by students progressing through the transition from second level to university and onwards to more advanced levels of mathematical learning, would be a natural next step in this line of research. Such a project would incorporate theoretical, practical and pedagogical dimensions and would involve extensive collaboration between researchers and lecturers, for example on the development of materials for use in the curriculum. This would yield further insights into many of the findings documented in this thesis, in particular those concerned with learning effects stimulated by the practice of proof evaluation and comparison. As discussed in Section 8.4.3, evidence was found in this study that students engaged in reflections about the nature of proof in their proof evaluations. In particular, comparison of different answers seemed to encourage reflections of this kind, and made some features and purposes of proofs *visible* to the students. Some differences in performances in the two tasks also suggested possible learning effects for some of the students, caused by their reflections on their evaluations in the course of the interviews. A longitudinal project involving sustained practice of proof evaluations at advancing levels of mathematical complexity, and regular assessments and interviews, would enhance existing understanding of these learning effects. The interview data from this study do not give any specific information on what the students thought about their interactions with the researcher or about the tasks that were presented. *Student G*'s comment at the end of her interview, saying that this meeting has changed her mind about the need to do more than Leaving Certificate mathematics, indicates her appreciation of this form of exercise. This comment may not indicate a specific learning event arising from the interview experience. Nevertheless it shows that the meeting at least initiated for this student an awareness of the limitations of her mathematical experience to date, and perhaps helped to expand her mathematical view. Being the interviewer, I recall that all of the participating students enjoyed these sessions. My impression was that they felt they were taken seriously and saw themselves in the role of critics or teachers whose opinions were valued. It was striking that the participants engaged so fully in this activity, even though the sessions were quite long and possibly tiring. Subjective impressions suggest that such high levels of engagement and enthusiasm were not always evident from the students' participation in the workshop sessions and in the lecture courses. This may be explained by the hypothesis that the practice of proof evaluation provides an opportunity for newcomers (novice students) to enhance their access to mathematical practice and continue to develop their identities from 'student' to 'mathematician' or maybe 'teacher'. A longitudinal research project, engaging students in focus groups and involving collaboration with lecturers, would provide clear evidence for this personal impression. It would further provide insight into the nature of the learning effects prompted by the practice of proof evaluation and comparison, and their potential to persist through advancing stages of mathematical study.

Bibliography

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

Alcock, L. & Weber, K. (2004), 'How do mathematicians validate proofs?', Paper presented at the annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Toronto, Canada.

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

Alibert, D. & Thomas, M. (1991), Research on mathematical proof, *in* D. Tall, ed., 'Advanced Mathematical Thinking', Kluwer, Dordrecht.

Aristotle (n.d.), Physica, *in* S. D. Ross, ed., 'The Works of Aristotle', Vol. II, Oxford: Clarendon Press.

Avigad, J. (2006), 'Mathematical method and proof', *Synthese* **153**(1), 105–159.

Bass, H. (2009), 'How do you know that you know? making believe in mathematics', http://deepblue.lib.umich.edu/handle/2027.42/64280/Bass-2009.pdf.

Bell, A. (1976), 'A study of pupil's proof-explanations in mathematical studies', *Educational Studies in Mathematics* **7**, 23–40.

Bereiter, C. (2002), *Education and Mind in the Knowledge Age*, Lawrence Erlbaum Ass.

Bills, E. & Tall, D. (1998), Operable definitions in advanced mathematics: The case of least upper bound, *in* 'Proceedings of PME', Vol. 22, Stellenbosch, South Africa, pp. 104–111.

Bonica, L. & Sappa, V. (2008), Cooperative learning and support in the transition from school to university: an integrated research-intervention aimed at first-year psychology students, *in* "Cooperative Learning in Multicultural Societies: Critical Reflections", IAIE–IASCE Conference, Turin, Italy.

Davis, P. & Hersh, R. (1981), *The Mathematical Experience*, Viking Penguin Inc., New York.

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

Dreyfus, T. (1990), Advanced mathematical thinking, *in* P. Nesher & J. Kilpatrick, eds, 'Mathematics and Cognition: A Research Synthesis by the International Group for the Psychology of Mathematics Education', ICMI Study Series, Cambridge University Press, pp. 113–134.

Dreyfus, T. (1999), 'Why Johnny can't prove', *Educational Studies in Mathematics* **38**, 85–109.

Fischbein, E. (1982), 'Intuition and proof', *For the Learning of Mathematics* **3**(2), 9–18.

Fischbein, E. & Kedem, I. (1982), Proof and certitude in the development of mathematical thinking, *in* A. Vernandel, ed., 'Proceedings of the Sixth International Conference of the Psychology of Mathematics Education', Universitaire Instelling Antwerpen, Belgium, pp. 128–131.

Glaser, B. & Strauss, A. (1967), *The discovery of Grounded Theory: Strategies for quaitative research*, De Gruyter, New York.

Hanna, G. (1991), Mathematical proof, *in* D. Tall, ed., 'Advanced Mathematical Thinking', Kluwer, Dordrecht.

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

Hanna, G. & Barbeau, E. (2008), 'Proofs as bearers of mathematical knowledge', *ZDM Mathematics Education* **40**, 345–353.

Harel, G. & Sowder, L. (1998), Students' proof schemes: Results from explatory studies, *in* J. K. A.H. Schoenfeld & E. Dubinsky, eds, 'Research in Collegiate Mathematics Education III', Providence, R.I.: American Mathematical Society.

Hart, E. (1994), A conceptual analysis of the proof-writing performance of expert and novice students in elementary group theory, *in* J. Kaput & E. Dubinsky, eds, 'Research Issues in Undergraduate Mathematics Learning: Preleminary Analysis and Results', MAA Notes 33, Washington, D.C.

Hemmi, K. (2006), Approaching Proof in a Community of Mathematical Practice, Doktoral thesis, Stockholm University.

Hemmi, K. (2008), 'Students' encounter with proof: the condition of transparency', *ZDM Mathematics Education* .

