Technical Report
for

Longitudinal Proof Project

Year 8 Survey 2000

D Küchemann & C Hoyles

ACKNOWLEDGEMENTS

We acknowledge the financial support of ESRC who funded this research project, Grant Number R000237777. We would like to thank the teachers in our design schools who assisted us in design and piloting. Finally we would like to express our thanks to all the schools and teachers who agreed to participate in this three year longitudinal study.

In this technical report we present the design of the Year 8 survey and the data collected. The aims were, first, to identify patterns of mathematical reasoning in high-attaining Year 8 students, as a precursor to studying how these patterns might evolve in the subsequent two years; and, second, to identify schools which seemed to be highly successful in developing students' mathematical reasoning, with a view to examining these schools more closely in order to identify factors that might have contributed to their success.

CONTENTS

1. AIMS

The general aim of the project is to advance understanding of how students learn to reason mathematically by studying their progress over time.

The more specific aims of the project are to:-
a. identify, through large-scale longitudinal study, individual, school and teacher factors which are predictors of secondary school students' competencies in mathematical reasoning;
b. identify and describe different school and classroom approaches that are effective in promoting the development of mathematical reasoning;
c. investigate how individual students progress in mathematical reasoning and trace the factors that shape this progression;
d. contribute to guidance for the teaching profession on ways to promote the development of mathematical reasoning.

2. THE DATA

Sample and administration

Four questionnaires were developed: a Baseline Mathematics Test, a student Proof Survey, a Teacher Questionnaire and a School Questionnaire. The full questionnaires are shown in Appendix A, B, C and D respectively.

Students were given the Year 8 Proof Survey in June 2000, at which time their mathematics teachers completed the Teacher Questionnaire and the head of mathematics completed the School Questionnaire. Students had been given the Baseline Mathematics Test a few weeks earlier, in May 2000.

Altogether, 3083 Year 8 students participated in the study, of whom 2799 took the Proof Survey and 2663 took both the Baseline Mathematics Test *and* the Proof Survey. The students came from a total of 63 schools spread across England.

Most schools in the project have a single top mathematics set in Year 10, and where this is the case, our intention was to test just this set when our Year 8 students reached Year 10. The Year 8 students were therefore selected in such a way that as many as possible of those who were likely to end up in the top set in Year 10 were chosen. Normally, whole classes were selected, unless students were taught mathematics in mixed attainment groups. So, for example, if students in a particular school were taught in two parallel mathematics classes in Year 8 but a single top set in Year 10, both Y8 classes were selected. On the other hand, if students were in mixed attainment groups in Y8 but there was a single top set in Y10, we asked the school to select the 30 or so Y8 students who were thought most likely to end up in the Y10 top set.

Six of our 63 schools have mixed attainment maths groups in Y8. A single Y8 class was selected in 25 of the remaining schools and 2 classes in the other 32 schools.

Selection of sample schools

Schools invited to participate in the project were chosen at random from within certain preselected regions. The selection process was started by dividing England into 9 geographical regions, covering south-, central- and north-England and east-, central- and west-England. An LEA was then chosen within each region (or two LEAs, in the case of one urban conurbation), making sure that between them the LEAs covered rural, semi-rural and urban areas. A mathematics inspector or advisor in each LEA was then contacted, each of whom agreed to support the project and whose endorsement was included in the initial contact letter sent to schools. For each LEA, we then compiled a randomly ordered list of all the medium or large 11 or 12 to 16 or 18 secondary schools (using the DFEE 1999 school performance tables, we excluded schools with 120 or less 15 year olds, on the basis that the Proof Survey might well not be suitable for many of the Year 8 students in their top set or sets). As this is a longitudinal study we wished to avoid approaching schools that were too unsettled to be relied upon to stay with the project (for example, schools without a head of mathematics). We sought guidance from the advisors on this and consequently eliminated a total of 9 schools from our lists. We then wrote to the first 8 schools in each LEA list, inviting them to take part in the project. (In the case of the two LEAs in the same conurbation, we wrote to the first 4 schools in each list; also, where an authority had, say, 11-16 and 11-18 schools, we made sure that the schools we approached reflected the actual proportion of such schools.)

Where a school declined our invitation, we wrote to the next school in that LEA's list. Initially, we wrote to 72 schools, of which 53 said yes, 18 said no and 5 never replied, despite reminders. In all we wrote to 23 further schools, of which 11 said yes, 6 said no and 5 did not reply. Of the 65 schools that agreed to take part, one fell out in June, after having received the scripts, and in another the students had such difficulties with the questions in the Proof Survey that we felt it was not fair to ask them to participate further. This left us with 63 sample schools.

Further information about the sample schools can be found in Section 8.

3. BASELINE MATHEMATICS TEST

In a previous project conducted by Hoyles and Healy in 1996 (see Healy & Hoyles, 2000), students' Y9 Key Stage 3 scores were used as an indicator of general mathematics attainment. Such an indicator is important if one is trying to identify other factors, be they at the student, class or school level, which might influence students' mathematical reasoning. It was felt that it would not be appropriate to use equivalent scores for the Year 8 students, as they were derived from tests taken 2 years previously. We therefore developed a test to provide us with comparable information, the final version of which is shown in Appendix A.

The Baseline Mathematics Test was constructed from items used and released by the Third International Mathematics and Science Study (TIMSS, see IEA, 1996). This satisfied our wish for multiple choice items, for ease of marking. Also, the items had been thoroughly tested and we had facility and discrimination data for Y8/9 students in England. We also had 'East Atlantic English' versions of the items.

Our aim was to devise a 30 minute test and we were advised to have about 20 items, mostly in the 30 % - 60 % facility range, with short wording, no tricks, not just involving definitions and with good discrimination. Also, they should span all the TIMSS areas (Fraction & Number, Algebra, Measurement, Geometry, Data & Probability, Proportion), but with an emphasis on number as we were advised that these items tended to be more searching.

