INVESTIGATING FACTORS THAT INFLUENCE STUDENTS' MATHEMATICAL REASONING

Dietmar Küchemann and Celia Hoyles

Institute of Education, University of London

We report on the responses of high attaining groups of fourteen-year-old students to one written algebra item and two written geometry items that formed part of a nationwide survey designed to test mathematical reasoning. Preliminary findings suggest that responses are influenced by topic (algebra or geometry), gender and familiarity, as well as by general mathematical attainment. Additionally responses to familiar (algebra) items appear to be subject more to the influence of textbook than of general mathematical attainment, while responses to unfamiliar (geometry) questions are more subject to variation between classes.

In this paper we report some preliminary findings from a written survey conducted in the first year of a three-year longitudinal study of mathematical reasoning. The aim of this research is to advance understanding of how students learn to reason mathematically by analysing their progress over time, and specifically to identify through large-scale longitudinal study, individual, school and teacher factors that are predictors of secondary school students' competence in mathematical reasoning[1].

A 50 minute survey was administered in June 2000 to 2797 high attaining fourteen-year-old students from 63 randomly selected schools within nine geographical areas that spanned England. The items were iteratively designed and tested over a period of three months. The starting point for the construction of each item was an issue concerned with proving, followed by a trawl of the literature around this issue and a search for relevant tasks in the curriculum. We report here on students' answers to one open-response algebra item (A1), one open-response geometry item (G2a), and one multiple-choice geometry question (G3).

Frequencies for the sample as a whole are given as well as for four groups of students (P1, P2, Q and R) in order to illustrate some trends in the data. Groups P1 (N = 30) and P2 (N = 28) are parallel top mathematics sets from a non-selective suburban school, Q is a group of 25 of the best students selected from four mixed ability classes in a highly selective school, and R (N = 31) is the top set from an urban comprehensive school.

Students were given a Baseline Mathematics Test a few weeks before taking the proof survey, to provide a measure of their general mathematical attainment. This test consisted of 22 multiple-choice items selected from the Third International Mathematics and Science Survey (IEA, 1996). There were no questions on mathematical reasoning. The mean scores on the test, for the total sample and for groups P1, P2, Q and R were 15.3, 16.4, 16.2, 20.0 and 15.2 respectively. Thus the mean scores for P1, P2 and R were roughly similar to the mean for the (high attaining) sample as a whole, while the mean for Q was substantially higher.

Pattern spotting responses to a question about generalising a structure

Question A1 (shown in Figure 1) is concerned with generalisation in a setting (tile patterns) familiar to English students. (There is extensive work on a generalisation perspective to introducing algebra; see Mason, 1996).

