UNDERGARDUATE STUDENTS' USE OF DEDUCTIVE ARGUMENTS TO SOLVE "PROVE THAT…" TASKS

Paola Iannone[1] and Matthew Inglis[2]

[1] School of Education and Lifelong Learning, University of East Anglia, UK

[2] Mathematics Education Centre, Loughborough University, UK

In this paper we report findings from an investigation of 222 proof attempts produced by 74 year-one undergraduate mathematics students at a university in the UK. We classify the proofs according to an extended classification originally used by Stylianides and Stylianides (2009). We found that already at the beginning of Year 1 most undergraduate students in our sample associate the request for a proof of a statement to the production of a deductive argument. Moreover, when students failed to produce a correct proof this was mostly because of difficulties in producing deductive arguments. We suggest that more attention should be given to the process of production of deductive argument rather than to the types of non-deductive arguments that some students produce as proofs.

INTRODUCTION

Undergraduate students' difficulties with proof production at university level have been widely documented in the literature (Moore, 1994; Harel and Sowder, 1998; Weber, 2001; Selden and Selden, 2003; just to cite a few examples). Indeed in 2009 the ICMI Study 19 was dedicated to "Proof and proving in mathematics education" with one study group solely focusing on proof and proving at tertiary level. Much of the attention of researchers in this field has been devoted to the types of arguments that undergraduate students produce when asked to solve "prove that…" tasks. One of the most popular theoretical frameworks for analysing students' proofs (not only at undergraduate level) has been proposed by Harel and Sowder (1998, 2007). This framework offers a comprehensive taxonomy of students' Proof Schemes, where

> a person's proof scheme consists of what constitute ascertaining and persuading for that person. (Harel and Sowder, 1998, p.244)

It is beyond the scope of this paper to give a detailed description of this framework. For the moment it will suffice to say that Harel and Sowder (1998) call an Analytic Proof Scheme

> … one that validates conjectures by means of logical deduction. (ibid. p.258)

and indicate that this is the proof scheme that students should aspire to and the one generally held by the community of mathematicians. Because of the prominence that Harel and Sowder (1998) give to deductive arguments, and the statement that this is indeed the proof scheme shared by professional mathematician, we will take deductive arguments to be the desired outcome of "prove that…" tasks. In the studies cited by Harel and Sowder (2007) a picture emerges of university students still uncertain about what the role and place of proof at university level is, with students gaining conviction about mathematical statements through various types of arguments, including many students relying solely on empirical arguments. To cite just one more example of research in this direction, Recio and Godino (2001) reported that only very few of the university students in their sample were successful in proof tasks and that 40% of the students in the sample relied on empirical arguments as proof. Whilst proof schemes offer a "truly comprehensive" (Harel and Sowder, 2007) perspective on learning and teaching proof, it can be argued that much of the focus of this framework is on the *types* of arguments that students find convincing when asked to solve proof tasks rather than the *process* they use to produce such arguments. In this paper we examine what types of arguments year-one students at a high-ranking university in the UK produce when asked to solve proof tasks and we argue that, at least in this case, students are aware of the requirement of proof but at time have their difficulties lie in the process of producing correct deductive arguments.

THE STUDY

The main aim of the study reported in this paper was to find out whether self-efficacy is an accurate predictor of academic performance, and in particular of proof production. For the scope of this study we have defined self-efficacy, following Bandura (1977), as the judgment students make of their own capability of performing a given task (in our study the task is proof production). The results of this part of the study were presented by Iannone and Inglis (2010). For the scope of his paper we have analysed the proofs that the students produced in the second part of the questionnaire they were asked to fill in. Seventy-six first year students in mathematics (or on a joint degree with a substantial mathematics component such as computer science or natural sciences) at a high-ranking university in the UK took part in this study. Data collection took place in week 8 of the first semester during one of the Linear Algebra lectures. A booklet of questions was given to each participant, and the students were asked to work through the questions at their own pace. The first section of the booklet consisted of 28 statements; ten consisted of our Proof Self-Efficacy Scale (Iannone and Inglis, 2010), designed following Bandura's (2006) guidelines. A further ten statements were formed from the items of the General Self-Efficacy Scale (Schwarzer and Jerusalem, 1995). Eight extra items were also introduced to readdress the balance between forward- and reverse-scored items (taken from the Need for Cognition scale, Cacioppo and Petty, 1982). The participants were asked to read and decide the extent to which they believed the statements were characteristic of them, using a five-point Likert Scale (from "extremely uncharacteristic" through to "extremely characteristic"). The order in which the statements appeared was randomised for each participant. The second part of the booklet consisted of four proof construction tasks novel to the students and designed in collaboration with a mathematics lecturer so that they would represent appropriate tasks for this cohort, in the sense that they were tasks similar to ones that the students had encountered during the lectures. The order in which the proof tasks appeared was again randomised for each participant. The participants were given 20 minutes to work on the proof tasks.