Our first draft had 20 items, most in the 30 % - 60 % facility range but with one item at 20 % and three just below 70 %, and all with a discrimination of at least 0.40. The draft was trialled with groups of Y8 students from each of our five design schools (we received back 139 completed scripts). The timing seemed nearly right, with high attaining students finishing the test in about 20 minutes, but we felt we should try to improve the test's discrimination at the top end (the marks of students from an upper band class in one of the design schools were all in the range 10 to 19). We did this by adding two more difficult items (facility 26 % and 36 %), though one of these had a discrimination of only 0.38 (ie just below our target of 0.40). We also replaced a middle facility item that did not seem to be working very well, with an item of comparable facility. We changed the time allowed to 25 minutes. The revised, 22 item test, was trialled with another upper band class from the same school. As hoped for, the resulting marks had a (slightly) greater range (11 - 22), and there was also a more even spread within the range. This revised test became the final version of the Baseline Mathematics Test (see Appendix A). The table below shows the distribution of scores for all the Y8 students who took the test and the Proof Survey (N = 2663; mean = 15.3, SD = 3.79).

score	frequency	% of total
1	1	0.0
2	2	0.1
3	2	0.1
4	11	0.4
5	15	0.6
6	25	0.9
7	30	1.1
8	48	1.8
9	64	2.4
10	93	3.5
11	147	5.5
12	165	6.2
13	221	8.3
14	210	7.9
15	236	8.9
16	289	10.9
17	260	9.8
18	269	10.1
19	226	8.5

20	160	6.0
21	131	4.9
22	58	2.2
Grand Total	2663	100

Table 3.1: Y8 students' scores on Baseline Mathematics Test

4. STUDENT PROOF SURVEY

The Y8 Proof Survey is a 50 minute written test contain 4 arithmetic/algebra questions, 4 geometry questions and one logic question. To avoid giving priority to the algebra or the geometry questions, two versions, A and B, of the test were produced containing identical questions in different orders. The final version of the survey is shown in Appendix B.

One of the algebra (A3) and one of the geometry (G3) questions has a multiple-choice format similar to the multiple choice questions used in the previous project: students are presented with a range of answers regarding the truth of a given conjecture, and are asked to choose which answer is nearest to the one that they would give, and which would receive the best mark from their teacher. They are then asked to rate the validity and explanatory power of each answer.

The survey questions were developed by interviewing individuals or small groups of Year 8 students in the five design schools (and one other school) and by then compiling the questions into a survey and trailing this with groups or classes of Y8 students. In all, 33 students were interviewed and successive drafts of the written survey were given to 6 groups or classes, including one group (of 72 students) from outside London.

Rationale for each question

The following categories were considered for devising items for the proof survey. Items were drafted for each category, although in the end we were not able to devise items that were suitable for (or would fit into) the survey for every category.

Making conjectures Many papers suggest that conjecturing should be part of the proof process, in contrast to how Euclid used to be presented in a polished, sterile, unengaging way [eg de Villiers, Barbin, Hersh, Hanna, Haddas&Herskovitz].

Turning conjectures into (conditional) statements [This is discussed as an important transitional stage in Boero et al, 1999.]

Making generalisations This is important, in particular contrasting structural generalisations (eg Bills & Rowland) with purely numerical ones (eg Coe & Ruthven's observation that students home in on numbers and forsake the situation that generated them; Hewitt, 'Train spotters' paradise'].

Using generic examples [Is discussed for example by Van Dormolen, Balacheff, Pimm&Mason, Tall, Rowland.]

Crucial experiment [This process is discussed by Balacheff.]

General cases which are then limited In part this grows out of dynamic geometry considerations, in part this might be a way of 'unfreezing' or relaxing the relationships that occur especially in

geometry, so that one can impose an order, ‘if this, then that’.
Given a statement, find (deduce) the value of an unknown or derive another statement This allows one to test deduction in a more concrete way, rather than proving from properties [see Paul Clifford]
Truth statements and general logic [see work of Almeida; Clifford; Watson: converse, counter example; Werner Walsch: false assumption, reverse reasoning; Flag tasks, Mars99; inductive reasoning: Bell, Van Dormolen, Balacheff.]
Using definitions This is clearly an important part of proving, especially for older students [see Alcock].
Transformational reasoning [eg Fischbein, Simon, Harel & Sowder; also action proof, eg Bills & Rowland, Semadeni, Walter; notion of proof by induction; reasoning about increments.]
Specialisation after a proof [eg Fischbein, Vinner, Porteous, Healy and Hoyles.]
Scrambled proof
Perceptual proof [Harel & Sowder; contrast with transformational reasoning.]
Multiple choice items [own choice v best mark; narrative v formal; empirical v crucial v generic v thought experiment; ordered v muddled; Proof recognition as opposed to proof construction.]

The final Y8 proof survey contained one or more items for most but not all categories, as can be seen from the list below.

Making conjectures	-
Turning conjectures into conditional statements	-
Making generalisations	A1, A2
Using generic examples	A2, A4
Crucial experiment	L1d
General cases which are then limited	-
Given a statement, find (deduce) the value of an unknown, or derive another statement	A3, G2a, G2b, G3, G4
Truth statements and general logic	L1, G1
Using definitions	A4
Transformational reasoning	G1, G2b, G3
Specialisation after a proof	L1b
Scrambled proof	G4
Perceptual proof	G1, G2a

Coding scheme and scoring scheme

For the open response questions in the proof survey, where students were asked for an answer and an explanation or justification, a coding scheme was devised for students' responses which had the general format shown in the table below. Often codes were divided into two or more subcodes (eg code 31, code 32), depending on the variety of student responses. Where a code ended in 0 (eg code 30), this indicated that it had *not* been subdivided. For items involving simpler responses, for example just a yes/no response, an attempt was made to match the general format as closely as possible.

code 1	Correct or incorrect decision; no valid justification
code 2	Correct or incorrect decision; incomplete or flawed justification
code 3	Correct decision; valid and adequate justification (eg at a specific level)
code 4	Correct decision; valid, higher level justification (eg at a general level)
code 9	Miscellaneous wrong answers

In general, code 3 was used for adequate correct answers, and code 4 (and sometimes also code 5) was used for responses that were at a higher level than was needed to answer the item correctly. Usually, code 4 (and code 5) answers were expressed in more general terms than code 3 answers.

The codes used for each item are shown in the coding scheme in Appendix E.

For the multilevel modelling, scores were used rather than codes. For most items, it was possible to score 0, 1, 2 or 3, and usually a score of 3 was given to any answer that was adequate (ie any code 3 answer). It was decided that answers that were better than adequate (ie code 4 and above) should also score 3, rather than more than 3, as the items did not require students to give such answers.

The scores used for each item are also shown in the coding scheme in Appendix E.

5. TEACHER AND SCHOOL QUESTIONNAIRES

A Teacher Questionnaire was given to all the mathematics teachers of the students in the Year 8 sample and a School Questionnaire was given to the head of mathematics in each school. The questionnaires were developed in consultation with the teachers in our design schools and with the mathematics advisors in the 10 LEAs. Many of the factors that were included in the corresponding questionnaires in the previous project did not show any significant influence on students' patterns of reasoning. However, it was felt that the same factors should be included in the current questionnaires to see whether these findings would be replicated. The Teacher Questionnaire asks teachers about their age, teaching experience and qualifications, about their continuing professional development, their involvement in extra-curricular mathematics activities and their use of computer software in teaching. Teachers are also asked to evaluate the choices in the two multiple-choice questions in the students' Proof Survey. The School Questionnaire asks for basic information about the school, about mathematics teaching in Year 8 and about Y8 students' involvement in extra curricular mathematics activities. We have supplemented this with certain information available from the DfES school performance tables.

