In Stylianou, D.A, Blanton, M. L. & Knuth, E.J. (Eds) *Teaching and Learning Proof Across the Grades K-16 Perspective* Lawrence Erlbaum Associates pp171- 191

CHAPTER 10: From Empirical To Structural Reasoning In Mathematics: Tracking Changes Over Time

Dietmar Küchemann and Celia Hoyles

Institute of Education, University of London

A major challenge in mathematics education is to develop students' abilities to reason mathematically, that is to make inferences and deductions from a basis of mathematical structures, henceforth referred to as structural reasoning, rather than by arguing for example from perception, the assertion of authority, or, in particular, from empirical cases—henceforth referred to as empirical reasoning (for a comprehensive perspective on proof that takes account of all its cognitive, social and well as mathematical constraints, see Harel & Sowder, 1998). Lampert (1990) argues that a common view of mathematics, in the world at large and in most mathematics classrooms, is one "in which *doing* mathematics means following the rules laid down by the teacher; *knowing* mathematics means remembering and applying the correct rule when the teacher asks a question; and mathematical *truth is determined* when the answer is ratified by the teacher" (emphasis in original). Formal proofs and consistent mathematical argument both require the ability to reason by appealing to the logical structure of the system, that is to engage in structural reasoning. This is a core component of being able to prove mathematically and of developing mathematical understanding.

If, therefore, it is the case that most students have "never learned what counts as a mathematical argument" (Dreyfus, 1999), it is perhaps not surprising that a considerable body of research has accumulated which indicates that school students tend to argue at an empirical level rather than on the basis of mathematical structure (Bell, 1976; Balacheff, 1988; Coe & Ruthven, 1994; Bills & Rowland, 1999; Healy & Hoyles, 2000).

This chapter describes patterns in high-attaining students' mathematical reasoning in the domain of number/algebra and traces development over time in their use of structural reasoning. The analysis presented forms part of The Longitudinal Proof Project (Hoyles & Küchemann: http://www.ioe.ac.uk/proof), which analysed the development of students' mathematical reasoning over three years. Before describing the research and its findings we briefly summarise the approach to mathematical reasoning adopted in England, which is rather different from that of other countries.

Learning To Prove: A Perspective From The English Curriculum

In the 1950s and 60s, academic students, i.e., the 20-30% of secondary school students who were in selective (grammar) schools, met proof mainly in the context of classic Euclidean geometry. However this systematic treatment of proof more or less disappeared from mathematics curricula for 11-16 year olds during the 1970s and 80s. Thus, for example, Pythagoras' Theorem became known as Pythagoras' Rule, which students were no longer required to prove but only to apply, perhaps after verifying it through examples drawn on squared paper. In this century, proof has started to make a comeback. Thus, for example, the current National Curriculum for students in English schools requires that most Year 8 students (12+ year olds) should:

> Understand a proof that the sum of the angles of a triangle is 180˚ and of a quadrilateral is 360˚, and that the exterior angle of a triangle equals the sum of the two interior opposite angles (DfEE, 2001, p183).

Here the suggested approach is exploratory, and the proof-construction informal. Thus, for *Angle sum of a triangle*, it is suggested that students "Consider relationships between three lines meeting at a point and a fourth line parallel to one of them," and that they "explain using diagrams." There is no requirement that the components of a proof be made explicit or that the argument is set out in a formal way, although more recent government-sponsored (but non-statutory) support material is encouraging teachers to do so. Despite the increasing emphasis on proof in the English national curriculum for 11-16 year olds, it is still likely to remain very different from tightly regulated activities in high school geometry in the USA, as described in Chapter 12 (McCrone & Martin) of this volume. Reference for example the kind of formal proof shown in "Linda's answer" in Appendix A of that chapter. It is inconceivable that English 11-16 year olds will be required to construct or even consider proofs with statements as formal as "All right angles are congruent" and "By the reflexive property, XY = YX."

As well as being less formal, proving in England tends first to be encountered in the number/algebra domain, rather than in geometry as in most other countries (Hoyles, 1997). In fact even after a systematic treatment of proof had virtually disappeared from the English school curriculum, students did have the opportunity (potentially at least) to engage in explanation and (informal) proof through extended 'investigations,' usually in number/algebra. During the 1990s investigational work was incorporated into the national mathematics examination at age 16. Marking schemes were devised to describe the characteristics of a 'good investigation' and as a consequence, the task of 'doing' investigations became increasingly procedural and routinised, with an emphasis on generating data and looking for number patterns, even at the upper end of secondary school (Morgan, 1997). Of course, inductive reasoning based on specific cases can be important and fruitful in mathematics (Polya, 1954). It can help students develop a feel for a mathematical situation and to form conjectures. It also provides a test for the validity of a general proof, especially where students are uncertain about the scope and logic of their argument (Jahnke, 2005), which is an issue we return to later in this chapter. Cuoco, et al. (1996), in their discussion of mathematical 'habits of mind,' suggest that students should learn to become 'pattern sniffers,' 'experimenters' and 'describers,' amongst other things. However, they make the point that the most important habit is to understand "when to use what." We suggest that a crucial habit is to look for mathematical structure, or, as Dreyfus (ibid) puts it, to move "from a computational view of mathematics to a view that conceives of mathematics as a field of intricately related structures." This shift would seem to be particularly important in England where, as we have argued above, the curriculum tends to emphasise data and computation and teaching does not emphasise the importance of structural reasoning (see Healy & Hoyles, 2000). This chapter seeks to throw light on this shift and how it is exemplified in any changing patterns of student response to our proof items.