The proof tasks included in the booklet were:

A. Prove that the sum of two odd numbers is even.

B. Prove that the sum of the first *n* natural numbers is equal to $\frac{1}{2}n(n+1)$.

C. Let *d*, *a* and *b* be integers. Prove that if $d \mid a$ and $d \mid b$ then $d^2 \mid (a^2 + b^2)$.

D. Prove that if the sum of the digits of a natural number is divisible by 3 then the number itself is divisible by 3.

The proof Task D was introduced even if after the meeting with the mathematics lecturer it was deemed to be too demanding for this cohort. As we will see in the analysis of the data this was indeed the case. A marking scheme was devised to evaluate the proofs. Each proof was marked out of five. In Appendix we report the questions in the questionnaire, the marking scheme and the model solutions to the tasks. We analysed the proofs and classified them following the refinement of a classification originally developed by Stylianides and Stylianides (2009). We refined the classification because we not only wanted to find out which proofs were correct, but we also wanted to investigate the types of arguments that the students used when asked to produce a proof. The categories used in our classification were:

Type	Description
M1	Correct proof
P1	The last part of the proof is missing
P2	The hypothesis are expressed correctly but the proof stops after the statement of the hypothesis
P3	The solution does not represent the most general case
P4	The solution resembles the correct proof but there are not enough details to see whether this is correct or not
P5	The solution follows a correct deductive argument but some mistakes in the calculations occur
M4	The solution consists of an empirical argument
M5	Some mathematical statement is presented but this is unrelated to the proof requested
0	Task left blank

Table 1: Responses categories

Before we discuss the classification of the proof tasks[1] we give here some examples to clarify the categories mentioned above. We assume that the categories M1 and 0 are self-explanatory so we will very briefly discuss the other categories giving some examples from the students' work.

P1 – The last part of the proof is missing

Solutions in this category comprise deductive arguments without conclusions. Typical examples here are proofs by induction (Task B) which stop at the end of the inductive step.

P2 - The hypothesis are expressed correctly but the proof stops after the statement of the hypothesis

In several of the solutions for the proofs Tasks A and C students wrote the hypothesis correctly (for example stated correctly that $d \mid a$ means $a=kd$ with k an integer number) but were unable to proceed. Attempts of proof of this kind resonate with Moore's (1994) findings who observed that students can sometimes state the hypothesis in the theorem to prove, but are unable to start the proof.

P3 - The solution does not represent the most general case

This category was mostly found in the solution of Task A. An example is reported below:

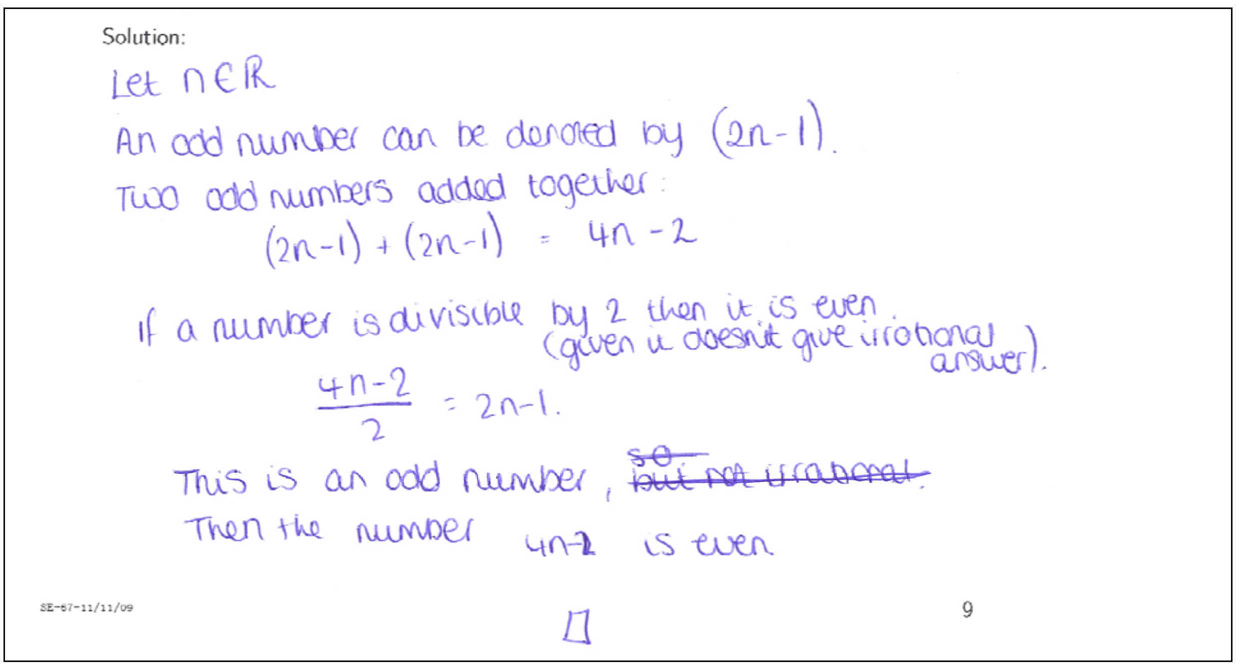


Fig. 1: Example of category P3

[1] We note that the denomination of the M tasks is the same denomination used in Stylianides and Stylianides (2009). We only have categories M1, M4 and M5 as the others in that classification are not relevant to our study. The P-denominations are ours.

In this case the student failed to express the most general case, proving instead the statement: the sum of two equal odd numbers (i.e. twice an odd number) is even.

P4 - The solution resembles the correct proof but there are not enough details to see whether this is correct or not

An example in this category is the following:

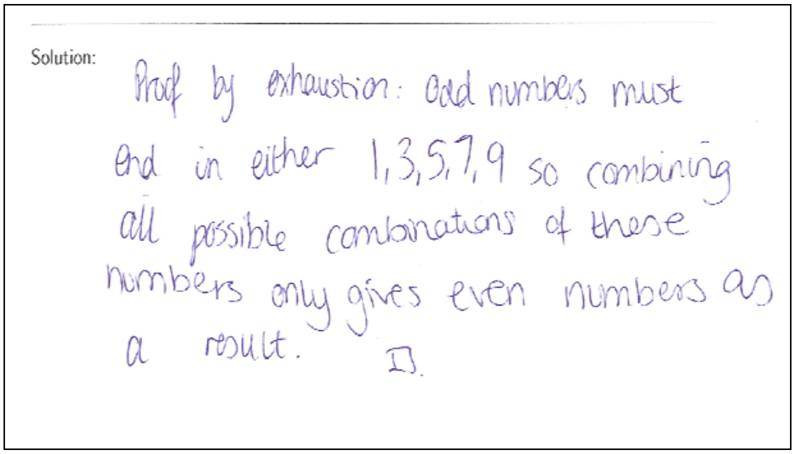

Solution: Proof by exhaustion: Odd numbers must end in either 1,3,5,7,9 so combining all possible combinations of these numbers only gives even numbers as a result. □.

Fig. 2: Example of category P4

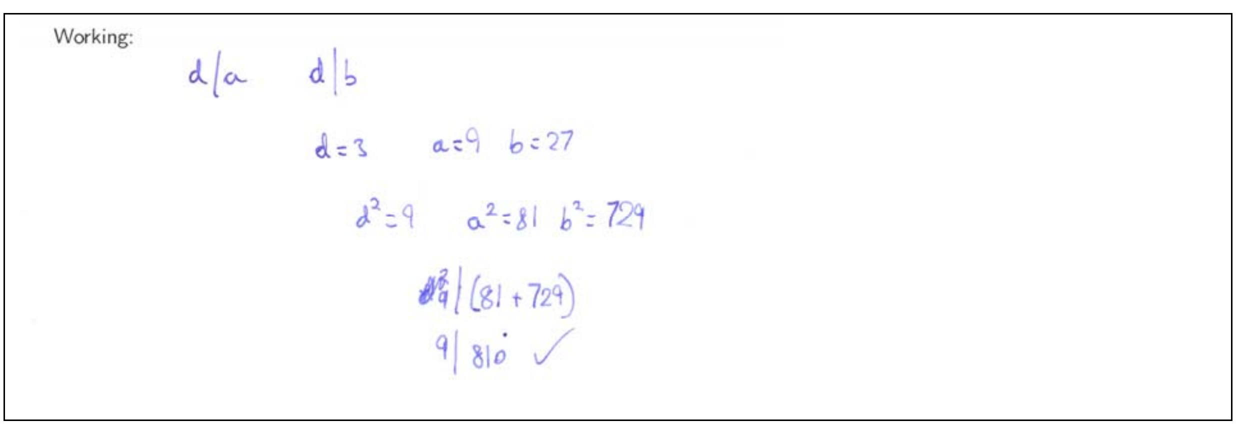
In this example the student has tried to give a verbal argument for the proof. However lack of clarity and details mean that it is not possible to ascertain whether the argument is correct or not.