Henkel, M. (2000), *Academic identities and policy change in higher education*, Jessica Kingsley Publishers, London.

Hersh, R. (1993), 'Proving is convincing and explaining', *Educational Studies in Mathematics* .

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

Jawitz, J. (2009), 'Academic identities and communities of practice in a professional discipline', *Teaching in Higher Education* **14**(3), 241–251.

Küchemann, D. & Hoyles, C. (1999–2003), The longitudinal proof project. a longitudinal study of mathematical reasoning: Student development and school influences, Technical report, Institute of Education, University of London.

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

Lindlof, T. & Taylor, B. (2002), *Qualitative Communication Research*, second edition edn, Sage.

Mariotti, M. (2006), Proof and proving in mathematics education, *in* A. Gutierrez & P. Boero, eds, 'Handbook of Research on the Psychology of Mathematics Education: Past, present and future', Sense Publishers, Rotterdam, Netherlands, pp. 173–204.

Martin, G. & Harel, G. (1989), 'Proof frames of preservice elementary teachers', *Journal for Research in Mathematics Education* **20**, 41–51.

Mason, J., Burton, L. & Stacey, K. (1982), *Thinking Mathematically*, Addison Wesley, London.

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

NCCA (2005), Leaving certificate examination 2005: Mathematics, *in* 'Chief Examiner's Report', Department of Education and Science, Dublin Ireland.

Powers, A., Craviotto, C. & Grassl, R. (2010), 'Impact of proof validation on proof writing in abstract algebra', *International Journal of Mathematical Education in Science and Technology* **41**(4), 501–514.

Rav, Y. (1999), 'Why do we proof theorems?', *Philosophia Mathematica* **7**(1), 5–41.

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

Schoenfeld, A. (1985), *Mathematical Problem Solving*, Orlando, Academic Press.

Schutt, R. K. (2009), *Investigating the Social World. The process and practice of research*, 6th edn, Pine Forge Press.

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

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

Senk, S. L. (1985), 'How well do students write geometry proofs?', *Mathematics Teacher* **78**(6), 448–456.

Solomon, Y. (2006), 'Deficit or difference? The role of students' epistemologies of mathematics in their interactions with proof', *Educational Studies in Mathematics* **61**(373–393).

Steiner, M. (1978), ‘Mathematical explanation’, *Philosophical Studies* **34**, 135–151.

Tall, D. (1989), ‘The nature of mathematical proof’, *Mathematics Teaching* **128**, 28–32.

Thurston, W. (1994), ‘Letter to the editors’, *Scientific American 270* **1**(5).

Thurston, W. (2006), On proof and progress in mathematics, *in* R. Hersh, ed., ‘18 Unconventional Essays on the Nature of Mathematics’, Springer, pp. 37–55.

Traweek, S. (1988), *Beamtimes and lifetimes: The world of high energy physicists*, Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (1978), *Mind in Society*, Harvard University Press.

Weber, K. (2001), ‘Student difficulty in constructing proofs: The need for strategic knowledge’, *Educational Studies in Mathematics* **48**, 101–119.

Weber, K. (2002), ‘Beyond proving and explaining: Proofs that justify the use of definition and axiomatic structures and proofs that illustrate technique’, *For the Learning of Mathematics* **22**(3), 14–17.

Wenger, E. (1991), Toward a theory of cultural transparency: elements of a social discourse of the visible and the invisible, Dissertation, University of Palo Alto.

Wenger, E. (1998), *Communities of Practice*, Cambridge University Press, Cambridge.

Appendices

Appendix A

‘End-of-the-Year-test’ 2008: The Task

Question 5. Aoife, Barry, Cathy, Dan, Eve and Finn are students trying to prove whether the following statement is true or false.

When you add any two even numbers, your answer is always even.

The responses given by each of the six are below. For each response, give a mark out of five, and give a line of advice for the student.

1. *Aoife's answer*: a is any whole number.
   b is any whole number.
   $2a$ and $2b$ are any two even numbers.
   $2a + 2b = 2(a + b)$.
   So the statement is true.

   Mark out of five: ______

   ____

   ____

2. *Barry's answer*:

   $$\begin{array}{ll} 2+2=4 & 4+2=6 \\ 2+4=6 & 4+4=8 \\ 2+6=8 & 4+6=10 \end{array}$$

   So it's true.

   Mark out of five: ______

   ____

   ____

3. *Cathy's answer*:
   Even numbers are numbers that can be divided by 2. When you add two numbers with a common factor, 2 in this case, the answer will have the same common factor. So the statement is true.

   Mark out of five: ______

   ____

   ____

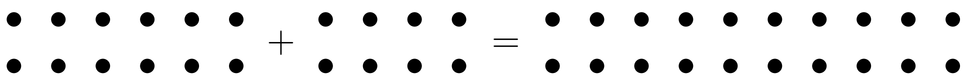
4. *Dan's answer*:
Even numbers end in $0, 2, 4, 6$ or 8. When you add any two of these the answer will still end in $0, 2, 4, 6$ or 8. So it's true.

Mark out of five: ______

5. *Eve's answer*:
Let $x =$ any whole number, $y =$ any whole number.
$x + y = z$
$z - x = y$
$z - y = x$
$z + z - (x + y) = x + y = 2z$.
So the statement is true.

Mark out of five: ______

6. *Finn's answer*:



So the statement is true.