The Longitudinal Proof Project

The analysis presented here forms part of The Longitudinal Proof Project (Hoyles

& Küchemann: http://www.ioe.ac.uk/proof), which analysed the development of students' mathematical reasoning over three years. Data were collected through annual administration of a proof test completed by classes of high-attaining students[1] in randomly selected schools within nine geographically diverse English regions. The items in all three proof tests were devised after reviews of the literature and of the curriculum, followed by extensive discussions with teachers. They ranged over the following proof 'categories' (although after piloting some categories were dropped as the items turned out to be unsatisfactory).

> Making conjectures; turning conjectures into conditional statements; making and expressing generalisations by engaging in structural reasoning; using generic examples; crucial experiment; general cases which are then limited; given a statement, find (deduce) the value of an unknown, or derive another statement; logical implication; using definitions and structures; transformational reasoning; specialisation after a proof; scrambled proof; reasoning from perception.

Items were piloted with up to about 200 students and overall, 1512 students from 54 schools completed all three tests. The tests comprised items in number/algebra and in geometry, some in open response format and some multiple choice. Each new proof test included some items that were identical or very similar to items from the previous test (*core* items), together with new items. The project used a combination of quantitative and qualitative methods. The quantitative methods included the identification of trends in

[1] Students in England are setted (or tracked) and we targeted students who would be in top sets in Year 10 (age 14+ years).

hierarchically-ordered categorical data obtained by coding students' responses to each item in each proof test, and multilevel analyses of student scores in geometry and in algebra to identify significant predictors of progress. The qualitative methods included analyses of interviews with selected students in schools identified from the multilevel modeling as those in which students performed significantly better than predicted (see Hoyles, et al., 2005, for a detailed description of the methodology adopted).

Overall in the area of number/algebra, we found an improvement in the use of algebra, though many in our sample of high attaining students were quite strongly attracted to pattern spotting and computation (as we shall see in this chapter). Also, we found a large gulf between success in a numerical and an equivalent algebraic task. Thus in Yr 10, 88% of students could solve a problem of calculating angles but only 21% were able to re-express the calculation as an algebraic relationship.

In this chapter we focus on two questions in number/algebra, A1 (see Figure 1) and A4 (see Figure 3), that both sought to assess whether students engaged in structural reasoning as opposed to appealing to computation or empirical data. As stated earlier, we regard structural reasoning as a core component of mathematical proof. Both items featured in more than one proof test so it was possible to identify changes in patterns of response between annual surveys.

Question A1 (Figure 1) presents a pattern of white tiles and grey tiles that has to be generalised. It is a familiar type of question in the English school curriculum - apart from the fact that it immediately asks students to make a 'far generalisation' (Stacey, 1989), rather than making a series of 'near generalisations' (ibid) first. It was devised to see whether students could make the generalisation on the basis of the pattern's structure or whether, on the basis of spurious number patterns, they would resort to a 'function' or 'whole-object scaling' strategy (ibid) that was inappropriate.

Brown, et al. (2002), in their work with student teachers, found that the students quite often chose to "perform computations when reasoning about computations would suffice." Question A4 (Figure 3), was based on a suggestion made by Ruthven (1995), and explores this tendency to perform calculations, in a less familiar setting. It uses factorials (a notation that would not be known by students of this age) and concerns divisibility. The first part of the question can be solved by simple computation while in the final part the dividend is too large for this to be a viable strategy, thus (we hoped) forcing a different, more structural approach.

As well as presenting cross-sectional quantitative analyses of responses to both questions, we have gathered together for this chapter a selection of extracts from the interviews with Yr 10 students that involved discussion of A1 or A4, and during which we probed the reasons behind their responses by asking them to compare and evaluate the responses they had made to the same question on different occasions. Unfortunately these interviews could not be systematic, as students were not necessarily selected for their performance on A1 or A4, but analysis of the relevant extracts does throw light on students' thinking on the items and why progress may or may not have been made.

Generalising A Number Pattern

Question A1 is a standard number/algebra task involving a tile pattern that was familiar to English students. It was designed to test whether students could discern, use and describe a structural relationship between a number of white and grey tiles. A question of this type was included in all the proof tests but took two forms, with the same question used in the proof tests in Yr 8 and Yr 10 (with an additional part in Yr 10), and a different but parallel question used in Yr 9 (not shown here). The Yr 10 version is shown in Figure 1 and consisted of two parts, A1*a* and A1*b*. The students were given one example of the relationship (showing 6 grey tiles and 18 white tiles), and in part *a* asked to generalise this to another number (60) of white tiles and to explain their numerical calculation. In part *b*, the students were asked to write a general relationship involving *n* white tiles. The Yr 8 version was identical to part *a* but consisted only of this part, as most students of that age would not yet have had much experience of algebraic symbolisation. We deliberately built numerical distracters into the item, in the form of simple, but irrelevant, relationships between the numbers of white tiles mentioned in the item (namely, $6 \times \underline{10} = 60$) and between the number of white and grey tiles in the given configuration ($6 \times \underline{3} = 18$).

<<INSERT FIGURE 1>>

We recall that an aim of the project was to map out the different kinds of responses to the proof items and to capture progress in reasoning by an analysis of how the frequencies of the different codes changed over time. To achieve this aim, we coded the responses to A1*a* into 5 broad categories that were based on an *a priori* analysis of possible response-types (using prior research and our own pilot study with over 150 students). Code 1 was given to responses that were incorrect and based solely on spotting number patterns. Code 2 responses showed some recognition of structure but were incomplete or incorrect. Code 3 responses were correct showing a recognition of the structure of the numerical calculation performed. (When we developing the coding scheme, we observed that most students who correctly structured the tile pattern saw the pattern, implicitly or explicitly, as of the form ‘double and add 6’. There are of course other equivalent forms, such as ‘add one at each end, double, add one at each end,’ but these seemed to be rare and so we did not differentiate between them in our coding scheme.) Code 4 was given to responses that included some explicit description of a general relationship between the different coloured tiles and Code 5 was used if this general rule was expressed correctly with variables. Finally, a miscellaneous code was used where students gave no response or responses that did not fit the other codes. Codes 1 to 5 are summarised below.

