IDENTIFYING DIFFERENCES IN STUDENTS' EVALUATION OF MATHEMATICAL REASONS

Dietmar Küchemann and Celia Hoyles
Institute of Education, University of London

We report on responses of high attaining thirteen-year-old students to a multiple-choice geometry question (G3) that formed part of a written survey designed to test mathematical reasoning. We describe trends in the choices made and put forward some suggested reasons for these trends.

INTRODUCTION

A 50 minute written survey designed to test mathematical reasoning was administered in June 2000 to 2799 high attaining Year 8 students from 63 randomly selected schools within nine geographical areas across England. All the questions in the survey were developed over a period of three months, the starting point in each case being an issue concerned with proving, followed by a trawl of the literature around this issue and a search for relevant tasks in the curriculum. The survey contained 9 questions in all, of which two were in multiple-choice format, including the question that is the subject of this paper.

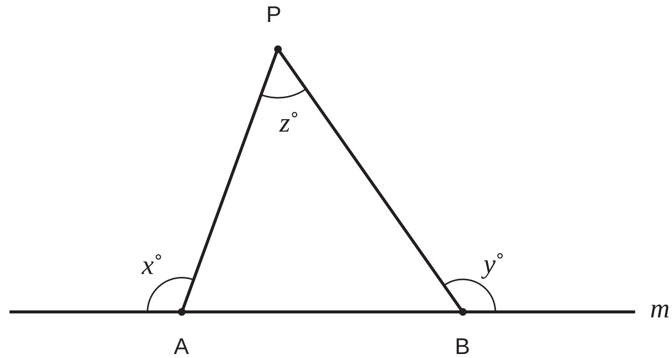
CHOICES OF ARGUMENT TO EXPLAIN A CONJECTURE

In question G3 (Figure 1, 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 be nearest to their own approach and the argument they believed would receive the best mark from their teacher. The question matches the format of the multiple-choice questions used by Healy and Hoyles (2000) in a survey of high attaining Year 10 students, though the current question offers fewer choices. As with the previous research, during pre-piloting the question was given to groups of high-attaining students in an open format during task based interviews with the researchers. We then collected a range of response-types that helped us to devise the choices presented in G3. Also, we made sure we presented pragmatic and more conceptual choices following Balacheff's distinctions (Balacheff, 1988): Avril's answer involves measuring, Bruno's the use of some geometrical knowledge in a specific case, Chandra's involves a general, conceptual argument based on the introduction of parallel lines and knowledge of alternate angles and Don's provides a counter example based on an incorrect diagram.

The question was deliberately couched in dynamic terms ("Point P can move …") to invite students to adopt a dynamic approach. Fischbein (1982) suggests that this 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

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:

$x° + y°$ is equal to $180° + z°$.

Avril's answer

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

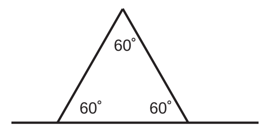
110˚ + 125˚ = 235˚,

and 180˚ + 55˚ = 235˚.

So Avril says it's true

Bruno's answer

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



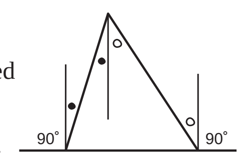
So *x* is 120˚ and *y* is 120˚.

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

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.



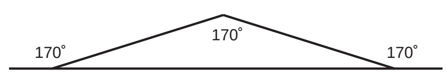
Angle *x* is 90˚+ ● and angle y is 90˚+ ○.

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 170˚.



So in my diagram *x* + *y* is not equal to 180 + *z*.

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 1: Question G3 parts a) and b)

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).

As mentioned above, students were asked to select the choice that was closest to their own approach, and then to select the choice that they thought would get the best mark from their teacher. Table 1 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 pragmatic and particular arguments, with only 10 % choosing the conceptual argument, C. On the other hand, 50 % chose C for best mark, which suggests that many of the students could appreciate the power of the conceptual argument, even if they felt that they would not have been able to devise such an argument themselves. Broadly, of the students who opted for choice A, B or D for own approach, about a quarter (slightly more in the case of D) stuck with it for best mark and about a half abandoned it in favour of the conceptual argument, C. Most of those who chose C for own approach, stuck with it for best mark. A similar response pattern occurred with a parallel algebra question, and with the questions used by Healy and Hoyles (ibid). The difference between the distribution of choices for own approach and for best mark was highly significant: $\chi^2 = 1759.5$, df = 16, $p < 0.0001$, perhaps not surprising given the large number of students involved, but nonetheless supporting the conjecture that students' conceptions of a proof can vary according to the question they are asked.