The questionnaires can be seen in Appendix C and D, and findings are shown in Sections 7 and 8.

6. DATA FROM THE STUDENT PROOF SURVEY

Descriptive Statistics

The tables below give the following information for each item in the Proof Survey:
the frequency and mean score on the Baseline Mathematics Test for all the girls in the sample, all the boys in the sample, and for the total sample (N = 2663) for
• each specific response code
• each broad response category
• each score.
The meaning of the codes and the scores assigned to the codes are shown on the coding sheet in Appendix E.

Where appropriate, frequency tables are also given for particular response-types such as "use of tables" in Y8A1 and "use of diagram" in Y8G2b.

Note: Percentages are rounded to nearest whole number; mean baseline scores are rounded to 1 DP.

Frequencies For Item Y8A1

Y8A1		Girls	Boys	All
11	Number	15	16	31
	%	1	1	1
12	Number	350	343	693
	%	26	26	26
12E	Number	7	6	13
	%	1	0	0
13	Number	137	87	224
	%	10	7	8
13E	Number	2	1	3
	%	0	0	0
14	Number	7	5	12
	%	1	0	0
21	Number	63	63	126
	%	5	5	5
22	Number	25	26	51
	%	2	2	2
22E	Number	5	3	8
	%	0	0	0
23	Number	12	14	26
	%	1	1	1
23T	Number	1	1	2
	%	0	0	0
24	Number	1	0	1
	%	0	0	0
30	Number	400	408	808
	%	29	31	30
30E	Number	10	15	25
	%	1	1	1
30T	Number	5	5	10
	%	0	0	0
41	Number	5	9	14
	%	0	1	1
41E	Number	0	0	0
	%	0	0	0
41T	Number	1	0	1
	%	0	0	0
42	Number	41	31	72
	%	3	2	3
42E	Number	1	2	3
	%	0	0	0
42T	Number	3	2	5
	%	0	0	0
50	Number	30	42	72
	%	2	3	3
50E	Number	2	1	3
	%	0	0	0
50T	Number	2	0	2
	%	0	0	0
50L	Number	31	58	89
	%	2	4	3
50LE	Number	1	2	3
	%	0	0	0
50LT	Number	10	9	19
	%	1	1	1
50LTE	Number	2	0	2
	%	0	0	0
91	Number	1	4	5
	%	0	0	0
93	Number	190	150	340
	%	14	12	13
Total	Number	1360	1303	2663

Table 6A1.1 Y8A1 code frequencies (number and percent)

Y8A1		Girls	Boys	All
11	Number	15	16	31
	Mean Baseline Score	12.1	14.1	13.2
12	Number	350	343	693
	Mean Baseline Score	14.5	15.3	14.9
12E	Number	7	6	13
	Mean Baseline Score	12.3	13.8	13.0
13	Number	137	87	224
	Mean Baseline Score	14.6	14.8	14.7
13E	Number	2	1	3
	Mean Baseline Score	11.0	14.0	12.0
14	Number	7	5	12
	Mean Baseline Score	16.7	15.6	16.3
21	Number	63	63	126
	Mean Baseline Score	15.6	16.3	15.9
22	Number	25	26	51
	Mean Baseline Score	13.9	14.2	14.0
22E	Number	5	3	8
	Mean Baseline Score	14.4	13.7	14.1
23	Number	12	14	26
	Mean Baseline Score	14.2	16.5	15.4
23T	Number	1	1	2
	Mean Baseline Score	13.0	12.0	12.5
24	Number	1	0	1
	Mean Baseline Score	17.0	x	17.0
30	Number	400	408	808
	Mean Baseline Score	15.7	16.5	16.1
30E	Number	10	15	25
	Mean Baseline Score	16.0	16.7	16.4
30T	Number	5	5	10
	Mean Baseline Score	14.8	18.0	16.4
41	Number	5	9	14
	Mean Baseline Score	15.8	16.7	16.4
41E	Number	0	0	0
	Mean Baseline Score	x	x	x
41T	Number	1	0	1
	Mean Baseline Score	10.0	x	10.0
42	Number	41	31	72
	Mean Baseline Score	16.2	16.6	16.4
42E	Number	1	2	3
	Mean Baseline Score	18.0	19.0	18.7
42T	Number	3	2	5
	Mean Baseline Score	21.3	21.5	21.4
50	Number	30	42	72
	Mean Baseline Score	16.6	17.7	17.2
50E	Number	2	1	3
	Mean Baseline Score	14.5	21.0	16.7
50T	Number	2	0	2
	Mean Baseline Score	17.5	x	17.5
50L	Number	31	58	89
	Mean Baseline Score	19.8	18.2	18.8
50LE	Number	1	2	3
	Mean Baseline Score	20.0	19.0	19.3
50LT	Number	10	9	19
	Mean Baseline Score	16.3	17.2	16.7
50LTE	Number	2	0	2
	Mean Baseline Score	17.5	x	17.5
91	Number	1	4	5
	Mean Baseline Score	6.0	10.0	9.2
93	Number	190	150	340
	Mean Baseline Score	12.4	13.5	12.9
All	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6A1.2 Y8A1 mean baseline score per code

Y8A1 T		Girls	Boys	All
30T	Number	5	5	10
	Mean Baseline Score	14.8	18.0	16.4
41T	Number	1	0	1
	Mean Baseline Score	10.0	x	10.0
42T	Number	3	2	5
	Mean Baseline Score	21.3	21.5	21.4
50T	Number	2	0	2
	Mean Baseline Score	17.5	x	17.5
50LT	Number	10	9	19
	Mean Baseline Score	16.3	17.2	16.7
50LTE	Number	2	0	2
	Mean Baseline Score	17.5	x	17.5
Total	Number	23	16	39
	Mean Baseline Score	16.6	18.0	17.2

Table 6A1.3 Y8A1 use of table (correct)

Y8A1		Girls	Boys	All
Pattern spotting (code 1)	Number	518	458	976
	%	38	35	37
Partial recognition of structure (code 2)	Number	107	107	214
	%	8	8	8
Correct, specific (code 3)	Number	415	428	843
	%	31	33	32
Correct, general (code 4)	Number	51	44	95
	%	4	3	4
Correct, name variable (code 5)	Number	78	112	190
	%	6	9	7
Other incorrect (code 9)	Number	191	154	345
	%	14	12	13
Total	Number	1360	1303	2663