Code 1:	Spotting number patterns ($6\times \underline{10} = 60$, so there are $18\times \underline{10} = 180$ grey tiles; or $6\times\underline{3} = 18$ so there are $60\times\underline{3} = 180$ grey tiles), no structure
Code 2:	Some recognition of structure (incomplete or draws & counts)
Code 3:	Recognition and use of structure: specific (correct answer, e.g., showing 60+60+3+3)
Code 4:	Recognition and use of structure: general (correct answer and general rule, e.g., $\times$ 2, +6)
Code 5:	Recognition and use of structure: general, with use of variables

(correct answer and general rule, with naming of variables in words or letters, e.g., multiply the number of white tiles by 2 and add 6, or $2n + 6$).

We judged that responses according to these codes were hierarchically ordered in terms of mathematical 'quality,' and thus were of the view that, broadly speaking, as students developed mathematically, i.e., became mathematically more capable and aware, they would tend to give higher quality responses. At the same time, a Code 3 response is sufficient to answer the item successfully and thus students who were capable of giving higher quality responses than Code 3 might not have felt the need to do so.

Pattern Spotting And Structural Reasoning

Our first attempt to capture changes in response patterns over time was to record the frequencies of response to A1*a* classified according to the codes. These frequencies are shown in Table 1.

<<INSERT TABLE 1>>

Table 1 shows that, despite some improvement between Years 8 and 9, a substantial minority of students continued to use 'number pattern spotting' strategies giving an incorrect solution of 180 grey tiles. Altogether 35% of the total sample gave such responses in Yr 8. This fell to 21% in Yr 9 but stayed at 21% in Yr 10. Thus the cross-sectional analyses indicated modest improvement followed by plateau. Of course, the fact that the proportions of students giving number pattern responses were the same in Yrs 9 and 10, does not mean these proportions consisted of exactly the same students. We discuss this further below, when we look at the data longitudinally.

Complementing the changes in frequency of pattern spotting responses, the frequency of correct responses (Codes 3, 4 and 5) went up from 48% in Yr 8 to 68% in Yr 9 but only to 70% in Yr 10. However, this small increase from Yr 9 to Yr 10 masks a substantial rise in the appropriate use of variables (expressed in words or with letters) in students' explanations (Code 5 responses), from 16% in Yr 9 to 26% in Yr 10 (and starting from just 9% in Yr 8).

We now turn to our longitudinal analyses of these data. We focus on changes in patterns of response, according to the codes, between Yr 8 and Yr 10 only, since the first part of A1 was identical in those years but not in Yr 9 (the complete longitudinal analyses for item A1 are available in the Year 10 Technical Report of the project—see Küchemann & Hoyles, 2003, pp 10 - 18). Table 2a shows the code frequencies for A1*a* longitudinally, in that it cross-tabulates individual students' responses in Yr 8 with their responses in Yr 10. However, as our purpose here is to consider 'progress' and as a Code 3 response is sufficient to answer the item successfully, we have grouped the Code 3, 4 and 5 responses. Table 2b is derived from Table 3b and assumes the validity of the hierarchical ordering of Codes 1, 2 and 3-4-5 combined. Based on this assumption, it shows the percentage of students who 'progressed' or 'regressed' in their responses from Yr 8 to Yr 10 (we have ignored all students who gave a miscellaneous non-correct response in either or both years).

<<INSERT TABLE 2.a>>

<<INSERT TABLE 2.b>>

As can be seen from Tables 2a and 2b, the improvement in students' responses from Yr 8 to Yr 10 was not entirely smooth, with 25% of the sample progressing on A1*a* but 7% regressing, giving a 'net progress' of 18%. (This is comparable to most of our other core items, where net progressed ranged from 2% to 32.) Focussing on the pattern spotting responses (Code 1) in particular, Table 3b show that over the two-year period from Yr 8 to Yr 10, well over one-half of those who gave a number pattern response (Code 1) in Yr 8 progressed to a correct (Code 3, 4 or 5) or partially correct (Code 2) structural response in Y10, but that one-quarter of those who gave a number pattern response in Yr 10 had given a higher quality response in Yr 8.

Expressing Structure In Algebra

The Yr 9 and 10 versions of question A1 had an added part, A1*b*, where students were asked to write an expression for the number of grey tiles needed for *n* white tiles (see Figure 1 for the Yr 10 version). We were interested in whether students were able to express any relationship they discerned in the tiling pattern in algebra, and indeed whether this was consistent with their explanations of structure given in words or numbers. We have noted in a previous study (Healy & Hoyles, 2000) that Yr 10 students in England rarely used algebraic symbolisation as a language with which to described mathematical structure, even though they had been taught to do this and indeed they accorded high status to algebraic 'proofs'.

A1*b* asks students to map the number of white tiles onto grey, i.e. it requires a function approach ($n \rightarrow 2n + 6$ for the Yr 10 item). Thus we were curious to see whether this would provoke a rethink from some of the students who had used a number pattern approach, namely those whose approach involved scaling (18 grey tiles × 10 = 180 grey tiles in Yr 10). In the event, in Yr 10 more than half of this subgroup of students (N = 197) switched to a function response, although most stuck to a number pattern approach in part *b*, with 43% giving a response of the form $n \rightarrow 3n$ (which gives 180 when n = 60); however, 13% produced a correct algebraic expression, of the form $n \rightarrow 2n + 6$ (with a further 6% producing a partially correct algebraic expression). (Not surprisingly, of the 92 students who in Yr 10 had given a function number pattern response to A1*a*, ie 60 white tiles × 3 = 180 grey tiles, the vast majority, 87%, gave a response of the form $n \rightarrow 3n$ to part *b*, with only 1% producing a correct algebraic expression.)