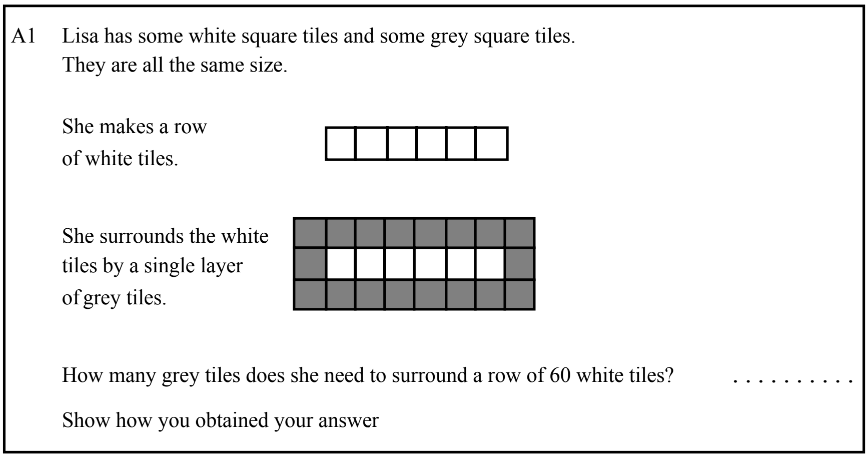


Figure 1: Question A1

As well as providing a numerical answer, students were asked to show how they had obtained their answer. Responses were coded into 5 broad categories (Table 1, below). We discuss just one category here, which we name pattern spotting. In this students obtained the incorrect answer 180 by applying a number pattern without recognition of the structure of the question (see Hoyles and Küchemann, 2000, for a description of the other categories). The diagram in question A1 shows 6 white tiles surrounded by 18 grey tiles. Students were asked for the number of grey tiles needed to surround a row of 60 white tiles. The incorrect answer of 180 grey tiles was found by deriving a (false) relationship, either between the given and required number of white tiles (there are 10 times as many, so there will be 10 times 18 grey tiles) or, less frequently, between the number of white and grey tiles (there are 3 times as many, so there will be 3 times 60 grey tiles).

Code 1	Incorrect answer (180); use of an incorrect number pattern
Code 2	Incorrect answer (eg 120); partial use of correct structure (eg doubles but does not add 6)
Code 3	Correct answer (126); use of correct structure in the specific case of the question with no indication of generality
Code 4	Correct answer (126); use of correct structure indicating its generality
Code 5	Correct answer (126); use of correct structure (expressed in variables)
Code 9	Miscellaneous incorrect answers (including no response)

Table 1: Response codes for question A1

It was not unexpected that some students would resort to using number patterns and thereby attempt to make an empirical, as opposed to a structural, generalisation (Bills and Rowland, 1999). Despite the drawbacks of such an approach (see Hewitt, 1992), the use of number patterns has been widely advocated in UK curriculum materials for many years, even in such stimulating materials as the DIME project (Giles, 1984). However, in such materials students are asked to produce a systematic list of numerical data from which to induce a rule. Here, for reasons as yet unclear, many students were attempting to generalise from a single numerical instance; that is, not only were they ignoring the spatial structure of the situation, but they were not testing their rule on other numerical data.

Figure 2 shows the frequency of the pattern spotting response for question A1. The solid black column shows that over one third of all the students gave this response (total sample, $N = 2797$). This is far larger than we had anticipated, and almost as large as the proportion of students who answered the question correctly (43 %).

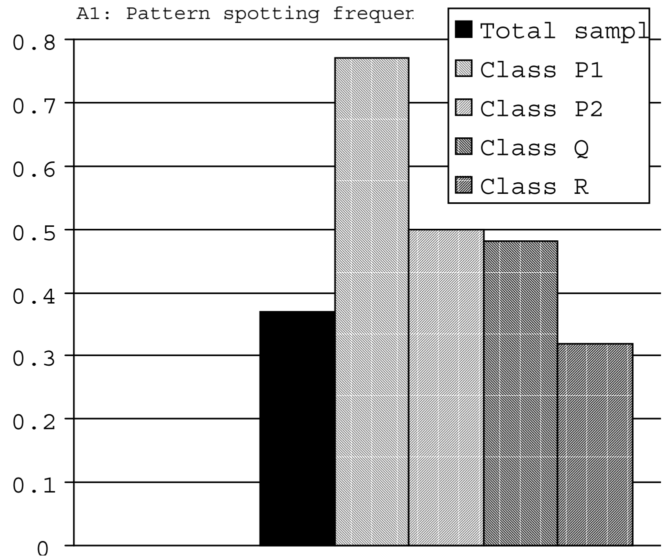


Figure 2: A1 pattern spotting frequencies for total sample and for four groups

Figure 2 also shows the responses for groups P1, P2, Q and R. We draw attention to two noteworthy features. One is the difference between classes P1 and P2. These are parallel classes from the same school, so the finding strongly points to the operation of teacher influences, though at this stage we do not know what these might be. A second interesting feature is the relatively high pattern spotting frequency (48 %) for group Q. The students in this group were selected for their high mathematical attainment within an already selective school. Many of those who gave this pattern spotting response to question A1 gave high level responses to other items on the proof survey (including the geometry item to be discussed below); many also scored highly on the Baseline Maths Test. Again, the reasons for the high frequency of the pattern spotting response for group Q are not yet known, though the textbook used in the school might provide a clue[2].