Table 6A1.4 Y8A1 broad-code frequencies

Y8A1		Girls	Boys	All
Pattern spotting (code 1)	Number	518	458	976
	%	38	35	37
	Mean Baseline Score	14.5	15.2	14.8
Partial recognition of structure (code 2)	Number	107	107	214
	%	8	8	8
	Mean Baseline Score	15.0	15.7	15.3
Correct, specific (code 3)	Number	415	428	843
	%	31	33	32
	Mean Baseline Score	15.7	16.6	16.1
Correct, general (code 4)	Number	51	44	95
	%	4	3	4
	Mean Baseline Score	16.4	16.9	16.6
Correct, name variable (code 5)	Number	78	112	190
	%	6	9	7
	Mean Baseline Score	17.9	18.0	17.9
Other incorrect (code 9)	Number	191	154	345
	%	14	12	13
	Mean Baseline Score	12.4	13.4	12.8
Total		1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6A1.5 Y8A1 broad-code frequencies/baseline

Y8A1		Girls	Boys	All
0 (code 1) Pattern spotting; incorrect answer	Number	518	458	976
	%	38	35	37
1 (code 21) Some correct data; incorrect/no answer	Number	63	63	126
	%	5	5	5
2 (code 22,23,24) Correct deduction from incorrect structure, or correct answer with no/incorrect reason	Number	44	44	88
	%	3	3	3
3 (code 3,4,5) Correct answer and correct structure (specific or general)	Number	544	584	1128
	%	40	45	42
0 (code 91,92,93) Other incorrect answers	Number	191	154	345
	%	14	12	13
Total	Number	1360	1303	2663
	Mean Total Score	1.31	1.46	1.38

Table 6A1.6 Y8A1 scores

Y8A1		Girls	Boys	All
0 (code 1) Pattern spotting; incorrect answer	Number	518	458	976
	%	38	35	37
	Mean Baseline Score	14.5	15.2	14.8
1 (code 21) Some correct data; incorrect/no answer	Number	63	63	126
	%	5	5	5
	Mean Baseline Score	15.6	16.3	15.9
2 (code 22,23,24) Correct deduction from incorrect structure, or correct answer with no/incorrect reason	Number	44	44	88
	%	3	3	3
	Mean Baseline Score	14.1	14.8	14.5
3 (code 3,4,5) Correct answer and correct structure (specific or general)	Number	544	584	1128
	%	40	45	42
	Mean Baseline Score	16.1	16.9	16.5
0 (code 91,92,93) Other incorrect answers	Number	191	154	345
	%	14	12	13
	Mean Baseline Score	12.4	13.4	12.8
Total		1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3
	Mean Total Score	1.31	1.46	1.38

Table 6A1.7 Y8A1 scores/baseline

Frequencies For Item Y8A2

Y8A2		Girls	Boys	All
10	Number	112	106	218
	%	8	8	8
10E	Number	7	9	16
	%	1	1	1
21	Number	53	44	97
	%	4	3	4
22	Number	25	36	61
	%	2	3	2
22E	Number	42	48	90
	%	3	4	3
23	Number	6	19	25
	%	0	1	1
23E	Number	0	2	2
	%	0	0	0
31	Number	182	185	367
	%	13	14	14
31AE	Number	10	7	17
	%	1	1	1
31A	Number	7	10	17
	%	1	1	1
31E	Number	52	32	84
	%	4	2	3
32	Number	310	269	579
	%	23	21	22
32E	Number	88	66	154
	%	6	5	6
40	Number	45	40	85
	%	3	3	3
40E	Number	17	9	26
	%	1	1	1
50	Number	17	13	30
	%	1	1	1
50E	Number	1	6	7
	%	0	0	0
50LE	Number	1	3	4
	%	0	0	0
50L	Number	11	12	23
	%	1	1	1
91	Number	150	154	304
	%	11	12	11
92	Number	6	8	14
	%	0	1	1
93	Number	218	225	443
	%	16	17	17
Total		1360	1303	2663

Table 6A2.1 Y8A2 code frequencies (number and percent)

Y8A2		Girls	Boys	All
10	Number	112	106	218
	Mean Baseline Score	13.9	14.2	14.1
10E	Number	7	9	16
	Mean Baseline Score	12.9	13.8	13.4
21	Number	53	44	97
	Mean Baseline Score	13.2	13.9	13.5
22	Number	25	36	61
	Mean Baseline Score	15.6	16.0	15.9
22E	Number	42	48	90
	Mean Baseline Score	13.0	14.8	14.0
23	Number	6	19	25
	Mean Baseline Score	13.3	16.1	15.4
23E	Number	0	2	2
	Mean Baseline Score	x	19.5	19.5
31	Number	182	185	367
	Mean Baseline Score	15.6	16.4	16.0
31AE	Number	10	7	17
	Mean Baseline Score	12.7	14.4	13.4
31A	Number	7	10	17
	Mean Baseline Score	17.9	15.6	16.5
31E	Number	52	32	84
	Mean Baseline Score	14.0	15.7	14.6
32	Number	310	269	579
	Mean Baseline Score	16.5	17.8	17.1
32E	Number	88	66	154
	Mean Baseline Score	14.9	16.7	15.7
40	Number	45	40	85
	Mean Baseline Score	17.8	18.3	18.0
40E	Number	17	9	26
	Mean Baseline Score	16.2	17.3	16.6
50	Number	17	13	30
	Mean Baseline Score	17.6	19.0	18.2
50E	Number	1	6	7
	Mean Baseline Score	16.0	14.8	15.0
50LE	Number	1	3	4
	Mean Baseline Score	16.0	20.0	19.0
50L	Number	11	12	23
	Mean Baseline Score	19.6	20.6	20.1
91	Number	150	154	304
	Mean Baseline Score	13.2	14.3	13.8
92	Number	6	8	14
	Mean Baseline Score	12.3	14.3	13.4
93	Number	218	225	443
	Mean Baseline Score	13.2	13.9	13.6
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6A2.2 Y8A2 mean baseline score per code

Y8A2		Girls	Boys	All
Spotting number patterns (code 1)	Number	119	115	234
	%	9	9	9
Recognition of structure but method ignored (code 2)	Number	126	149	275
	%	9	11	10
Recognition of structure and method used (code 3)	Number	649	569	1218
	%	48	44	46
Structure made explicit and method used (code 4)	Number	62	49	111
	%	5	4	4
Structure made explicit using variable names; method used (code 5)	Number	30	34	64
	%	2	3	2
Other incorrect (code 9)	Number	374	387	761
	%	28	30	29
Total	Number	1360	1303	2663