It is also worth noting, that of the 615 students who gave a code 3 response to A1*a* in Yr 10 (i.e., a correct, but specific and non-algebraic response), 93% gave a correct algebraic response to part *b*. Thus part *b* was effective in prompting students to use algebra where they had not felt compelled to do so in part *a*, and, as discussed above, it provoked a substantial proportion of students to switch from a scalar to a function approach, even if in most cases, probably, this did not lead to a correct restructuring of their answers.

Some Illustrative Interview Extracts

Our analyses of the quantitative data suggests that, for some students at least, there is an element of chance about their responses: rather than being wedded to a particular way of construing such tile patterns (with some going for structural reasoning and some for number pattern spotting) they seemed to hit upon one way on one occasion and another way on another occasion; furthermore there is not necessarily any consistency between their non-algebraic and algebraic expressions of the relationship. An examination of individual scripts also shows that some students flipped between approaches on a given occasion.

We now turn to student interviews to help us understand the students' interpretation of the question. The interview extracts used here were all with Yr 10 students who were asked to compare the responses they had made to A1 in Yr 8 and in Yr 10 and to explain any inconsistencies and changes. It is worth recalling that all these interviews were with students whose schools were singled out as exceptionally 'good' in developing mathematical reasoning. We have selected to report written responses and interview extracts for three students with different response patterns. First, student MS who persisted in making pattern spotting approaches in both Yr 8 and Yr 10, and then students EC and JG who both used pattern spotting in Yr 8 and showed awareness of structure in Yr 10 and had thus apparently made progress, but who differed in their responses to being asked to explain their reasoning. We were not able to interview any student who clearly appeared to regress.

Student MS gave a pattern spotting response in Yr 8 and Yr 10 (and also in Yr 9). However in Yr 8 he used a '×10' scalar strategy while in Yr 10 he used a '×3' function strategy (together with a function response, '$3n$', that was consistent with this in part *b*).

MS was interviewed a few days after taking the Yr 10 test and was asked about his Yr 8 and Yr 10 responses which were placed on the desk in front of him. At first he seemed to feel the responses were essentially the same and, as this extract indicates, he remained unperturbed when the interviewer read through the responses again and suggested that they were different:

> I: ...Here (Yr 10) you do something quite different by saying…I've got 6 white tiles and if I multiply by 3 I get 18 grey tiles.
>
> MS: ...there's 6 there and 18…altogether times it by 3, then I thought it would be the same if you wanted to find out how many grey tiles would be in 60 so I timesed by 3…
>
> I: Did you, did you think about checking it in any way? Or…
>
> M: I was quite confident on this stuff…I went onto the next question.

Thus MS seemed confident that the numerical relationships he had found were right, perhaps because of their simplicity, and seemed content to ignore the suggestion that they might be different.

We now turn to student EC who gave a '×10' pattern spotting response in Yr 8 (and Yr 9). Initially, he seems to have embarked on a similar response in Yr 10, in that he has written '6 × 10 = 60'. However this is crossed through and replaced by '60 × 2 = 120' and '+3 +3 = 126', indicating a correct answer based on the geometric structure. EC was interviewed about a week after taking the Yr 10 test and asked to compare his Yr 8 and Yr 10 responses. We were surprised that he chose his incorrect Yr 8 response as the one he now believed was correct:

> I: Which one do you think is the right one?

EC: I think this one is [pointing to his Yr 8 response].

I: The Yr 8 one?

EC: Yeah, kind of the first instinct I had.

I: You go by instinct.

EC: Yeah, I think, sort of, in the majority of the time the first instinct is right, so, I think maybe that one looks right.

The interviewer then attempted to probe further and asked EC to explain his Yr 10 response in more detail. EC seemed able to do this but still did not change his mind about the relative merits of his Yr 8 and Yr 10 responses:

I: ...So, you ended up here in Year 10 with this double thing and then add 6.

EC: Yeah.

I: So, how did that come to you, I mean, why would you have done that?

EC: I think, because we needed 60 and there was 6 along each row, each of the white things, so that means 12, so I just thought that doubling it, and then there's 3 left over, so I just plussed 3 on one, so, I'm not really sure.

I: Okay. I mean, that sounds sensible enough, so, the trouble is, we've still got these two different answers. So are you going to stick with your instinct, your Yr 8 instinct?

EC: I think so, yes.

It seems surprising that EC was so ready to abandon his correct structural approach in Yr 10 for the simplicity of his '×10' approach in Yr 8, especially when one notes that his Yr 10 (and Yr 9) response to A1b was also correct. Perhaps at this stage, through lack of experience or guidance, EC does not have the meta knowledge needed to classify his different responses in an appropriate way and to recognise their positive or negative qualities.

We look finally at student JG had a similar set of responses to EC, in that she gave a '×10' number pattern response in Yr 8 and a structural response in Yr 10: '$60 \times 2 = 120$, $3 \times 2 = 6$, $120 + 6 = 126$' (see Figures 2a and 2b). JG was interviewed about a week after taking the Yr 10 test. At first she could make no sense of her Yr 8 response ("I have no idea why I wrote that in Yr 8"), though she comes up with an interpretation eventually:

I: I mean, say someone else had done it, not you…could you sort of try and figure out why on earth they did it?

JG: (Long pause) No.

I: No, you can't see any logic in it?

JG: Well…yeah. I can now. It's because there's 6 there so, I figured 6 times 10 would be the 60 that they were talking about in the question, and so I just had to times the amount around the outside by the same number. Oh yeah.

However, unlike EC, she prefers her structural Yr 10 response, which she feels makes more sense:

I: Can you say a bit more why it makes more sense?

JG: I don't know…just a couple more years' practice of finding patterns and stuff.

I: So how does this answer sort of fit the pattern, the Yr 10 answer fit the pattern better?