Responses to an item to distinguish perceptual from logical reasoning

Question G2a (Figure 3) is based on an item by John Gardiner (personal communication). We used it to investigate whether students use perception or logical reasoning in explaining their answers to a simple (but unfamiliar) figural question (see Lehrer and Chazan, 1998; Harel and Sowder, 1998).

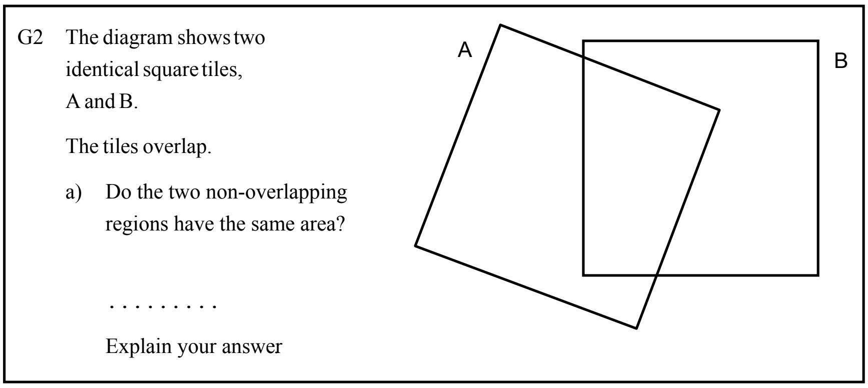


Figure 3: Question G2a

Students are presented

with identical overlapping squares and are asked whether the two non-overlapping regions have the same area, and why.

Responses were coded into 3 broad categories, which are shown below (Table 2)[3]. For code 1 responses, students gave no reason or only mentioned the non-overlapping areas and deemed them to be the same (or not the same) because they looked to be the same (or not). Code 3 responses focussed on the overlap, with an argument along the lines: 'The squares overlap the same amount (and have the same area)[4], so the non-overlapping regions have the same area'. The code 2 responses were similar but with the area of the overlap treated in a specific rather than a general way; for example, students might state that if one third of each square overlaps, then the same amount, two thirds, of each square does not overlap.

Code 1	Correct or incorrect answer (Yes/No); no logical explanation
Code 2	Correct answer (Yes); logical explanation based on specific example
Code 3	Correct answer (Yes); logical argument
Code 9	Miscellaneous incorrect answers (including no response)

Table 2: Response codes for question G2a

Figure 4 shows the frequencies for each code, for the total sample (the solid black columns) and groups P1, P2, Q and R. Perhaps the greatest surprise was that over half the total sample gave a code 1 (perceptual) response. The relatively poor response of class R is also of interest. It matches that class's performance on another geometry item (question G1) where the students made an incorrect response based on perception rather than mobilising a simple geometric argument (see Hoyles and Küchemann, 2000). On the other hand, there is less difference between classes P1 and P2 than on question A1, and the performance of group Q is much as one might expect from their general mathematical attainment, with relatively few code 1 responses and relatively many code 3 responses.