Table 6A2.3 Y8A2 broad-code frequencies

Y8A2		Girls	Boys	All
Spotting number patterns (code 1)	Number	119	115	234
	%	9	9	9
	Mean Baseline Score	13.8	14.2	14.0
Recognition of structure but method ignored (code 2)	Number	126	149	275
	%	9	11	10
	Mean Baseline Score	13.6	15.1	14.4
Recognition of structure and method used (code 3)	Number	649	569	1218
	%	48	44	46
	Mean Baseline Score	15.8	17.0	16.4
Structure made explicit and method used (code 4)	Number	62	49	111
	%	5	4	4
	Mean Baseline Score	17.4	18.1	17.7
Structure made explicit using variable names; method used (code 5)	Number	30	34	64
	%	2	3	2
	Mean Baseline Score	18.3	18.9	18.6
Other incorrect (code 9)	Number	374	387	761
	%	28	30	29
	Mean Baseline Score	13.2	14.1	13.7
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6A2.4 Y8A2 broad-code frequencies/baseline

Y8A2score		Girls	Boys	All
0 (code 10) Pattern spotting; incorrect answer	Number	119	115	234
	%	9	9	9
1 (code 21) Some recognition of structure; method ignored; no answer	Number	53	44	97
	%	4	3	4
2 (code 22,23) Some recognition of structure; method ignored; correct answer	Number	73	105	178
	%	5	8	7
2.5 (code 31) Recognition of structure; imprecise use of method; partially correct answer	Number	251	234	485
	%	18	18	18
3 (code 32,40,50) Recognition of structure; correct use of method; correct answer	Number	490	418	908
	%	36	32	34
0 (code 91,92,93) Other incorrect answers	Number	374	387	761
	%	28	30	29
Total	Number	1360	1303	2663
	Mean Total Score	1.69	1.61	1.65

Table 6A2.5 Y8A2 scores

Y8A2score		Girls	Boys	All
0 (code 10) Pattern spotting; incorrect answer	Number	119	115	234
	%	9	9	9
	Mean Baseline Score	13.8	14.2	14.0
1 (code 21) Some recognition of structure; method ignored; no answer	Number	53	44	97
	%	4	3	4
	Mean Baseline Score	13.2	13.9	13.5
2 (code 22,23) Some recognition of structure; method ignored; correct answer	Number	73	105	178
	%	5	8	7
	Mean Baseline Score	13.9	15.6	14.9
2.5 (code 31) Recognition of structure; imprecise use of method; partially correct answer	Number	251	234	485
	%	18	18	18
	Mean Baseline Score	15.2	16.2	15.7
3 (code 32,40,50) Recognition of structure; correct use of method; correct answer	Number	490	418	908
	%	36	32	34
	Mean Baseline Score	16.5	17.8	17.1
0 (code 91,92,93) Other incorrect answers	Number	374	387	761
	%	28	30	29
	Mean Baseline Score	13.2	14.1	13.7
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3
	Mean Total Score	1.69	1.61	1.65

Table 6A2.6 Y8A2 scores/baseline

Frequencies For Item Y8A3

Y8A3a (own approach)		Girls	Boys	All
Choice A	Number	537	377	914
	%	39	29	34
Choice B	Number	313	388	701
	%	23	30	26
Choice C	Number	374	367	741
	%	28	28	28
Choice D	Number	94	120	214
	%	7	9	8
c9	Number	42	51	93
	%	3	4	3
Total	Number	1360	1303	2663

Table 6A3.1 Y8A3a choice frequencies (number and percent)

Y8A3a (own approach)		Girls	Boys	All
Choice A	Number	537	377	914
	Mean Baseline Score	15.1	16.3	15.6
Choice B	Number	313	388	701
	Mean Baseline Score	15.4	16.0	15.8
Choice C	Number	374	367	741
	Mean Baseline Score	14.7	15.7	15.2
Choice D	Number	94	120	214
	Mean Baseline Score	13.3	14.2	13.8
c9	Number	42	51	93
	Mean Baseline Score	13.1	14.0	13.6
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6A3.2 Y8A3a mean baseline score per choice

Y8A3a (own approach)		Girls	Boys	All
Choice A	Number	537	377	914
	%	39	29	34
	Mean Baseline Score	15.1	16.3	15.6
Choice B	Number	313	388	701
	%	23	30	26
	Mean Baseline Score	15.4	16.0	15.8
Choice C	Number	374	367	741
	%	28	28	28
	Mean Baseline Score	14.7	15.7	15.2
Choice D	Number	94	120	214
	%	7	9	8
	Mean Baseline Score	13.3	14.2	13.8
c9	Number	42	51	93
	%	3	4	3
	Mean Baseline Score	13.1	14.0	13.6
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6A3.3 Y8A3a choice frequencies/baseline

Y8A3b (best mark)		Girls	Boys	All
Choice A	Number	268	220	488
	%	20	17	18
Choice B	Number	694	667	1361
	%	51	51	51
Choice C	Number	287	308	595
	%	21	24	22
Choice D	Number	35	50	85
	%	3	4	3
c9	Number	76	58	134
	%	6	4	5
Total	Number	1360	1303	2663

Table 6A3.4 Y8A3b choice frequencies (number and percent)

Y8A3b (best mark)		Girls	Boys	All
Choice A	Number	268	220	488
	Mean Baseline Score	14.4	15.5	14.9
Choice B	Number	694	667	1361
	Mean Baseline Score	15.2	16.1	15.6
Choice C	Number	287	308	595
	Mean Baseline Score	15.0	15.8	15.4
Choice D	Number	35	50	85
	Mean Baseline Score	12.9	14.7	14.0
c9	Number	76	58	134
	Mean Baseline Score	13.8	14.2	14.0
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6A3.5 Y8A3b mean baseline score per choice

Y8A3b (best mark)		Girls	Boys	All
Choice A	Number	268	220	488
	%	20	17	18
	Mean Baseline Score	14.4	15.5	14.9
Choice B	Number	694	667	1361
	%	51	51	51
	Mean Baseline Score	15.2	16.1	15.6
Choice C	Number	287	308	595
	%	21	24	22
	Mean Baseline Score	15.0	15.8	15.4
Choice D	Number	35	50	85
	%	3	4	3
	Mean Baseline Score	12.9	14.7	14.0
c9	Number	76	58	134
	%	6	4	5
	Mean Baseline Score	13.8	14.2	14.0
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6A3.6 Y8A3b choice frequencies/baseline

	Y8A3a/A3b	A3a: Own approach					
		Choice A	Choice B	Choice C	Choice D	other	Total
A3b: Best mark	Choice A	10	3	3	2	0	18
	Choice B	18	19	10	4	0	51
	Choice C	5	3	13	1	0	22
	Choice D	1	1	0	1	0	3
	other	1	1	1	0	3	5
	Total	34	26	28	8	3	100

Table 6A3.7 Y8A3a/A3b choice frequencies (percent)