Mark out of five: ______

Appendix B

‘End-of-the-Year-test’ 2008: Coding Scheme

Code	Description of response
MS VerwithEx	Misunderstanding of the statement The student verifies the proposed proof with examples.
NoUnd	The student does not seem to understand the proposed proof.
PC NoPr PrNotEnough	The student regards the proposed proof as *"correct"*. The student does not regard the proposed argument as a *"correct"* proof of the statement. The student regards the approach as not enough to be a proof.
VG NotGood	The student expresses in some way that she likes the answer. (The comment does not indicate whether the proof is regarded correct and sufficient.) The student does not like the answer. (The comment does not indicate whether it is regarded a correct proof.)
InclCrit NotInclCrit	The student's comment includes some reasons for approval or disapproval of the proposed proof, indicating evaluation criteria. The student does not give any reasons for approval or disapproval of the proposed proof indicating evaluation criteria.
A-C:S	The student interprets Aoife's and Cathy's answer as similar.
Cl NotCl NotClEnough	The student appreciates the 'clearness' of the proposed answer. The student criticises, that the proposed proof is not 'clear'. The student would prefer a clearer appearance or structure.
CritLackofForm MForm	The student criticises a lack of mathematical formalism (e.g. 'equations' or 'algebraic formula') in the proposed proof. The student would prefer to see more mathematical formalism.
NoEx MEx PrAsEx PrbyEx	The student criticises that the suggested answer does not include an example. The student would prefer to see more examples. The student regards a general argument as an example. The student regards a selection of examples as "proof by example".

Code	Description of response
NoExpl MExpl WellExpl NoExplWhySt	The student criticises a lack of explanation (of the **proof**). The student would prefer to see more explanations (of the **proof**). The student appreciates the quality of explanation in the proposed answer. The student criticises that the proposed proof does not explain why the **statement** is true.
NotGen Gen	The student criticises a lack of generalisation. The student mentions explicitely the general applicity of the answer.
Compl Conc	The student finds the proposed answer more complicated than necessary. The student appreciates that the proposed answer is concise.
NoDef PoorDef	The student criticises a lack of definitions. The student criticises a mistake in the proposed definition of an even number.
AffNotJust	The student criticises that in the proposed argument affirmations are being stated without justification.
AlgWr	The student notices wrong algebraical manipulations.
Conv	The student appreciates in the proposed proof that it carries some potential to convince someone of the truth of the statement.
BC	The student regards the proposed proof as 'basically correct'. (The student does not recognize that the proposed argument is wrong.)
Int-Ex Int-Pr Int-?	The student interprets Finn's visual approach as one example. The comment indicates that the student interprets Finn's approach as visual demonstration of a general proof. The comment does not indicate how the student interprets Finn's visual approach.
IncorrectCrit	The student criticises the proposed proof by using mathematically incorrect arguments.
NoComment	The student does not comment on the proposed proof.
Misc	The comment differs from the coded attitudes listed above.

Appendix C

‘End-of-the-Year-test’ 2008: Students’ responses

EoYTest07/08: Student evaluations of 'Aoife's answer'

Marks:	Comments on Aoife's answer:	Code	
5	"Proof is correct."	PC	NotInclCrit
5	"Define an even number before them in the proof."	NoDef	InclCrit
5	"Aoife has a very clear and straightforward answer."	Cl	InclCrit
5	"Proof is correct. Should also point out that a number is even if divisible by 2 more clearly."	PC, NoDef	InclCrit
4	"didn't explain why 2a and 2b are even numbers or why the last line was true. "	NoExpl	InclCrit
2	"Did not state a,b must be positive. Did not explain line 3 that the nth even number is 2n. Did not state why the statement is true eg 2(a+b) is divisible by 2 ??even number."	NoUnd, NoExpl, IncorrectCrit	InclCrit
5		No Comment	NotInclCrit
4	"Needs to state what an even no. is, i.e is divisible by 2, can be written in form 2 (something ?)"	NoDef	InclCrit
5	"well done"	VG	NotInclCrit
5	"This is a perfect example."	PrAsEx	NotInclCrit
3	"I believe Aoife's answer would have been more acceptable had she defined an even number"	NoDef	InclCrit
5	"Very good but would like a practical as well as a conceptual answer"	VG, NoEx	InclCrit
4	"State definition of even no at start."	NoDef	InclCrit
5	"very clear"	Cl	InclCrit
5		No Comment	NotInclCrit
2	"This doesn't prove for any numbers a and b. Therefore not proving the statement."	NoPr	InclCrit
0		No Comment	NotInclCrit
3	"Aoife should have stated that the whole numbers she used were even numbers, instead of 'any whole number' as the statement she is trying to prove is only concerned with even numbers."	NoUnd	InclCrit
5	"Keep up the good work, Aoife!"	VG	NotInclCrit
4	"Could have outlined it more clearly but basically sound."	NotClEnough	InclCrit
0		No Comment	NotInclCrit
5	"Well explained answer"	VG	InclCrit
4	"It shows that when you add two even numbers together you get an even number back."	WellExpl	InclCrit
4	"If a=-b then 2(a+b)=0 [not even]. She should either make her answers strictly for positive numbers or say a+b#0."	IncorrectCrit	InclCrit
4	"Give an example to back up proof."	NoEx	InclCrit
0	"Proves nothing"	NoPr	InclCrit
3	"Aoife is using clear and simple language to get her answer across but instead of writing whole number she should	Cl, CritLackofForm, Mexpl	InclCrit
	write a,b in N. And also justify with one extra line why she thinks that 2a+2b = 2(a+b) is a right statement to use."		InclCrit
5	"This is valid and in proper form."	PC, Cl	InclCrit
4	"She should have described what an even number is."	NoDef	InclCrit
4	"Good idea, but asked to add not multiply"	VG, NoUnd	InclCrit
5	"I think this explains it well but it would have been better if some number egs were used as well"	WellExpl, NoEx	InclCrit
5		No Comment	NotInclCrit
4	"No example"	NoEx	InclCrit
5		No Comment	NotInclCrit
4	"Aoife has proved the statement is true, I would just advice her to give an example to finish the proof."	PC, NoEx	InclCrit
3	"What happens when a = -b"	IncorrectCrit	InclCrit
5	"Good way of proving this statement and it will work for all possible numbers."	PC, VG, Gen	InclCrit

EoYTest07/08: Student evaluations of 'Barry's answer'