P5 - The solution follows a correct deductive argument but some mistakes in the calculations occur

Typically this occurred in the proofs by induction where the students made a mistake in the calculations for the inductive step.

M4 - The solution consists of an empirical argument

In this category we have grouped examples of "proof by example" where students gave a numerical example as a proof. This is clear in the following example:


Working:

$d|a \quad d|b$

$d=3 \quad a=9 \quad b=27$

$d^2=9 \quad a^2=81 \quad b^2=729$

$9|(81+729)$

$9|810$ ✓

Fig. 3: Example of category M4

Here the student gave one numerical example and terminated the answer with a tick – perhaps signaling that s/he was satisfied with the answer or was satisfied with having found a confirming example.

M5: Some mathematical statement is presented but this is unrelated to the proof requested

Responses in this category included all sorts of unrelated mathematical facts, but no recognisable deductive arguments.

ANALYSIS AND FINDINGS

Two participants were removed from the analysis of the data as they attempted none of the proof tasks. This left us with 74 responses.

A classification of the proof tasks excluding Task D, according to the categories outlined above yield to the following table:

	A	**B**	**C**	**Total**
M1	28	34	21	83
P1	0	11	5	16
P2	3	8	8	19
P3	30	0	1	31
P4	3	2	2	7
P5	0	3	6	9
M4	0	0	2	2
M5	7	4	6	17
0	3	12	23	38
Total	74	74	74	222

Table 2: Classification of the proof tasks excluding Task D

We excluded Task D as it was solved correctly only two times across the sample, with only five students accruing marks for it.

The distribution of the marks over the three remaining proof tasks had mean 7.76 and standard deviation 4.35, suggesting that the three tasks were suitable for this cohort of students (i.e. not at ceiling or floor levels). As for the classification of the proof tasks Table 2 shows that 37% of the total number of tasks attempted by the students were correct, with the task solved correctly the most times being task B (solved correctly 34 times) followed by Task A (solved correctly 28) times and Task C (solved correctly 21 times). Considering categories P1, P5 and P3 as (attempts at producing) deductive arguments and category P2 as manifestation of the inability to produce a deductive argument (we infer this by the formal statement of the hypothesis but the lack of any other writing to follow) the distribution gives the following frequencies:

Deductive arguments (M1, P1, P2, P3, P5)	158
P4	7
M4	2
M5	17
0	38
Tot	222

Table 3: Summary of the types of arguments used across the tasks

Table 3 shows that more than two thirds of the arguments that the students produced (or attempted to produce) were in fact deductive arguments. Only two arguments produced out of 222 were empirical arguments.

CONCLUSIONS

In this paper we have classified 222 solutions to proof tasks collected during a study involving year-one mathematics students in a university in the UK. Our classification of proof tasks aimed at investigating what type of arguments students produce when asked to give a proof a mathematical statement. We found that the students in our sample, at a very early stage in their university degree, know largely what the result of a proof construction task ought to be (at least in algebra), namely a deductive argument. Their difficulties were mostly related to failure to produce a correct deductive argument rather than failure to recognise that a proof involves some kind of deductive argument. Remarkably, in our sample, we found only 2 "empirical proofs" pointing to the fact that our students on the whole did not regard investigating one or more examples of a conjecture to be sufficient proof. Our findings resonate with other findings in the research literature (e.g. Weber, 2001; Weber & Alcock, 2004) where researchers have repeatedly pointed out to the difficulties students have in producing correct deductive arguments. Our findings also seem to be in contrast with other research findings on proof production (for example Recio and Godino, 2001) where researchers have pointed out the reliance of university students on empirical arguments to solve "prove that…" tasks. One possible explanation for this is that the students in some of these research projects are not always mathematics students. For example in the paper by Recio and Godino (2001) we have cited previously, participants were "students who took a mathematics subject in different faculties and polytechnic" (pg. 84). Although the authors do not give more details about their sample, we argue that this could include students from many departments and not necessarily sciences. The participants in our sample were mostly mathematics students, with some students coming from other disciplines with a strong mathematics component (e.g. computer science, natural sciences, mathematics with economics). Perhaps it is the background of the students that accounts for such discrepancies in the findings. The data we have presented seem to indicate that the students in our sample (e.g. students with a strong mathematics background) already operate with an analytical proof scheme, and therefore the goal of instruction in this case should be to help the students become versed in the production of deductive arguments rather than the elaboration of their proof schemes. If we confront this with Harel and Sowder (2007) where they write:

> We emphasize again that despite this subjective definition the goal of instruction must be unambiguous—namely, to gradually refine current students' proof schemes toward the proof scheme shared and practiced by contemporary mathematicians.
>
> (Harel and Sowder, 2007, pg. 7)

we would argue that perhaps more attention (in research and in teaching) should be given to the *process* by which students produce evidence to gain conviction about statements and write deductive arguments to produce proofs rather than to what are the *types* of evidence that students at university level offer as proofs.