JG: Well because, I didn't just times the ones round the outside…by the same number as the ones on the inside…I worked out sort of a rule for it, rather than just a rule for that, that number.

I: Right. How did you get the rule for the Yr 10 answer?

JG: Well, the three at each end won't change, it's a single row of tiles say… you just use the top…the grey tiles above and below the white…

I: Right, okay…

<<INSERT FIGURE 2a & 2b >>

JG's replies here are interesting in several respects. First she justifies her preference for the structural response with an external reason ("more years' practice"), which, though perfectly plausible, has nothing to do with the quality of the actual response and which is certainly no more valid than EC's quest for simplicity. However, she is able to describe the structure itself very nicely ("three at each end... ...grey tiles above and below...") and she does in fact say something, albeit in a rather cryptic way, about the quality of this explanation, namely about it being *general*: thus she found "sort of a rule" in contrast to "just a rule for that ... number." Notice also her statement that "I didn't just times the ones round the outside," which potentially provides a compelling visual test for this number pattern spotting strategy, since the outcome would be a set of grey tiles that no longer fits snugly around the white tiles. Further, she gives a correct response to A1*b,* which shows she is able to express the structure using algebra. All in all, we seem here to be witnessing the beginnings of a meta knowledge about structural reasoning, even though the concepts and language may not yet be well-formed. Unfortunately, we do not know how this knowledge has arisen—although from the limited information we managed to gather about JG's Yr 10 mathematics class, the quest for structural explanations was not a strong feature of its socio-mathematical norms.

Although we can not say how representative these three interviewees are, their responses do suggest that, for some students at least, the simplicity of number pattern responses may have a stronger appeal than the insight that might be gained from taking a structural approach. The analysis suggests possible discontinuities between modelling with numbers, narrative descriptions of these models and modelling with algebra which in turn might lead some students to re-organise their thinking. It also suggests fragility in appreciating the power of a structural approach, and rather limited ability to describe the characteristics of even correct structural reasoning. This is a phenomenon we have found elsewhere and which may well be widespread even amongst the highest of our high attaining students, since they will generally have had little experience of producing mathematical explanations and reflecting upon them.

From Calculating To Structural Reasoning

We found a strong tendency for students to work at an empirical level on our proof test items. In the case of A1 discussed above, working at an empirically was largely manifested by pattern spotting responses, which were given by just over one-third of our high attaining sample in Yr 8 and still by just over one-fifth in Yr 10. This empirical tendency was particularly strong, though manifested rather differently in responses to another question, A4, which we used in the Yr 8 and Yr 9 proof tests (but not in the Yr 10 test, for reasons of time and space). A4 had distinctive characteristics in that unlike A1, the mathematical content would not have been familiar to students and a solution strategy that appealed to structure rather than calculation would not have been taught. The question is shown in abbreviated form in Figure 3 (the original question was spread over an entire A4-size[2] page, with blank space after each part for students to write their responses). The question has three parts, but our interest here is in parts *a* and *c*. In A4*a* students are asked about the divisibility of 5! by 3 and in A4*c* about the divisibility of 100! by 31 (or 50! by 19 in the Yr 9 version).

<<INSERT FIGURE 3>>

Responses to A4*a* were coded into three broad categories, to capture whether students gave incorrect or irrelevant reasons (Code 1), or determined the divisibility of 5!

[2] A4 is comparable in size to US letter, but with the property that an A4 sheet can be folded in half to produce a rectangular shape that is mathematically similar to the original rectangular shape (the edge-lengths are in the ratio 1 to root 2).

by calculating the value of 5! (Code 3), or gave a correct reason based on the 'divisibility principle' (Code 4).

The Codes 1 and 4 were again used in A4*c*, but there was no Code 3 as students did not have the means to calculate the given factorial (100! in Yr 8 or 50! in Yr 9). However, some students (albeit very few) used an inductive reason to argue that the given factorial was divisible by 31 (Yr 8) or 19 (Yr 9), based usually on the observation that in part *a* 120, i.e., 5!, is divisible by 5, 4, 3, 2 and 1. Such responses were given the Code 2.

As with A1, we again judged that the codes were hierarchically ordered in terms of mathematical 'quality' and that as students developed mathematically they would give higher quality responses.

The full code frequencies for A4*a* and A4*c* can be found in Küchemann and Hoyles (2003, pp 23 - 24). Regarding A4*a*, the Code 3 and 4 frequencies show that most students could correctly determine the divisibility of 5! by 3 (76% in Yr 8, 83% in Yr 9) but the overwhelming majority gave a Code 3 rather than Code 4 response (74% of the total sample in Yr 8, 77% in Yr 9), that is, they did so by calculating 5! and then calculating 120 ÷ 3, and thus essentially by multiplying by 3 and dividing by 3 again. These students showed that they understood what was meant by the notions of factorial and divisibility but it seems they could not put them together to form an explanation. (It is of course possible that some did not give such an explanation because they felt that a demonstration was good enough; however the responses to part *c* suggest that this would have been rare.) Only 2% of the sample in Yr 8 and only 6% in Yr 9 gave a Code 4 response, i.e., based their argument on the fact that 3 was a *given* factor of 5!.

While the latter (Code 4) kind of argument is not required in part *a*, it is essential for part *c*, since students were not allowed calculators. However, as mentioned above, only a slightly larger proportion used such an argument in part *c* (3% of the sample in Yr 8, and 9% in Yr 9). Most students wanted to evaluate the factorial, and had no viable alternative strategy. Some students wrote statements like "That would take years to work out and if there is some short cut I don't know it"; some evaluated 100! as being 2400, on the basis that $100 = 20 \times 5$ so $100! = 20 \times 5!$, and so answered 'No'; others answered 'No' because 100! is even and 31 is odd or a prime.