Marks:	Comments on 'Barry's answer':	Code:	
3	"Not a general solution."	NotGen	InclCrit
1	"This isn't general. It only deals with a small, finite number of integers."	NotGen	InclCrit
2	"Mark only gives 6 reasons even though there could be an infinite amount that disprove his statement. He needs to give a general Statement that answers the statement."	NotGen	InclCrit
2	"Good example, but that does not prove it for every number."	NotGen	InclCrit
2	"It is not proved because he did not try all the combinations of even numbers."	NotGen	InclCrit
1	"Not proved in general using something like induction. Contradiction could still be found."	NotGen	InclCrit
0	"You need to prove it not just give example. You need to use two variables like n+m that can be represent any number."	NotGen	InclCrit
1	"Although the examples prove it works for these specific values of evens, it doesn't prove for all of them. Needs to show for all evens."	NotGen	InclCrit
0	"This only proves it is true for some even numbers not all even numbers"	NotGen	InclCrit
2	"Barry does not generalise the formula, he gives examples but every member is not covered thus he doesn't prove it."	NotGen	InclCrit
2	"Barry's answer is not answering the question he has simply 'shown' a friend he has yet to prove anything "	NoPr, Conv	InclCrit
1	"There is no proof, if you worked to prove this statement this way you have to do it for all even integers"	NotGen	InclCrit
1	"Yes it's true but this method of proof only proves for specific values. Need to find a proof for all values."	NotGen	InclCrit
2	"He has just given examples, this is not a proof."	NotGen	InclCrit
2	"Need to show a general term."	NotGen	InclCrit
1	"This is only using a handful of examples. You need to prove for all even numbers."	NotGen	InclCrit
4.5	"Give a general equation which satiesfies other examples not shown."	NotGen	InclCrit
1	"Barry started off well, but he can't prove a statement with examples, he must use logic to find an arbitary (or general) case, in which the statement is true."	NotGen	InclCrit
2	"I gave him a few marks just because he's my friend but his enemy will never believe him because he only chose 6 examples out of an infinite amount."	NotGen, Conv	InclCrit
1	"Only shows it for a few selected values."	NotGen	InclCrit
0		NoComment	NotInclCrit
3	"Only done by examples, try to show proof by a general formula"	NotGen, (PrbyEx)	InclCrit
2	"It shows that by adding 2 even numbers together in certain cases you get an even number back but not good for proving all cases."	NotGen	InclCrit
3	"It's not bad. But it's only proved by example."	PrbyEx	NotInclCrit
2	"Didn't prove answer just showed it was true for some examples."	NotGen	InclCrit
2	"Just examples, no proof!!"	NotGen	InclCrit
2	"Barry should try & come up with a formula to prove this statement. Giving examples is not a sufficient answer. Although it is easy to see that he has grasped the concept of the question."	NotGen	InclCrit
2	"All this prove shows in adding numbers and it's not proving anything. There need to be a written explanation to accompany it."	NoExpl	InclCrit
2	"He hasn't proved that it's true for all even numbers."	NotGen	InclCrit
3	"Give general formula. Give more information on what your doing"	NotGen, Mexpl	InclCrit
3	"Although this does prove the statement, it only does so for a few egs and therefore the rest is left to assumption rather than proven fact"	NotGen, (PrbyEx)	InclCrit
2.5	"Instead of using specific examples use a more general formula to prove the statement, so it is true for all possibilities"	NotGen	InclCrit
2	"Not a proof."	NoPr	NotInclCrit
2	"It is true for those given examples, your argument is not convincing however as there is no general solution which shows me the statement will always be true."	NotGen	InclCrit
1	"All Barry has done here is give examples of the statement, he must however prove if the statement is true for all even numbers."	NotGen	InclCrit
2	"This doesn't prove it for every even integer. There could be even integers beyond 6 where this doesn't apply. "	NotGen	InclCrit
2	"What happens when you work with larger numbers? This method might not work!"	NotGen	InclCrit

EoYTest07/08: Student evaluations of 'Cathy's answer'

Marks:	Comments on 'Cathy's answer':	Code:	
4	"Not proven by numerical example"	CritLackofForm	InclCrit
4	"The proof should be show mathematically as well as in words."	CritLackofForm	InclCrit
5	"This is a to the point answer and basically the same as in Aoife's answer but in word form."	A-C:S	NotInclCrit
4	"Correct answer but show mathematically."	CritLackofForm	InclCrit
4	"She did not give an example"	NoEx	InclCrit
2.5	"Did not mention that numbers must be positive. Perhaps should explain with numbers like 2n."	IncorrectCrit	InclCrit
2	"Do mention of the numbers needing to be whole numbers. Any number can be divided by 2."	PoorDef	InclCrit
4	"Maybe [...] prove that the 2 No.s [...]have the same common factor ??2a+2b = 2(a+b). "	AffNotJust	InclCrit
5	"your making it sound complicated"	Compl	InclCrit
4	"Although Cathy's answer is true there is too much english and does not mathematically prove it unlike Aoife."	CritLackofForm	InclCrit
5	"I believe Cathy's answer is completely correct. She has both defined & logically stated mathematical reasoning behind a statement."	Cl, Def	InclCrit
4	"Would like to see this expanded with a general equation"	CritLackofForm	InclCrit
3	"Use mathematical notation to show this"	CritLackofForm	InclCrit
4	"well thought out."	VG	NotInclCrit
3	"Show an example."	NoEx	InclCrit
4	"Good intuitive answer but needs a mathematical proof."	CritLackofForm	InclCrit
5	"Give clear equation to support your answer"	CritLackofForm	InclCrit
4.5	"Cathy uses sound logic to prove her case, but proving a general case would have made her answer better."	CritLackofForm	InclCrit
0		NoComment	NotInclCrit
5	"gets to the heart of the matter"	VG	NotInclCrit
0		NoComment	NotInclCrit
3	"Need to show that the answer will have the same common factor."	AffNotJust	InclCrit
5	"This is a good answer. It's straight to the point and holds true."	VG, Conc	InclCrit
4	"Odd numbers can be divided by 2 as well, eg 7:2=3.5. If she said even numbers are numbers that are divisible by 2...."	PoorDef	InclCrit
5	"Give example to back up proof."	NoEx	InclCrit
5	"2+4 [2x2] = 6:2, 2+6 [2x3] = 8:2"	VerwithEx	NotInclCrit
4	"Cathy's answer is well written and easy to comprehent although she could sum her answer up by explaining it again at the end	VG, WellExp,	InclCrit
	using formal notation"	Mexpl, Mform	InclCrit
3	"This proof makes sense but she could have used a more mathematical approach."	CritLackofForm	InclCrit
5		NoComment	NotInclCrit
3	"Give an example: to show then 4 [...]"	NoEx	InclCrit
4	"examples"	NoEx	InclCrit
4		NoComment	NotInclCrit
3	"Needs more to be a proof."	PrNotEnough	NotInclCrit
4	"Didn't say what group the even number in question belongs to."	PoorDef	InclCrit
5	"Cathy has proven the statement clearly here in my opinion. She defined even numbers first and made her proof clear."	Cl, Def	InclCrit
4	"No proof given. Must backup the given statement."	NoPr	NotInclCrit
3	"Good but need examples and some way of showing proof."	NoEx, CritLackofForm	InclCrit