Y8A3a		Own	Best
Choice A	Number	17	19
Choice B	Number	77	70
Choice C	Number	0	1
Choice D	Number	0	0
c9	Number	10	14
Total	Number	104	104

Table 6A3.8 Y8A3 teachers' choice frequencies (number)

Y8A3		Choice			
		A	B	C	D
Rank	1st	7	97	1	0
	2nd	93	3	2	3
	3rd	1	1	10	94
	4th	0	0	88	4
	other	3	3	3	3
	Total	104	104	104	104

Table 6A3.9 Y8A3 teacher rankings of choices: frequencies (number)

Y8A3	Students who selected					
	Choice A for own approach N = 914	Choice B for own approach N = 701	Choice C for own approach N = 741	Choice D for own approach N = 214	other N = 93	Total N = 2663
Mean validity rating of choice A	0.72	1.03	1.11	0.62	0.25	0.88
Mean validity rating of choice B	0.84	1.21	0.46	0.62	0.19	0.79
Mean validity rating of choice C	0.26	0.31	0.06	0.25	0.02	0.21
Mean validity rating of choice D	1.13	1.14	1.15	0.68	0.32	1.07
Total mean validity rating	2.94	3.68	2.78	2.16	0.78	2.96

Note: Maximum possible validity rating for choice A, B, C, D is 2, 2, 1, 2 respectively

Table 6A3.10 Y8A3 mean validity rating of each choice, for students who chose A, B, C or D for own approach

Y8A3	Students who selected					
	Choice A for best mark N = 488	Choice B for best mark N = 1361	Choice C for best mark N = 595	Choice D for best mark N = 85	other N = 134	Total N = 2663
Mean validity rating of choice A	0.72	0.88	1.15	0.71	0.47	0.88
Mean validity rating of choice B	0.64	1.07	0.42	0.58	0.33	0.79
Mean validity rating of choice C	0.20	0.29	0.06	0.21	0.07	0.21
Mean validity rating of choice D	1.02	1.10	1.25	0.58	0.53	1.07
Total mean validity rating	2.58	3.34	2.87	2.07	1.40	2.96

Note: Maximum possible validity rating for choice A, B, C, D is 2, 2, 1, 2 respectively

Table 6A3.11 Y8A3 mean validity rating of each choice, for students who chose A, B, C or D for best mark

Y8A3 Mean explanatory power ([mean of N(agree) - mean of N(disagree)]	Students who selected					
	Choice A for own approach N = 914	Choice B for own approach N = 701	Choice C for own approach N = 741	Choice D for own approach N = 214	other N = 93	Total N = 2663
Mean explanatory power of A	-0.08	-0.23	-0.16	0.20	0.00	-0.12
Mean explanatory power of B	0.61	0.74	0.32	0.37	0.11	0.53
Mean explanatory power of C	-0.05	-0.13	0.45	0.15	0.01	0.09
Mean explanatory power of D	-0.30	-0.37	-0.32	0.09	0.03	-0.28

Table 6A3.12 Y8A3 mean explanatory power of each choice, for students who chose A, B, C or D for own approach

Y8A3 Mean explanatory power ([mean of N(agree) - mean of N(disagree)]	Students who selected					
	Choice A for best mark N = 488	Choice B for best mark N = 1361	Choice C for best mark N = 595	Choice D for best mark N = 85	other N = 134	Total N = 2663
Mean explanatory power of A	0.11	-0.22	-0.13	0.31	-0.01	-0.12
Mean explanatory power of B	0.49	0.71	0.25	0.16	0.25	0.53
Mean explanatory power of C	0.10	-0.11	0.50	0.27	0.05	0.09
Mean explanatory power of D	-0.20	-0.34	-0.34	0.13	-0.04	-0.28

Table 6A3.13 Y8A3 mean explanatory power of each choice, for students who chose A, B, C or D for best mark

Frequencies For Item Y8A4a

Y8A4a		Girls	Boys	All
10	Number	124	121	245
	%	9	9	9
30	Number	879	833	1712
	%	65	64	64
30E	Number	83	56	139
	%	6	4	5
41	Number	25	10	35
	%	2	1	1
41E	Number	0	1	1
	%	0	0	0
41N	Number	1	1	2
	%	0	0	0
42	Number	7	8	15
	%	1	1	1
42N	Number	1	3	4
	%	0	0	0
91	Number	60	79	139
	%	4	6	5
92	Number	2	6	8
	%	0	0	0
93	Number	178	185	363
	%	13	14	14
Total	Number	1360	1303	2663

Table 6A4a.1 Y8A4a code frequencies (number and percent)

Y8A4a		Girls	Boys	All
10	Number	124	121	245
	Mean Baseline Score	12.2	13.8	13.0
30	Number	879	833	1712
	Mean Baseline Score	15.5	16.5	16.0
30E	Number	83	56	139
	Mean Baseline Score	14.7	15.2	14.9
41	Number	25	10	35
	Mean Baseline Score	18.4	20.0	18.9
41E	Number	0	1	1
	Mean Baseline Score	x	17.0	17.0
41N	Number	1	1	2
	Mean Baseline Score	17.0	15.0	16.0
42	Number	7	8	15
	Mean Baseline Score	19.9	20.9	20.4
42N	Number	1	3	4
	Mean Baseline Score	11.0	18.0	16.3
91	Number	60	79	139
	Mean Baseline Score	12.7	14.1	13.5
92	Number	2	6	8
	Mean Baseline Score	17.5	14.8	15.5
93	Number	178	185	363
	Mean Baseline Score	13.6	13.9	13.8
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6A4a.2 Y8A4a mean baseline score per code

Y8A4a		Girls	Boys	All
Ignores ! sign (code 1)	Number	124	121	245
	%	9	9	9
Uses definition and calculates (code 3)	Number	962	889	1851
	%	71	68	70
Uses definition and understands divisibility (calculation not used) (code 4)	Number	34	23	57
	%	3	2	2
Other incorrect (code 9)	Number	240	270	510
	%	18	21	19
Total	Number	1360	1303	2663

Table 6A4a.3 Y8A4a broad-code frequencies

Y8A4a		Girls	Boys	All
Ignores ! sign (code 1)	Number	124	121	245
	%	9	9	9
	Mean Baseline Score	12.2	13.8	13.0
Uses definition and calculates (code 3)	Number	962	889	1851
	%	71	68	70
	Mean Baseline Score	15.4	16.5	15.9
Uses definition and understands divisibility (calculation not used) (code 4)	Number	34	23	57
	%	3	2	2
	Mean Baseline Score	18.5	19.7	19.0
Other incorrect (code 9)	Number	240	270	510
	%	18	21	19
	Mean Baseline Score	13.4	14.0	13.7
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6A4a.4 Y8A4a broad-code frequencies/baseline