The response below (Figure 4), given by a student in Yr 8, is a typical response to A4*c*. Interestingly this student gave a structural response in Yr 9, answering 'Yes' to "Is 50! divisible by 19?", because 'If you times it by a certain number, you will be able to divide by it'. However, we found that very few students managed to move from empirical to structural reasoning on this unfamiliar question. On A4*a* only 18 students used the divisibility principle in both years, and just 76 students (5% of the sample) shifted from calculating (or a lower level response or no response) to using the divisibility principle, with 19 of the 37 students who had used the divisibility principle successfully in Yr 8 reverting to calculating (or to a lower level response) in Yr 9. On A4*c*, where the divisibility principle is needed to answer the item successfully, the picture is not much better. Here only 22 students (1% of the sample) used the principle consistently, and 114 students (7%) progressed to using it in Yr 9 having not done so in Yr 8, with 20 of the 42 students who answered part *c* successfully in Yr 8, regressing (or omitting the item) in Yr

9. We found this regression, particularly on part *c*, surprising: surely once a student had understood that 'if you multiply by a number then the product is divisible by that number' then this would not be 'forgotten'?

To gain some insight into these issues, we turn again to our interview data. In the course of our visits to 'outlier' schools, we managed to interview 12 students on their Yr 8 and Yr 9 responses to A4. It turned out that none of these were students who had regressed on part *c*, which is unfortunate (though perhaps not surprising given that only 15 students regressed out of our total sample of 1512 students). However several students had successfully progressed from calculating in Yr 8 to using the divisibility principle in Yr 9 and we consider one of these students here.

Student AM was interviewed the day after he had taken the Yr 9 test. His written responses are summarised in Table 3, below, and are shown fully in Figures 5a and 5b. It appears that AM made considerable progress from Yr 8 to Yr 9: he would seem to have a clear understanding of the 'divisibility principle' by Yr 9.

<<INSERT TABLE 3>>

<<INSERT FIGURE 5a & 5b>>

As with the A1 interviews, we asked students to compare their responses in different years and to explain why they had written their particular responses.

In AM's Yr 9 written responses he gives an explanation based on the divisibility principle in part *a* as well as part *c*, rather than one based on evaluating 5!. Thus, in part *a* he had written "The number has been multiplied by 3, so it must be divisible by 3." In the interview, he is asked how he arrived at this explanation:

AM: I think I was just thinking about it. I was thinking 5! would be $5\times4\times3\times2\times 1$ so if it's been timesed by 3 you can almost certainly divide it by 3 and I was just thinking sort of that was that. Also I was saying because it's only then been timesed by 2 and by 1, because 1's obviously not going to change and 2's just going to double it. So you're just going to be able to divide it by 3.

AM does not talk of calculating 5! but his explanation seems grounded in the steps of the potential calculation. AM feels that it is not enough that $\times3$ is one element in a string of terms, and he carefully checks that the subsequent terms, $\times2$ and $\times1$, will not affect the divisibility by 3. This nicely illustrates the fact that a notion like 'divisibility' involves a whole nexus of ideas (see Brown, et al., 2002), including some awareness of the associative and commutative properties of multiplication, especially when considering the divisibility of a long string of terms as in the case of 100! and 50!.

Later, AM says he is "almost certain" but not entirely that 5! is divisible by 3. The interviewer explores this further:

I: You said just now it's almost certainly divisible by 3, you're hesitating slightly.

AM: I think it is divisible by 3, I think at the time I wasn't completely certain.

I: What yesterday, but you're certain today?

AM: Yes, after all that I'm almost certain that's good.

I: Why is there still this edge of doubt? You say *almost* certain.

AM: Don't know. The thing is, I am certain, but not quite... I can't see why, I can see slightly why it works but not entirely, I haven't thought 'suppose you had a bigger number, would it…'

Thus it would seem that for AM, this lack of certainty about the 'divisibility principle' (the argument that if you multiply by a certain number the result is divisible by that number) is not because he does not understand the basic argument or appreciate its power, but because of some awareness that other features of the situation (e.g., that in 100! the term $\times$ 31 is followed by a long string of other terms) might render the principle invalid. From this (and other interviews) it appears that students who correctly answered part *c* of A4, yet still expressed a need to calculate, may be expressing this need not because they reject or do not value their attempts at a structural explanation, but to *check* that the explanation is valid because they are unsure about parts of their argument. This insecurity about using number relationships because of the possible influence of 'other' factors (even when the relationships seem to be understood), might help to explain why some students 'regressed' in Yr 9.

Conclusions

We have reported some findings from a longitudinal study of high attaining students' conceptions of mathematical reasoning. In this study, students' responses to specially designed annual written proof tests were coded and the code frequencies analysed cross sectionally and longitudinally. Total scores on each proof test were subjected to multilevel analyses, which were used in part to identify schools in which students performed significantly better than predicted. Interviews were conducted with students from these schools, who were asked to explain their written responses. Although not systematic, extracts from these interviews serve to throw light on students' thinking on the items and why progress may or may not have been made.

In this chapter we have focussed on two number/algebra questions both of which were designed to investigate patterns in the use of empirical and structural reasoning. One question, A1, involved a familiar task of generalising a tile pattern (although it was unusual in involving a 'far generalisation' only); the other question, A4, was less familiar and concerned factorials and divisibility. We found that the use of structural reasoning increased over the years, albeit at a modest rate. This improvement indicates a general cognitive shift from reliance on empirical reasoning to more theoretical reasoning based on the development of meta knowledge about structures that is doubtless interlinked with the effects of teaching for this high attaining group. However, despite this effect of teaching, the use of empirical reasoning was still widespread in the form of inappropriate number pattern spotting for A1 and through the desire to perform rather than analyse a calculation for A4.