REFERENCES

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review*, 84, 191-215.

Bandura, A. (2006). *Self-Efficacy Beliefs of Adolescents.* Information Age Publishing.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42, 116-131.

Harel, G., & Sowder, L. (1998). Students proof schemes. *Research in collegiate mathematics education,* 3, 234-282.

Harel, G., & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. In F. K. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (p. 805-842). Charlotte, NC: Information Age Publishing.

Iannone, P., & Inglis, M. (2010). Self-efficacy and mathematical proof: are undergraduate students good at assessing their own proof production ability? *Proceeding of the 13th Conference on Research in Undergraduate Mathematics Education*. Raleigh, North Carolina. Available at http://sigmaa.maa.org/rume/crume2010/Abstracts2010.htm

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

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

Schwarzer, R., & Jerusalem, M. (1995). Generalized self-efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A users portfolio. causal and control beliefs* (p. 35-37). Windsor, UK: NFER-Nelson.

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.

Stylianides A. J., & Stylianides G. J. (2009). Proof constructions and evaluations. *Educational Studies in Mathematics* 72(2), 237-253.

Weber, K. (2001). Students difficulty in constructing proofs: The need for strategic knowledge. *Educational Studies in Mathematics*, *48*(1), 101-119.

Weber, K., & Alcock, L. (2004). Semantic and syntactic proof productions. *Educational Studies in Mathematics*, 56, 209-234.

APPENDIX

In this appendix we report the questions included in the questionnaire we gave to the students (labelled here Task X), a solution designed according to what the students would have seen in the lectures and the marking scheme (the last two items not included in the questionnaire).

Task A: Prove that the sum of two odd numbers is even.

Proof: Let $n=2k_1-1$ and $m=2k_2-1$ be odd numbers. We have

$m+n = 2k_1-1 + 2k_2-1 = 2(k_1 + 2k_2)-2$ even.

2 points for expression of odd number; 2 for expression of sum; 1 for conclusion

Task B: Prove that the sum of the first n natural numbers is equal to $\frac{1}{2}n(n+1)$.

Proof: Base For $n=1$ $\frac{1}{2}n(n+1) = 1$.

Suppose this is true for $n-1$ show this is true for n. We have

$S_{n-1} = \frac{1}{2}(n-1)(n-1+1) = \frac{1}{2}(n-1)(n)$.

Add n and obtain $S_n = \frac{1}{2}(n-1)(n)+n = \frac{1}{2}n(n+1)$.

Hence this is true for all n in N.

2 points for base step; 2 for inductive step; 1for conclusion.

Task C: Let d, a and b be integers. Prove that if $d \mid a$ and $d \mid b$ then $d^2 \mid (a^2 + b^2)$.

Proof: $d \mid a \Rightarrow a = kd$ and $d \mid b \Rightarrow b = md$, hence

$(a^2 + b^2) = (kd)^2 + (md)^2 = k^2d^2 + m^2d^2 = d^2(m^2 + k^2) \Rightarrow d^2 \mid (a^2 + b^2)$.

2 points for expression of divisibility; 2 for expression of sum of squares; 1 for conclusion.

Task D: Prove that if the sum of the digits of a natural number is divisible by 3 then the number itself is divisible by 3.

Proof: Let d be a natural number. Let $d = d_k d_{k-1} \ldots d_0$ its expression in digits. If we expand this expression in base 10 we have

$d = 10^k d_k + 10^{k-1} d_{k-1} + \ldots + d_0$ we can write this as

$d = (10^k - 1)d_k + (10^{k-1} - 1)d_{k-1} + \ldots 9d_1 + d_k + d_{k-1} \ldots + d_1 + d_0$

Note that all the items in the form $10^x - 1$ are divisible by 3 (they are in fact a string of 9s). So we can write

$d = 3A + d_k + d_{k-1} \ldots + d_1 + d_0$.

Hence d is divisible by 3 if and only if $d_k + d_{k-1} \ldots + d_1 + d_0$ is divisible by 3.

2 points for expression base 10; 2 for changing into $(10^{k-1} - 1)$ etc; 1 for conclusion.