Y8A4a		Girls	Boys	All
0 (code 10) Ignores ! sign	Number	124	121	245
	%	9	9	9
3 (code 30,41,42) Uses definition and calculates or understands divisibility	Number	996	912	1908
	%	73	70	72
0 (code 91,92,93) Other incorrect	Number	240	270	510
	%	18	21	19
Total	Number	1360	1303	2663
	Mean Total Score	2.2	2.1	2.1

Table 6A4a.5 Y8A4a scores

Y8A4a		Girls	Boys	All
0 (code 10) Ignores ! sign	Number	124	121	245
	%	9	9	9
	Mean Baseline Score	12.2	13.8	13.0
3 (code 30,41,42) Uses definition and calculates or understands divisibility	Number	996	912	1908
	%	73	70	72
	Mean Baseline Score	15.5	16.5	16.0
0 (code 91,92,93) Other incorrect	Number	240	270	510
	%	18	21	19
	Mean Baseline Score	13.4	14.0	13.7
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3
	Mean Total Score	2.2	2.1	2.1

Table 6A4a.6 Y8A4a scores/baseline

Frequencies For Item Y8A4b

Y8A4b		Girls	Boys	All
11	Number	23	22	45
	%	2	2	2
12	Number	1	3	4
	%	0	0	0
31	Number	369	405	774
	%	27	31	29
31N	Number	45	30	75
	%	3	2	3
32	Number	532	483	1015
	%	39	37	38
32N	Number	62	57	119
	%	5	4	4
32Z	Number	17	14	31
	%	1	1	1
32NZ	Number	5	6	11
	%	0	0	0
32F	Number	131	72	203
	%	10	6	8
32FZ	Number	0	1	1
	%	0	0	0
41	Number	2	3	5
	%	0	0	0
41N	Number	9	7	16
	%	1	1	1
41L	Number	0	3	3
	%	0	0	0
42	Number	4	0	4
	%	0	0	0
42N	Number	6	3	9
	%	0	0	0
42L	Number	0	1	1
	%	0	0	0
91	Number	64	89	153
	%	5	7	6
92	Number	5	8	13
	%	0	1	0
93	Number	85	96	181
	%	6	7	7
Total	Number	1360	1303	2663

Table 6A4b.1 Y8A4b code frequencies (number and percent)

Y8A4b		Girls	Boys	All
11	Number	23	22	45
	Mean Baseline Score	13.3	14.9	14.0
12	Number	1	3	4
	Mean Baseline Score	12.0	11.3	11.5
31	Number	369	405	774
	Mean Baseline Score	14.6	16.0	15.3
31N	Number	45	30	75
	Mean Baseline Score	16.2	16.4	16.3
32	Number	532	483	1015
	Mean Baseline Score	15.1	16.4	15.7
32N	Number	62	57	119
	Mean Baseline Score	15.7	15.9	15.8
32Z	Number	17	14	31
	Mean Baseline Score	14.9	15.4	15.1
32NZ	Number	5	6	11
	Mean Baseline Score	17.2	15.3	16.2
32F	Number	131	72	203
	Mean Baseline Score	15.4	15.8	15.5
32FZ	Number	0	1	1
	Mean Baseline Score	x	12.0	12.0
41	Number	2	3	5
	Mean Baseline Score	13.0	18.0	16.0
41N	Number	9	7	16
	Mean Baseline Score	17.2	16.7	17.0
41L	Number	0	3	3
	Mean Baseline Score	x	19.3	19.3
42	Number	4	0	4
	Mean Baseline Score	18.3	x	18.3
42N	Number	6	3	9
	Mean Baseline Score	18.8	19.3	19.0
42L	Number	0	1	1
	Mean Baseline Score	x	20.0	20.0
91	Number	64	89	153
	Mean Baseline Score	13.1	14.1	13.7
92	Number	5	8	13
	Mean Baseline Score	16.0	15.8	15.8
93	Number	85	96	181
	Mean Baseline Score	13.2	13.2	13.2
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6A4b.2 Y8A4b mean baseline score per code

Y8A4b		Girls	Boys	All
Partial statement only of definition (code 1)	Number	24	25	49
	%	2	2	2
Correct statement of definition (code 3)	Number	1161	1068	2229
	%	85	82	84
Correct statement of definition; structure made explicit (code 4)	Number	21	17	38
	%	2	1	1
Other incorrect (code 9)	Number	154	193	347
	%	11	15	13
Total	Number	1360	1303	2663

Table 6A4b.3 Y8A4b broad-code frequencies

Y8A4b		Girls	Boys	All
Partial statement only of definition (code 1)	Number	24	25	49
	%	2	2	2
	Mean Baseline Score	13.2	14.4	13.8
Correct statement of definition (code 3)	Number	1161	1068	2229
	%	85	82	84
	Mean Baseline Score	15.0	16.1	15.6
Correct statement of definition; structure made explicit (code 4)	Number	21	17	38
	%	2	1	1
	Mean Baseline Score	17.5	18.1	17.7
Other incorrect (code 9)	Number	154	193	347
	%	11	15	13
	Mean Baseline Score	13.2	13.7	13.5
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6A4b.4 Y8A4b broad-code frequencies/baseline

Y8A4b		Girls	Boys	All
0 (code 11,12) Partial statement only of definition	Number	24	25	49
	%	2	2	2
0.5 (code 31) Correct statement of definition but not completely explicit	Number	414	435	849
	%	30	33	32
1 (code 32,41,42) Correct and explicit statement of definition	Number	768	650	1418
	%	56	50	53
0 (code 91,92,93) Other incorrect answers	Number	154	193	347
	%	11	15	13
Total	Number	1360	1303	2663
	Mean Total Score	0.72	0.67	0.69

Table 6A4b.5 Y8A4b scores

Y8A4b		Girls	Boys	All
0 (code 11,12) Partial statement only of definition	Number	24	25	49
	%	2	2	2
	Mean Baseline Score	13.2	14.4	13.8
0.5 (code 31) Correct statement of definition but not completely explicit	Number	414	435	849
	%	30	33	32
	Mean Baseline Score	14.7	16.0	15.4
1 (code 32,41,42) Correct and explicit statement of definition	Number	768	650	1418
	%	56	50	53
	Mean Baseline Score	15.3	16.3	15.7
0 (code 91,92,93) Other incorrect answers	Number	154	193	347
	%	11	15	13
	Mean Baseline Score	13.2	13.7	13.5
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3
	Mean Total Score	0.72	0.67	0.69