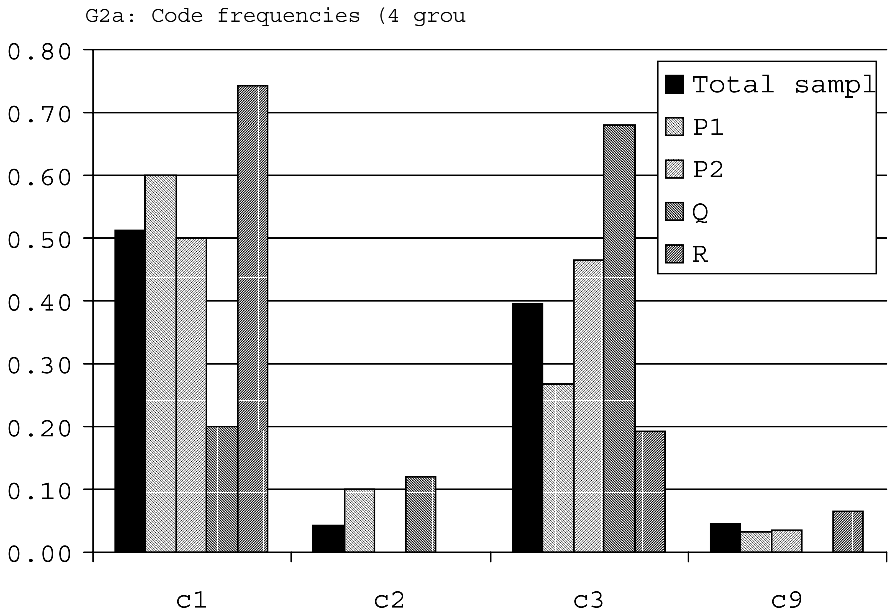


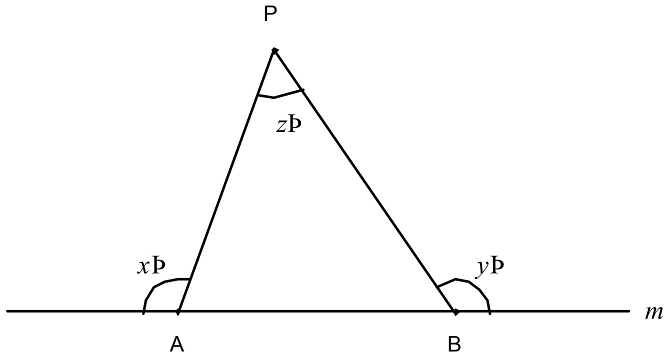
Figure 4: G2a code frequencies for total sample and for four groups

Choices of argument to explain a geometrical conjecture

In question G3 (Figure 5, below), students were presented with a mathematical conjecture and a range of arguments in support of it (options A, B, C and D). They were asked to make two selections from these arguments--the argument that would

G3 In the diagram, A and B are two fixed points on a straight line m.

Point P can move, but stays connected to A and B (the straight lines PA and PB can stretch or shrink).



Avril, Bruno, Chandra and Don are discussing whether this statement is true:

xÞ+ yÞ is equal to 180Þ+ zÞ

Avril's answer

I measured the angles in the diagram and found that angle x is 110Þ, angle y is 125Þ and angle z is 55Þ.

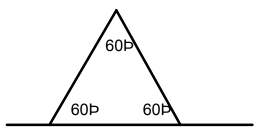
110Þ + 125Þ = 235Þ,

and 180Þ + 55Þ = 235Þ.

So Avril says it's true

Bruno's answer

I can move P so that the triangle is equilateral, and its angles are 60Þ.



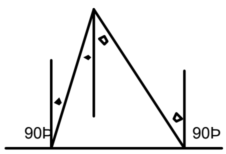
So x is 120Þ and y is 120Þ.

120Þ + 120Þ is the same as 180Þ + 60Þ.

So Bruno says it's true

Chandra's answer

I drew three parallel lines.
The two angles marked with a ◆ are the same and the two marked with a ◇ are the same.



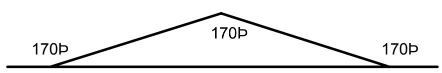
Angle x is 90Þ+◆ and angle y is 90Þ+◇.

So x plus y is 180 + ◆ + ◇, which is 180 + z.

So Chandra says it's true

Don's answer

I thought of a diagram where the angles x, y and z are all 170Þ.