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 1: G3 - frequencies for own approach and best mark (N = 2799)

The relative frequencies for choice A, which involved measurement, and B, which involved the special case of an equilateral triangle, are also of interest. Overall, choice B was rated more highly than A, since more students chose B (22 %) than A (12 %) for best mark, even though A was more popular than B for own approach (40 % compared to 35 %). Thus whilst students might well resort to measurement as a way of generating data, there is perhaps some awareness that measurement provides an unreliable basis for justifying arguments in geometry, and that in this respect choice B is superior to A. Choice B might also have been preferred because it involves a constructive use of geometric knowledge. However, in one respect at least, A can be regarded as superior to B: the essentially arbitrary position of the point P in the diagram used in A means that the diagram can be treated as a ‘crucial experiment’ (Balacheff, ibid) or as a generic example. This contrasts with the highly symmetrical case in B which might well possess properties that do not always hold.

The mathematics teachers of the students involved in the survey were also presented with question G3, and asked to mark each argument out of 10. These data provided us with the teachers' ranking of the choices A, B, C and D and these are shown in Table 2. As can be seen, there is a very clear

G3		Choice A	B	C	D
Rank	1st	4	7	96	0
	2nd	30	81	0	5
	3rd	61	11	3	7
	4th	4	0	0	85
	other	5	5	5	7
	total	104	104	104	104

Table 2: G3 - frequencies for teacher rankings of choices A, B, C and D (N = 104)

preference for C and a strong rejection of D. The ranking of A and B is not as clear cut, but the general preference for B over A matches that of the students.

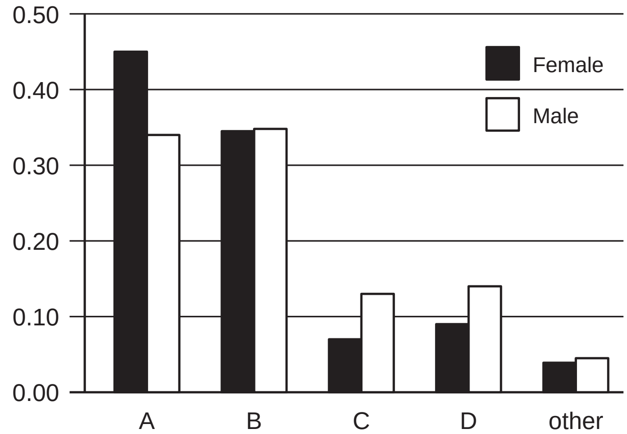


Figure 2: G3 - frequency of choices for 'own approach' for girls (N = 1430) and boys (N = 1369) in total sample

Gender differences in choices

Figure 2 shows the frequency of choices for own approach for girls and for boys in the total sample. Clearly there are gender differences with girls showing a bias towards choice A, boys a (lesser) bias towards C and D. Gender differences in choices were statistically significant for own approach and for best mark, though the latter were less pronounced, especially with regard to choice C. It is possible, therefore, that of the students who appreciate the merits of choice C, more boys than girls are prepared to claim it for their own approach. However, there may well be other reasons for the differences, and they would seem to be worth further investigation, especially as similar differences occurred in the parallel algebra multiple choice question and with some other items in the survey.

HOW THE STUDENTS EVALUATED THE ARGUMENTS PRESENTED

There may be many reasons why students selected an argument for own approach and for best mark. To shed light on this, students were asked to rate the *validity* of each argument and to indicate how well they felt each argument *explained* the truth (or in the case of argument D, the falsehood) of the original statement.

Validity rating

Regarding validity, students were asked whether they agreed, didn't know or disagreed with these two statements, for each of the choices A, B and C:

The answer shows you that the statement is **always true**

The answer **only** shows you that the statement is true for **some** examples.

For choice D, students were presented with this single statement:

The answer shows you that the statement is **not true**.

Students could achieve a total *validity rating* of 7 for the four choices, if they assessed each statement correctly. In the event, the mean validity rating for the total sample was less than 3.