EoYTest07/08: Student evaluations of 'Dan's answer'

Marks:	**Comment on Dan's answer:**	**Code:**	
3	"Not proven by numerical example"	CritLackofForm	InclCrit
3	"It should be proven formally even if it is intuitively true."	CritLackofForm	InclCrit
4	"A very short and simple answer which seems to cover all situations."	Conc, Gen, Cl	InclCrit
4	"Correct but no examples shown."	NoEx	InclCrit
3	"It is not proved mathematically or he did not show it using an example."	CritLackofForm, NoEx	InclCrit
4	"Does not account for negative numbers."	IncorrectCrit	NotInclCrit
5		NoComment	NotInclCrit
5		NoComment	NotInclCrit
4	"This is implying the numbers are positive"	IncorrectCrit	NotInclCrit
3	"Again, a good answer just not strong enough"	PrNotEnough	InclCrit
4	"I believe Dan's answer would be more acceptable had he defined an even number"	NoDef	InclCrit
3	"Needs further proof but onto something"	PrNotEnough	InclCrit
3	"True but incomplete. It needs the definition of an even no."	NoDef	InclCrit
4	"well spotted"	VG	NotInclCrit
4	"Need a more general term"	NotGen	InclCrit
2	"An unusual intuitive understanding but not based on any algebraic proof."	CritLackofForm	InclCrit
5	"Give examples to support your answer."	NoEx	InclCrit
4.5	"Examples may have helped."	NoEx	InclCrit
2	"rather than prove it, he's just explained it"	NoPr	InclCrit
4	"Intuitive"	VG	NotInclCrit
0		NoComment	NotInclCrit
4	"Need to show some examples that this is true"	NoEx	InclCrit
3	"This answer works but it doesn't explain why it works."	NoExpl	InclCrit
5		NoComment	NotInclCrit
3	"Not a proof, just a statement"	NoPr	InclCrit
3	"True but why?? Multiples of 2 is even"	NoExpl	InclCrit
1	"Dan has not proved that what he has said is true and is only using a small range of numbers, which is not sufficient."	NotGen	InclCrit
2	"This is a vague proof. He needs to make an example out of what he's saying."	NoEx	InclCrit
5		NoComment	InclCrit
2	"Give working samples. Try to come up with formula."	NoEx, CritLackofForm	InclCrit
4	"examples "	NoEx	InclCrit
4		NoComment	NotInclCrit
5	"It's clear to understand & to the point."	Cl, Conc	InclCrit
5		NoComment	NotInclCrit
3	"Dan's proof here is not complete. He has made a good point but the proof has not quite come full circle."	PrNotEnough	InclCrit
3	"There is no proof of this, it is just an observation."	NoPr	InclCrit
3	"Good but need worked examples."	NoEx	InclCrit

EoYTest07/08: Student evaluations of 'Eve's answer'

Marks:	Comments on Eve's answer:	Code:	
0		NoComment	NotInclCrit
5	"An even number should be clearly defined."	NoDef	InclCrit
1	"In her answer she says x+y=2. She then goes on to say x+y=2z, it seems she has gotten caught up in the algebra and made an error."	AlgWr	InclCrit
0	"Algebra is not correct. You already stated that x+y=z however have come to the conclusion that x+y=2z this is a contradiction."	AlgWr	InclCrit
2	"Didn't explain in words and it is not clear"	NoExpl, NotCl	InclCrit
4	"Needs to account for negative numbers also. Did not mention that even numbers are divisible by 2. "	Misc	InclCrit
4	"Could have made it clearer at the end why it is true."	MExpl	InclCrit
1	"I don't see how this proves at, first you state x+y=z & then x+y=2z, which can't be true unless z=0, x=-y and you are not proving for any combinations at integers, f e.g two [..]"	Misc	InclCrit
0	"There is no mention of even numbers and you made a mistake on your last line"	AlgWr	InclCrit
0	"This result has not even proven true as z+z-(x+y)=2z-z=z."	AlgWr	InclCrit
3	"I believe Eve's answer is lacking as she failed to define an even number."	NoDef	InclCrit
3	"Need more explanation"	MExpl	InclCrit
1	"Yes its true but proof is inaccurate."	AlgWr	InclCrit
5	"well worked out."	VG	NotInclCrit
1	"Need to explain whats happening. If x+y=z then x+y#2z unless z=0, must state this."	NoExpl	InclCrit
1	"This answer makes no sense."	NotGood	InclCrit
2	"More explanation. What about z+z-(x+y) = z+z-z=z since x+y = z"	NoExpl, AlgWr	InclCrit
5	"Maybe an extra line stating that 2z is divisible by 2, therefore it is even would have made the answer better."	MExpl	InclCrit
3	"First mark lost because she didn't say, why it's true. Second because it's unnecessarily complicated."	NoExpl, Compl	InclCrit
0		NoComment	NotInclCrit
0		NoComment	NotInclCrit
2	"Can't understand your explanation"	NoUnd	NotInclCrit
1	"This is a bad answer. It's very difficult to understand why this would work."	NotGood, NoUnd	NotInclCrit
0	"x+y=z not 2z."	AlgWr	InclCrit
1	"Proof is incorrect, x+y cannot be equal to z and 2z."	AlgWr	InclCrit
0	"Proves nothing."	NoPr	NotInclCrit
5	"This is well written as it shows that z is a multiple of 2. Any number which can be divided by 2 is even."	VG	InclCrit
2	"I find this a difficult proof to follow, there are much easier approaches."	Compl	InclCrit
3	"She should explain what she's doing."	NoExpl	InclCrit
3	"Show work more clearly, and Product is not show clear give general example"	NotCl, NoEx	InclCrit
3	"confusing"	Compl	InclCrit
0	"There is no mention of x and y being even numbers."	Misc	InclCrit
4	"A nice way to prove it but it doesn't show it's even. Could say it can't be divided by 2."	VG, Misc	InclCrit
3	"The equations contradict each other, i.e. First z=x+y but at the end x+y=2z, this isn't possible."	AlgWr	InclCrit
0	"Eve has not proven this statement correctly, she started with x+y=z and ended with x+y=2z. I would advise her to begin again."	AlgWr	InclCrit
0	"x+y is not = to 2z = z, all she has done here is reproduce the fact that x+y = z."	AlgWr	InclCrit
4	"Good but maybe try with numbers to further your proof."	VG, NoEx	InclCrit