The quantitative analyses of the longitudinal coded data showed a degree of turbulence in student responses, suggesting that for some students at least, there might be an element of chance about their responses: for example rather than being wedded to a particular way of construing a tile pattern, they seemed to hit upon the underlying structure on one occasion and a superficial number pattern on another occasion. the simplicity of number pattern responses appeared to have a continuing appeal often stronger than the insight that might be gained from taking a structural approach. We also noted inconsistencies in how students construed the tile pattern, most notably when asked to write an algebraic expression where their responses did not necessarily match the approaches used in their earlier reasoning, and though some students might well have been perturbed by this mismatch, there was evidence to suggest that others were not.

Our student interviews lent support to the view that students on the whole showed a lack of confidence in and a fragile grasp of structural reasoning. Initially, we interpreted these findings simply as further evidence of students' lack of appreciation of the power of structural reasoning and of the widespread tendency to use empirical methods—something that is perhaps particularly strong in English schools, because of the curriculum and classroom approaches (see for example Morgan, 1997). However, we have modified our views, having reflected on the tentative commitment to different methods exhibited by students in their written responses and interviews. We therefore offer an alternative explanation that distinguishes a more advanced use of empirical reasoning, namely to check the validity of a structural argument.

Jahnke (2005) puts forward the metaphor of 'theoretical physicist' as a way of describing students' behaviour as they learn to engage in mathematical proof. From this viewpoint, students' recourse to empirical evidence can be seen as a perfectly rational and meaningful attempt to test the validity of a proof argument. Indeed, if proof is seen as something undertaken by "fallible mathematicians ... as part of a quasi-empirical process" (Reid, 2005; see also for example: Lakatos, 1976; Lampert, 1990), then this metaphor might usefully be extended beyond the novice student of mathematics. This throws interesting light on an influential study by Fischbein and Kedem (1982), who might be said to have held a "traditional" concept of proof, whereby "a formal proof of a mathematical statement confers on it the attribute of a priori universal validity" (ibid, quoted in Reid, 2005). They found that even where students agreed that a given proof was correct, many endorsed the idea that further numerical checks would increase their confidence in the theorem. This finding was supported by the study of Healy and Hoyles (2000), who similarly used an item where students were presented with different 'proofs' of a statement. They reported that students simultaneously held two different conceptions of proof; those about arguments they considered would receive the best mark and those about arguments they would adopt for themselves. In the latter category, students chose arguments (usually empirical) that they could evaluate and which they found relatively convincing (in the sense that they found them more convincing than did students who chose other arguments) even if they recognised that their scope were limited.

For Fischbein, students who welcomed further numerical checks did not "really understand what a mathematical proof means" (Fischbein, 1982, p.16). Our evidence suggests that though this may be the case for some students, there is an alternative interpretation which may apply to others, namely that they do have some understanding of proof (as, say, a logically ordered set of reasons involving mathematical properties) but they are acknowledging that there might be flaws in the proof (e.g., in the logic or the reasons) which they have not spotted. So in response to our unfamiliar question about factorials and divisibility, even some students who had shown a basic understanding of the underlying structure and given structural reasons in their written response, seemed to need to calculate to achieve full closure. From the perspective of checking the validity of a structural argument, this can be seen as a rational way of coping with a degree of uncertainty about the influence of other features of the situation that might render their reasoning invalid, rather than as a lack of appreciation of such an argument's power. Thus, we suggest that though recourse to empirical data may in many cases indicate a naive understanding of proof, it need not do so.

We end by summarising our findings and briefly pointing to some implications for teaching. High attaining students in our large random sample made progress, albeit modest, in the use of structural reasoning over the three years of the project suggesting a positive and cumulative outcome of teaching. However progress in reasoning from structures was not linear, and was not necessarily retained. Thus despite an overall positive trend, there was unpredictable variation due to issues of interpretation of the task, to changes in curricular emphasis (such as the introduction of algebra), and to an individual student's confidence in their adoption of structural reasoning. We also found that the use of empirical reasoning, in the form of inappropriate number pattern spotting or through the desire to perform rather than analyse a calculation, remained widespread over the three years of the project although we identify a more advanced use of empirical reasoning, namely to check the validity of a structural argument.

Given the fragility of student responses, as our longitudinal analyses showed, we conclude that single snapshots of student understanding can be misleading, as students may not have a clear sense of what it means to progress in regard to the quality of a mathematical argument, or they may be trying to express their mathematical ideas in a new representational infrastructure (e.g., algebra). Also, though we do not discuss it in this chapter, we have observed a seeming u-shaped development, reminiscent of "errors of growth" (Bruner, et al., 1966, p199), where students appear to regress through applying ideas recently met at school when they are inappropriate, or in ways that are not yet effective.

Our analyses suggest that switching strategies (even between incorrect strategies) might be helpful in catalyzing a new perspective on a problem. This indicates that we can change students' habits of mind (Cuoco, et al., 1996). In our recent work on the Proof Materials Project (Küchemann, 2008; see also http://www.ioe.ac.uk/proof/PMPintro.html), we collaborated with teachers to see whether this change of habit could be put into effect more widely, in particular by helping students become more aware of different kinds of proof strategies and explanations. We used the student responses to our items as starting points for discussion among teachers, leading them to think about how to encourage students to use different representation for their ideas (e.g., narrative, algebraic and visual), to make connections between them and to justify their reasoning. This helped students focus on structure rather than just on outcomes, as well as to distinguish mathematical from non-mathematical reasons. It is notable that teaching strategies such as speaking about the relationship, using and comparing different representations, and taking a range of particular cases, tend to be used by teachers in lower sets in English schools and not in higher sets, given the twin fears of not covering the curriculum and of students becoming bored if invited to revisit work from a new perspective. A shift in teaching emphasis in this direction for high attainers would seem a useful way forward.