Question G3	Mean total validity rating for students in total sample who chose				
	A	B	C	D	other
	for own approach				
X	2.75	2.54	3.41	2.88	0.87
	for best mark				
Y	2.38	2.35	3.06	2.61	1.36

Table 3: G3 - mean total validity rating of all four choices, for students in the total sample who chose A, B, C or D, for own approach (X) and for best mark (Y) (N = 2799)

Table 3 shows the mean total validity rating for the four choices, for those students who chose A, B, C and D respectively for own approach (row X) and for best mark (row

Y). Those who chose C were generally best able to assess the validity of the choices as a whole, whether one considers those who chose C for own approach (who had a mean total validity rating of 3.41) or those who chose C for best mark (mean = 3.06). This is perhaps not surprising, but it lends support to the supposition that those choosing C tended to do so for sound mathematical reasons rather than just for some surface feature of the explanation. This supposition is further supported if one compares the mean validity rating of each choice, for all the students in the sample, with the ratings given by just those students who selected the particular choice for own approach and for best mark. Students who selected C, whether for own approach or for best mark, achieved a higher validity rating for that choice (mean = 1.02 and 0.80 respectively) than the students as a whole (mean = 0.62); on the other hand, those who selected any of the weaker choices (A, B or D), whether for own approach or best mark, achieved a lower validity rating for their choice than did the sample as a whole (the one exception was the mean for A, own approach, which was the same as for A, total sample). We hope to investigate whether the students who chose C in particular have any common characteristics, for example whether they performed in a distinct way on any of the other proof survey questions. All the students who completed the proof survey had also taken a baseline mathematics test, and we have examined the distribution of baseline scores. This does not suggest that the students who chose C performed any better than the others, so we have to look beyond general mathematical attainment to characterise this group.

Explanatory power

For each choice, students were also presented with the statement, “The answer shows you **why** the statement is true” (or, in the case of D, “not true”). Students were given an *explanatory power* score of 1, 0, or –1 for each choice, depending on whether they agreed, didn't know, or disagreed respectively, with the appropriate statement. It is worth pointing out that, in contrast to the validity rating, this score is subjective, ie it indicates how highly the students rated the explanatory power of an argument, rather than how well their rating might match that of an experienced mathematician. The mean explanation score for the total sample of students was highest for choice C. At the same time, students who selected a particular choice, be it for own approach or for best mark, on average rated the explanatory power of that choice more highly than did the sample of students as a whole. These two findings would seem to convey rather different messages. On the one hand, it is encouraging that the conceptual argument is the one that is rated most highly; on the other hand, it is disappointing that students are perhaps not as objective about the merits of the choices as the differences between the own approach and best mark frequencies (Table 1) had lead us to believe.

DISCUSSION

These simple statistics suggest that there are substantial differences in student choices of argument that most closely matches their own approach and for best mark. A pattern in choices can be identified, with students showing a clear preference for pragmatic approaches for own approach and for a more conceptual approach for best mark. Similarly, trends can be identified where for example students who chose an empirical argument for their own approach switched to a conceptual argument for best mark, while students who chose the latter for own approach stayed with the same choice for best mark. These findings were however subject to gender differences where boys were more likely than girls to claim the conceptual argument for their own approach. Similar findings can be found in algebra, but with rather more uniformity in response than in geometry.

There is evidence that students chose the more conceptual argument for best mark for sound mathematical reasons, although they might also have been influenced by surface presentation. It is worth pointing out that, despite the obvious drawbacks of using a multiple choice format, it allowed us to gather evidence about the more conceptual argument which we would not otherwise have obtained (it was clear from the interviews conducted during the development of question G3, that few of our Year 8 students would have generated such an argument themselves). Students who chose the more conceptual argument were also generally better able to assess the validity of all the arguments although these students were not necessarily better in general mathematics attainment.

Finally, students generally rated the explanatory power of their choice of argument (whether for own approach or for best mark) more highly than did the other students, which suggests that understanding as much as generality was important to students. Interestingly, students also rated the explanatory power of the conceptual argument more highly than the other arguments, which is a cause of some surprise and optimism.

ACKNOWLEDGEMENT

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

REFERENCES

Balacheff, N.: 1988, 'Aspects of proof in pupils' practice of school mathematics.' In D. Pimm (ed.), *Mathematics, teachers and children* (pp. 216-235). London: Hodder & Stoughton.

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.

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