EoYTest07/08: Student evaluations of 'Finn's answer'

Marks:	Comments on Finn's answer:	Code:		
3	"Proof is illustrated using graphics rather than numbers"	Misc	Int-Pr	NotInclCrit
1	"This proof should be generalized also."	NotGen	Int-Ex	InclCrit
2	"Again Finn's answer only covers 1 solution. He needs to give a general statement."	NotGen	Int-Ex	InclCrit
0	"Answer not clear. Shows it works for only one set of even n numbers, you need to prove it for all even numbers."	NotGen	Int-Ex	InclCrit
2		NoComment	Int-?	NotInclCrit
2	"Intuitively correct but needs to explain why the answer means the statement is true."	NoExplWhySt	Int-Pr	InclCrit
5		NoComment	Int-?	NotInclCrit
1	"Although intuitively you can see the idea, there is no definition of an even no, and also thus set only proves it rigourously for values 12 e 8."	NoDef, NotGen	Int-Pr	InclCrit
5	"Nice pictures you could have written a line explaining it though."	NoExpl	Int-Pr	InclCrit
2	"This is yet another example. It needs to be generalized like Aoife's, although the dots have proven a pattern there is just 1 example."	NotGen	Int-Ex	InclCrit
2	"Finn has failed to explain himself as he answers the question, he also failed to define an even number."	NoExpl, NoDef	Int-Pr	InclCrit
2	"Shows a different approach, between only show 1 case and not a general formula."	NotGen	Int-Pr	InclCrit
3	"Good visual representation but needs notational explanation."	NoExpl, Mform	Int-Pr	InclCrit
0	"I don't understand!"	NoUnd	Int-?	NotInclCrit
1	"Need general example. Not in primary school anymore."	NotGen	Int-Ex	InclCrit
3	"Good visual proof but don't use mathematics."	VG, Mform	Int-Pr	InclCrit
4.5	"Give general equation which suits for huge positive numbers."	NotGen, Mform	Int-Pr	InclCrit
3	"Finn should have given more examples and written that no matter how many even numbers of dots you add to LHS you still get an even Number of dots on the RHS."	Mex	Int-Pr	InclCrit
3	"it makes it clear but lacks explanation and is not a valid mathematical definition"	NoExpl, NoDef	Int-Pr	InclCrit
3	"Does not clearly move from the particular to the general but nice visual approach!"	NotGen, VG	Int-Pr	InclCrit
0		NoComment	Int-?	NotInclCrit
2	"Not really explaining the statement"	NoExplWhySt	Int-?	InclCrit
2	"This proves that it works for 12+8. It doesn't prove it for all cases."	NotGen	Int-Ex	InclCrit
4		NoComment	Int-?	NotInclCrit
1	"Not a proof, just an example"	NotGen	Int-Ex	InclCrit
3	"No explanation. Multiples of 2, therefore even, 2 is an even number"	NoExpl	Int-Pr	InclCrit
2	"Finn has just counted dots but has not given an explanation as to why he thinks the statement is true. Not a sufficient answer."	NoExplWhySt, NoPr	Int-?	InclCrit
0	"This does not prove anything, words and numbers are needed."	NoPr	Int-?	InclCrit
2	"He hasn't proved that it's true for all even numbers."	NotGen	Int-Ex	InclCrit
3	"Good idea, give explanation"	VG, NoExpl	Int-Pr	InclCrit
2	"Only shows 1 eg so doesn't prove that it's true for all"	NotGen	Int-Ex	InclCrit
5		NoComment	Int-?	NotInclCrit
2	"Not a proof."	NoPr	Int-?	InclCrit
3	"There are no words in this proof."	NoExpl	Int-Pr	InclCrit
1	"This is a poor effort to prove this statement. He has just given one example with no explanation. He needs to prove his example."	NoExpl, NotGen	Int-Ex	InclCrit
3	"No explanation given, there is no general formula given"	NoExpl, NotGen	Int-Pr	InclCrit
1	"Finn only gave one solution. Maybe find a way that will show a wider range of solutions."	NotGen	Int-Ex	InclCrit

Appendix D

Interviews 2009: Sketch

First Year Honours Mathematics - Interview(Sketch)
Semester 2 08-09

1. **Introduction**

   (a) *I.* **introduces herself**, **thanks** for the student's participation, offers tea, coffee and biscuits, etc. She tries to create a relaxed atmosphere.

   - *Thank you for agreeing to participate in this research project. I very much appreciate your help.*
   - *My name is .... I am working on a research project in Mathematical Education. I am interested in the transition from school to university maths. At the moment I focus on the criteria students use when validating mathematical arguments.*

   (b) The **students are being assured that the interviews are treated anonymously** and asked for permission to use the transcripts for the research project.