S i omn d y i xa+y ign sro qtt 1 m +8z 0l

So Don says it's not true

a) Whose answer is closest to what you would do? . . . .

b) Whose answer would get the best mark from your teacher? . . . .

Figure 5: Question G3 parts a) and b)

be nearest to their own approach and the argument they believed would receive the best mark from their teacher. The question was deliberately couched in dynamic terms (“Point P can move …”) to invite students to adopt a dynamic approach to the question. Fischbein (1982) suggests that such an approach can be an effective way of accessing generality and of gaining insight, and option C (Chandra's answer) is similar to an approach that he recommends for tackling the angle sum of a triangle. Frant and Rabello (2000) also suggest that a dynamic approach can be useful at an intuitive level and for forming conjectures, (though they seem to argue that a static approach is needed for a formal proof).

Two aspects of the students' responses are of particular interest. One is the marked difference between the choices for ‘own approach’ and for ‘best mark’; the other is the difference in choice between girls and boys, particularly for ‘own approach’.

G3		Own approach					
		A	B	C	D	other	total
Best mark	A	0.08	0.03	0.01	0.00	0.00	0.12
	B	0.09	0.10	0.01	0.01	0.00	0.22
	C	0.18	0.17	0.08	0.05	0.00	0.50
	D	0.03	0.03	0.01	0.04	0.00	0.11
	other	0.01	0.01	0.00	0.00	0.03	0.06
	total	0.40	0.35	0.10	0.11	0.04	1.00

Table 3: G3 - frequencies for 'own approach' and 'best mark' (N = 2797)

Table 3 shows the distribution of choices for the total sample. It indicates that by far the most popular choices for ‘own approach’ were A (40 %) and B (35 %), both of which are empirical arguments, with only 10 percent choosing the general argument, C. On the other hand, 50 percent chose C for ‘best mark’. This response pattern also occurred with a parallel numerical question (A3) and it echoes the findings of Healy and Hoyles (2000) in their survey with 16-year-old students. (Not surprisingly, given the large number of students involved, the difference between the choices for ‘own approach’ and ‘best mark’ was highly significant: $\chi^2 = 1759.5$, df = 16, $p < 0.0001$)

Figure 6 shows the frequency of choices for ‘own approach’ for girls and boys in the total sample. It can be seen that the girls show a clear bias towards choice A with the boys showing a lesser bias towards C and D. These differences were also significant ($\chi^2 = 63.8$, df = 4, $p < 0.0001$). At this stage the reasons for the differences are not clear, but they would seem to be worth investigating further. Similar differences occurred in question A3 and with some other items in the survey.

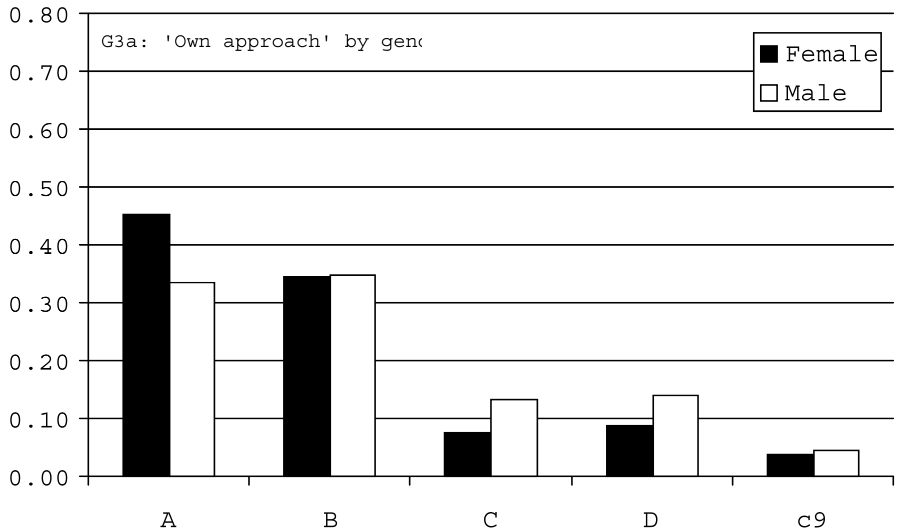


Figure 6: Question G3 – frequency of choices of 'own approach' for girls and boys in total sample