References

Balacheff, N. (198). Aspects of proof in students' practice of school mathematics. In D. Pimm (Ed.), *Mathematics, teachers and children* (pp. 216-235). London: Hodder and Stoughton.

Bell, A. W. (1976). A study of students' proof-explanations in mathematical situations. *Educational Studies in Mathematics, 7*, 23-40.

Bills, L. & Rowland, T. (1999). Examples, generalisation and proof'. In L. Brown (Ed.), *Making meanings in mathematics: Refereed Proceedings of the British Society for Research into Learning Mathematics* (pp. 103-116). York: QED Publications, York.

Boero, P. (Ed.). (in press). *Theorems in school*. Rotterdam, The Netherlands: Sense Publishers.

Brown, A., Thomas, K., & Tolias, G. (2002). Conceptions of divisibility: success and understanding. In S. R. Campbell & R. Zazkis (Eds.), *Learning and teaching number theory* (pp. 41-82). Stamford, CT: Ablex Publishing.

Bruner, J. S., et al. (1966). *Studies in cognitive growth*. New York: Wiley.

Coe, R., & Ruthven, K. (1994). Proof practices and constructs of advanced mathematics students. *British Educational Research Journal, 20*(1), 41-53.

Cuoco, A., Goldenberg, E. P., & Mark, J. (1996). Habits of mind: An organizing principle for mathematics curricula. *Journal of Mathematical Behavior,* 15, 375-402.

DfEE (Department for Education and Employment). (2001). *Key stage 3 national strategy: Framework for teaching mathematics: Years 7, 8 and 9*. London: DfEE Publications.

Dreyfus, T. (1999). Why Johnny can't prove. *Educational Studies in Mathematics, 38*(1-3), 85-109.

Fischbein, E. (1982). Intuition and proof. *For the Learning of Mathematics, 3*(2), 9-18, 24.

Fischbein, E. & Kedem, I (1982). Proof and certitude in the development of mathematical thinking. In A. Vermandel (Ed.), *Proceedings of the 6th Annual Conference of the International Group for the Psychology of Mathematics Education* (pp. 128-131). Antwerp, Belgium: Universitaire Instelling Antewerpen.

Godino, J. D. & Recio, A. M. (1997). Meaning of proofs in mathematics education. In E. Pehkonen (Ed.), *Proceedings of the 21st Conference of the International Group for the Psychology of Mathematics Education (*Vol. 2, pp. 313-320). Lahti, Finland: Program Committee.

Hanna, G. (1989). Proofs that prove and proofs that explain. In G. Vernaud, J. Rogalski & M. Artigue (Eds.), *Proceedings of the 13th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 45–51). Paris, France: G. R. Didactique, Laboratoire de Psycholgie du Development et de l'Education de l'Enfant.

Harel G. & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. In E. Dubinsky, A. Schoenfeld & J. Kaput (Eds.), *Research in collegiate mathematics education III* (pp. 234-282). Providence, RI: American Mathematical Society.

Healy, L. & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education, 31*(4)*,* 396-428.

Hoyles, C. (1997). The curricular shaping of students' approaches to proof. *For the Learning of Mathematics, 17*(1), 7-15.

Hoyles, C., Küchemann, D., Healy, L. & Yang, M. (2005). Students' developing knowledge in a subject discipline: Insights from combining quantitative and qualitative methods. *International Journal of Social Research Methodology, 8*(3), 225-238.

Jahnke, H. N. (2005). A genetic approach to proof. Paper presented at the *Fourth Congress of the European Society for Research in Mathematics Education*. Sant Feliu de Guixols, Spain.

Küchemann, D. (1989). The effect of setting and numerical content on the difficulty of ratio tasks. In G. Vergnaud, J. Rogalski & M. Artigue (Eds.), *Proceedings of the 13th Annual Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 180-186). Paris, France: G. R. Didactique, Laboratoire de Psycholgie du Development et de l'Education de l'Enfant.

Küchemann, D. & Hoyles, C. (2003). *Technical report for Longitudinal Proof Project: Year 10 survey 2002, Volume 1*. London: Institute of Education (http://www.ioe.ac.uk/proof).

Kücheman, D. (2008). *Looking for structure: a report of the Proof Materials Project*. London: Dexter Graphics (also http://www.ioe.ac.uk/proof/PMPintro.html).

Lakatos, I. (1976). *Proofs and refutations*. Cambridge: Cambridge University Press.

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: mathematical knowing and teaching. *American Educational Research Journal*, *27*(1), 29-63.

Morgan, C. (1997). The institutionalisation of open-ended investigation: some lessons from the UK experience. In E. Pehkonen (Ed.), *Use of open-ended problems in mathematics classroom* (Vol. 176, pp. 49-62). Helsinki: Department of Teacher Education.

Polya, G. (1954). *Mathematics and plausible reasoning, Volume 1: Induction and analogy in mathematics*. London: OUP.

In Stylianou, D.A, Blanton, M. L. & Knuth, E.J. (Eds) *Teaching and Learning Proof Across the Grades K-16 Perspective* Lawrence Erlbaum Associates pp171- 191

Reid, D. (2005). The meaning of proof in mathematics education. Paper presented at the *Fourth Congress of the European Society for Research in Mathematics Education.* Sant Feliu de Guixols, Spain.

Ruthven, K. (1995). The unbearable mightness of being. In L. Healy (Ed.), *Justifying and proving in mathematics: Draft activities*. Unpublished manuscript, London: Institute of Education.

Stacey, K. (1989). Finding and using patterns in linear generalising problems. *Educational Studies in Mathematics, 20*(2), 147-164.

Vergnaud, G. (1983). Multiplicative Structures. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematics concepts and processes* (pp. 128-175). London: Academic Press.

de Villiers, M. (1990). The role and function of proof in mathematics. *Pythagoras, 24,* 17-24