   - *This talk will be recorded and then transcribed.*
   - *What you say may be quoted in our research. However, should this occur, we want to reassure you that your identity will not be divulged - we will use a pseudo-name as appropriate. Are you happy with this?*
   - *We would ask that you don't discuss the questions you are asked with any of your classmates who haven't had their meetings with me yet.*

   (c) *I.* **tells the student what to expect of this meeting:** first talking about their experience when studying maths at NUI Galway so far, then validating and discussing some student generated answers.

   - *I am interested in your experience of studying maths at University. In the first part of this meeting I would like you to share your thoughts about this.*
   - *In the second part of our meeting I will show you two not too difficult mathematical tasks and a few solutions to those. My questions on each of those answers will be for example:*
     - *Does the answer prove the given statement?*
     - *What do you like/ not like about the particular answer?*
     - *What advice would you give that student to improve the answer?*

     Finally I will ask you to rank the given answers.
   - *I will try not to interfere unless I get the impression a hint would be helpful and not hinder your own work.*

2. **Reflections about the transition**

   - *When and where did you do your Leaving Cert? Pass or Honours?*
   - *Did you like maths in school?*
   - *Were you good at maths in school?*
   - *Your first year in college is almost finished. How did you like it?*
   - *Studying Maths at University: what is different from school mathematics?*
   - *To help the next incomers: what advice would you give them?*

3. **Validating of mathematical arguments**

   Task 1. Consider the following statement: The squares of all even numbers are even, and the squares of all odd numbers are odd.

   (a) *I.: Consider this statement for a few minutes, please.*

   (b) *I.* gives student the six prepared answers

   (c) (in a pile) and asks him/her about each of those:

   - *What do you think of this answer?*
   - *Does the answer prove the statement?*
   - *What do you like/ do you not like about this answer?*
   - *What advice would you give [name]? How could he/she improve the answer?*

   (d) *I.* asks the student to rank the different answers.

   (e) *I.* tries to to get some information about students' validation criteria.

   - *How would you describe a good proof?*
   - *Why did you [name1's] answer prefer to [name2's]?* [I. refers to certain answers.]

   Task 2. Let f be a quadratic function, $f(x) = ax^2+bx+c$ with $a, b, c \in \mathbb{R}$ and $a > 0$. Show: f can't have more than two common values with its derivative f'.

   (a) *I.* gives student the five prepared answers and asks him/her about each of those:

   - *What do you think of this answer?*
   - *Does the answer prove the statement?*
   - *What do you like/ do you not like about this answer?*
   - *What advise would you give [name]? How could he/she improve the answer?*

   (b) *I.* asks the student to rank the different answers.

   (c) *I.* tries to get some information about students' validation criteria.

      - *How would you describe a good proof?*
      - *Why did you [Name 1's] answer prefer to [Name 2's]?* [*I.* refers to certain answers]

4. At the end of the meeting *I.* tries to clarify certain terms the student used during the interviews, eg 'mathematical', and any outstanding or unusual opinions which occured during the interviews.

5. Finally the student is being asked to sign a **consent form**.

Appendix E

Interviews 2009: The Tasks

Interviews09: The Tasks

1. 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.

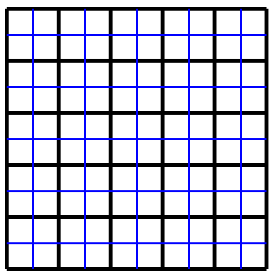
   Benny's answer:

$2^2 = 4$ even	$3^2 = 9$ odd
$4^2 = 6$ even	$5^2 = 25$ odd
$6^2 = 8$ even	$7^2 = 49$ odd
$20^2 = 400$ even	$23^2 = 529$ odd
$(-16)^2 = 256$ even	$(-19)^2 = 361$ odd

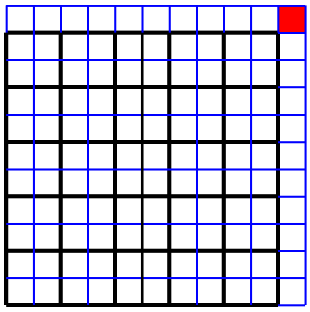
   The squares of even numbers are even, and the squares of odd numbers are odd.

   Ciara's answer:

   Square of an even number:

Square of an odd number:



Darragh's answer:

k is any whole number.
l is any whole number.
$2k$ is an even number.
$2l + 1$ is an odd number.
$(2k)^2 = 4k^2 = 2(2k^2)$ is even.
$(2l + 1)^2 = 4l^2 + 2l + 1 = 2(2l^2 + l) + 1$ is odd.
So the statement is true.

Elaine's answer:

Proof: Let x be an even number and let y be an odd number. Then $x + y$ and $x - y$ are both odd.

$$(x + y)^2 = x^2 + 2xy + y^2, \quad (x - y)^2 = x^2 - 2xy + y^2$$

So $(x + y)^2 + (x - y)^2 = 2x^2 + 2y^2 = 2(x^2 + y^2)$.
The number $2(x^2 + y^2)$ is even obviously. Dividing by 2 we find that $x^2 + y^2$ is odd.
Then x^2 and y^2 can't be both even or both odd. So x^2 is even and y^2 is odd.

Fintan's answer:

Let $2a$ and $2b$ be any even numbers.
$(2a) * (2b) = 4ab = 2(2ab)$ is even.
Therefore the product of two even numbers is even, and so in particular the square of an even number is even as well.

Let $(2a + 1)$ and $(2b + 1)$ be any odd numbers.
$(2a + 1) * (2b + 1) = 4ab + 2a + 2b + 1 = 2(2ab + a + b) + 1$ is odd.
Therefore product of two odd numbers is odd, and so in particular the square of an odd numbers is odd as well.

2. **Let f be a quadratic function, $f(x) = ax^2 + bx + c$ with $a, b, c \in \mathbb{R}$.**

   Show: f can't have more than two common values with its derivative f'.