Discussion

These simple statistics suggest that: *first*, type of response to a familiar algebra item about generalisation may be related to general mathematics attainment but may also be influenced by teaching and textbook: *second*, that since geometry is given rather little emphasis as a context for reasoning in the English curriculum, 'better' responses in geometry (that is, introducing a logical as opposed to a perceptual explanation) may be more strongly related to general mathematics attainment than in algebra and also geometry responses may be subject to more variation (as in class R) due perhaps to teacher belief and interest: *third*, that even students who have had rather little introduction to proving have already developed two different conceptions of mathematical reasoning, in that arguments that they assess would receive the best mark differ from arguments they would adopt for themselves: and *fourth*, gender might also be a factor influencing response. Many of these results are similar to those reported following the analysis of a survey conducted in 1998 of older (16-year-old) students' conceptions of proof (see Healy and Hoyles, 2000). These suggestive findings will be investigated further: statistically using multilevel modelling (Goldstein, 1995) and qualitatively through interviews with students and teachers selected on the basis of the profile of individual or class response.

Acknowledgement

We gratefully acknowledge the support of the Economic and Social Research Council (ESRC), Project number R000237777.

Notes

1. Information about the study can be found on the project's website at www.ioe.ac.uk/proof.
2. The Year 8 book devotes several pages to number sequences, and these are presented in a fairly open way; however, the setting is nearly always purely numerical, rather than involving spatial patterns as in A1.
3. In a fourth category, not listed in the table, students obtained the correct answer by measuring, for example by imposing a square grid on the diagram and counting squares. Such responses were given by only 0.5 % of the total sample and by none of the students in the four groups under discussion here.
4. Code 3 responses were subdivided into those that did and did not make explicit reference to the fact that the squares A and B had the same area.

References

Bills, L. and 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*. York: QED Publications.

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

Frant, J and Rabello, M.: 2000, 'Proofs in geometry: different concepts build upon very different cognitive mechanisms.' ICME9, Tokyo, Japan. www-cabri.imag.fr/Preuve/ICME9TG12/ICME9TG12Contributions/BoliteFrantICME00t.html.

Giles, G.: 1984, 'Number patterns 2 - further mappings.' *Dime Pre-Algebra Project*. Norfolk: Tarquin Publications.

Goldstein, H.: 1995, *Multilevel Statistical Models* (2nd ed.). London: Edward Arnold.

Harel, G., & Sowder, L.: (1998), 'Students' proof schemes: Results from exploratory studies.' In A. H. Schoenfeld, J. Kaput, and E. Dubinsky (eds.), *Research in Collegiate Mathematics Education, III* (pp. 234-283). Providence, RI: American Mathematical Society.

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

Hewitt, D.: 1992, 'Train spotters' paradise.' *Mathematics Teaching*, 140, 6-8.

Hoyles, C. and Küchemann, D.: (2000), 'Year 8 students' proof responses: some preliminary findings from four groups.' *Proceedings of the British Society for Research into Learning Mathematics* 20(3), pp. 43-48.

IEA.: 1996, *Third International Mathematics and Science Study*, TIMSS Mathematics Items, Released Set for Population 2 (seventh & eighth grades).

Lehrer, R. & Chazan, D.: 1998, *Designing Learning Environments for Developing Understanding of Geometry and Space*. London: LEA.

Mason, J.: 1966, 'Expressing generality and roots of algebra.' In N. Bednarz, C. Kieran and L. Lee (eds.), *Approaches to Algebra* (pp. 65-86). Dordrecht: Kluwer Academic Publishers.