   Gerard's answer:

   $f(x) = 2x^2 + 3x - 7$
   $f'(x) = 4x + 3$

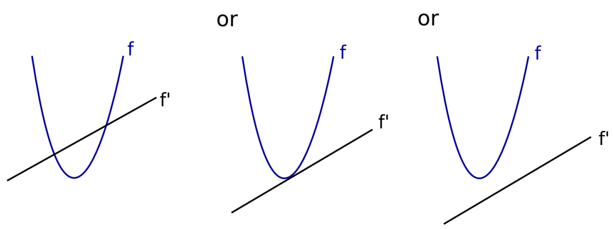
   $2x^2 + 3x - 7 = 4x + 3$
   $2x^2 - x - 10 = 0$
   $(2x - 5)(x + 2) = 0$
   $x = \frac{5}{2}$ or $x = -2$ $\longrightarrow$ 2 solutions

   So the statement is true.

   Helena's answer:

   The derivative of a quadratic function is a linear expression. To find common points I solve the equation I get from setting both expressions equal. Setting a quadratic expression equal to a linear expression gives a quadratic equation which can't have more than two solutions.

   Ian's answer:



   Joan's answer:

   Proof: $f(x) = ax^2 + bx + c$, so $f'(x) = 2ax + b$. We want solutions to $f(x) = f'(x)$, i.e.

$$ax^2 + bx + c = 2ax + b.$$

   Multiplying by x gives

$$\begin{aligned} ax^2 + bx + c &= 2ax^2 + bx \\ \Longrightarrow ax^2 &= c \\ \Longrightarrow x^2 &= \frac{c}{a} \end{aligned}$$

   So the only two possible solutions are given by $x = \pm\sqrt{\frac{c}{a}}$.

Kieran's answer:

$f'(x) = 2ax + b$
$f(x) = f'(x)$

$\Leftrightarrow ax^2 + bx + c = 2ax + b$
$\Leftrightarrow ax^2 + (b - 2a)x + c - b = 0$

is a quadratic equation which has at most two solutions.

Appendix F

Interviews 2009: Coding Scheme

Overall Adjustments:

Und	The student understands the approach.
NoUnd	The student does not understand the approach.
Und?	The students' comment does not indicate clearly whether he understands the approach.
Und-I	The student does not understand the proposed approach until it has been explained by the interviewer.
C-Ex	The student interprets Ciara's answer as description of one example.
VG	The student expresses in some way that she likes the answer. (The comment does not indicate whether the proof is regarded correct and sufficient.)
Conf	The student finds the proposed answer confusing.
GAppr	The student regards the proposed approach as good, but not accomplished.

Particular Judgements and Comments:

C	The student regards the answer as *"correct"*, meaning 'sufficient' .
NotC	The student regards the answer as **not** *"correct"*.
PC	The student regards the answer as **partly** *"correct"*.
C?	The student does not decide whether s/he regards the answer as *"correct"* or not.
(..)→(..)	The student changes his/her opinion during the interview-reflection.
PC−MMEnd	The student regards the answer as correct except minor mistakes in the end.
Pr	The student regards the answer as proof of the statement.
NoPr	The student does not regard the answer as proof of the statement.
PrNotEnough	The student appreciates the approach, but regards it as not enough to be a proof.
NotConvEnough	The student appreciates the approach, but regards it as not convincing enough to be a good proof.
PrNotImpl	The student interprets the answer as description of a proof, which is not implemented.
Pr?	The student is not sure whether the proposed approach is a sufficient proof of the statement or not.

MPr	The student notes that the proposed argument proves a more general statement.
Ex	The student mentions explicitely that the answer is just one or two examples.
NoEx	The student criticises the lack of examples.
St	The student regards the answer as a statement (in comparison to a proof).
T	The student detects typos.
Err	The student detects one or more errors.
Gen	The student mentions explicitely the general applicity of the answer.
Expl	The student appreciates the quality of explanation in the proposed answer.
Cl	The student appreciates the 'clearness' of the proposed answer.
Form	The student appreciates the formal appearance of the proposed answer.
Conc	The student appreciates that the proposed answer is concise.
NoGen	The student criticises a lack of generalisation.
NoForm	The student criticises a lack of formula.
NoDef	The student criticises a lack of definitions.
NoExpl	The student criticises a lack of explanation.
MExpl	The student would prefer to see more explanations.
MGr	The student would prefer to see a graph.
AffNotJust	The student criticises that in the proposed argument affirmations are being stated without justification.
ApprDid	The student appreciates a didactic value of the approach.

Comments to Approaches in Comparison:

A-B:S	The student interprets Anna's and Benny's answer as similar.
D(...)	The student prefers Darragh's answer to those in brackets.
A:MPr(B)	In the student's opinion Anna's answer is more like a proof than Benny's.
B:WR(A)	The student prefers Anna's approach to Benny's as her answer involves a wider range of examples.
C:MTh(AB)	The student prefers Ciara's answer to Anna's or Benny's because it is more substantial, (*"more thought in it"*).
MI	The students favors the proposed answer to others because she considers it more *"interesting"*.
D:ExplWhy(...)	The student prefers Darragh's to those in brackets because this approach involves more explanation **why** the the statement is true.
D:MSucc(F)	The student prefers Darragh's answer to Fintan's because s/he finds it more succinct.
H:MGen(G)	The student prefers Helena's answer to Gerard's because s/he finds it more general.
J:MImpl(H)	The student prefers Joan's answer to Helena's because in his/her opinion Helena did not accomplish the task.
MCompl(...)	The student finds the proposed answer more complicated than those in brackets.
MConf(...)	The student finds the proposed answer more confusing than those in brackets.
Why2Var	The student criticises that Darragh uses two variables.
R-Succ	When ranking the students appreciates that the proposed answer is succinct.
R-NotExpl	When ranking the student criticises that the proposed answer provides less explanation than the others.
R-Good	The student decides in the ranking process that the proposed answer is good.