Table 6A4b.7 Y8A4b scores/baseline

Frequencies For Item Y8A4c

Y8A4c		Girls	Boys	All
11	Number	195	169	364
	%	14	13	14
12	Number	31	48	79
	%	2	4	3
13	Number	349	389	738
	%	26	30	28
20	Number	4	2	6
	%	0	0	0
40	Number	34	28	62
	%	3	2	2
40N	Number	5	4	9
	%	0	0	0
40L	Number	0	0	0
	%	0	0	0
91	Number	430	347	777
	%	32	27	29
92	Number	22	17	39
	%	2	1	1
93	Number	290	299	589
	%	21	23	22
Total	Number	1360	1303	2663

Table 6A4c.1 Y8A4c code frequencies (number and percent)

Y8A4c		Girls	Boys	All
11	Number	195	169	364
	Mean Baseline Score	13.4	14.4	13.9
12	Number	31	48	79
	Mean Baseline Score	16.0	16.6	16.3
13	Number	349	389	738
	Mean Baseline Score	15.2	15.9	15.6
20	Number	4	2	6
	Mean Baseline Score	15.0	18.5	16.2
40	Number	34	28	62
	Mean Baseline Score	17.9	19.8	18.7
40N	Number	5	4	9
	Mean Baseline Score	16.0	17.3	16.6
40L	Number	0	0	0
	Mean Baseline Score	x	x	x
91	Number	430	347	777
	Mean Baseline Score	14.5	15.7	15.0
92	Number	22	17	39
	Mean Baseline Score	17.2	16.5	16.9
93	Number	290	299	589
	Mean Baseline Score	15.3	15.8	15.6
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6A4c.2 Y8A4c mean baseline score per code

Y8A4c		Girls	Boys	All
Incorrect or irrelevant reason or "can't explain" (code 1)	Number	575	606	1181
	%	42	47	44
Reasons by (scientific) induction (code 2)	Number	4	2	6
	%	0	0	0
Uses definition and understands divisibility (code 4)	Number	39	32	71
	%	3	2	3
Other incorrect (code 9)	Number	742	663	1405
	%	55	51	53
Total	Number	1360	1303	2663

Table 6A4c.3 Y8A4c broad-code frequencies

Y8A4c		Girls	Boys	All
Incorrect or irrelevant reason or "can't explain" (code 1)	Number	575	606	1181
	%	42	47	44
	Mean Baseline Score	14.6	15.6	15.1
Reasons by (scientific) induction (code 2)	Number	4	2	6
	%	0	0	0
	Mean Baseline Score	15.0	18.5	16.2
Uses definition and understands divisibility (code 4)	Number	39	32	71
	%	3	2	3
	Mean Baseline Score	17.7	19.4	18.5
Other incorrect (code 9)	Number	742	663	1405
	%	55	51	53
	Mean Baseline Score	14.9	15.8	15.3
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6A4c.4 Y8A4c broad-code frequencies/baseline

Y8A4c		Girls	Boys	All
0 (code 11,12,13) Incorrect or irrelevant reason or "can't explain"	Number	575	606	1181
	%	42	47	44
2 (code 20) Reasons by (scientific) induction	Number	4	2	6
	%	0	0	0
3 (code 40) Uses definition and understands divisibility	Number	39	32	71
	%	3	2	3
0 (code 91,92,93) Other incorrect answers	Number	742	663	1405
	%	55	51	53
Total	Number	1360	1303	2663
	Mean Total Score	0.09	0.08	0.08

Table 6A4c.5 Y8A4c scores (same as codes, table 3)

Y8A4c		Girls	Boys	All
0 (code 11,12,13) Incorrect or irrelevant reason or "can't explain"	Number	575	606	1181
	%	42	47	44
	Mean Baseline Score	14.6	15.6	15.1
2 (code 20) Reasons by (scientific) induction	Number	4	2	6
	%	0	0	0
	Mean Baseline Score	15.0	18.5	16.2
3 (code 40) Uses definition and understands divisibility	Number	39	32	71
	%	3	2	3
	Mean Baseline Score	17.7	19.4	18.5
0 (code 91,92,93) Other incorrect answers	Number	742	663	1405
	%	55	51	53
	Mean Baseline Score	14.9	15.8	15.3
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3
	Mean Total Score	0.09	0.08	0.08

Table 6A4c.6 Y8A4c scores/baseline (same as codes, table 4)

Frequencies For Item Y8G1

Y8G1		Girls	Boys	All
11	Number	670	496	1166
	%	49	38	44
12	Number	63	100	163
	%	5	8	6
21	Number	68	61	129
	%	5	5	5
22	Number	30	40	70
	%	2	3	3
31	Number	262	283	545
	%	19	22	20
32N	Number	28	44	72
	%	2	3	3
32D	Number	170	204	374
	%	13	16	14
41N	Number	7	7	14
	%	1	1	1
41D	Number	3	9	12
	%	0	1	0
42N	Number	14	14	28
	%	1	1	1
42D	Number	16	13	29
	%	1	1	1
91	Number	8	16	24
	%	1	1	1
92	Number	0	0	0
	%	0	0	0
93	Number	21	16	37
	%	2	1	1
Total	Number	1360	1303	2663

Table 6G1.1 Y8G1 code frequencies (number and percent)

Y8G1		Girls	Boys	All
11	Number	670	496	1166
	Mean Baseline Score	13.7	14.4	14.0
12	Number	63	100	163
	Mean Baseline Score	13.2	15.5	14.6
21	Number	68	61	129
	Mean Baseline Score	15.4	15.4	15.4
22	Number	30	40	70
	Mean Baseline Score	14.2	15.7	15.0
31	Number	262	283	545
	Mean Baseline Score	16.5	16.7	16.6
32N	Number	28	44	72
	Mean Baseline Score	16.1	17.7	17.0
32D	Number	170	204	374
	Mean Baseline Score	16.8	17.4	17.1
41N	Number	7	7	14
	Mean Baseline Score	15.4	16.9	16.1
41D	Number	3	9	12
	Mean Baseline Score	17.0	17.6	17.4
42N	Number	14	14	28
	Mean Baseline Score	16.9	16.9	16.9
42D	Number	16	13	29
	Mean Baseline Score	19.9	18.5	19.3
91	Number	8	16	24
	Mean Baseline Score	13.0	14.6	14.1
92	Number	0	0	0
	Mean Baseline Score	x	x	x
93	Number	21	16	37
	Mean Baseline Score	14.8	15.2	14.9
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G1.2 Y8G1 mean baseline score per code

Y8G1		Girls	Boys	All
21	Number	68	61	129
Correct, weak explanation (no diagram)	%	5	5	5
22	Number	30	40	70
Correct, weak diagram	%	2	3	3
31	Number	262	283	545
Correct, diagram showing clear counter example	%	19	22	20
32N	Number	28	44	72
Correct, description of 'absolute' counter example (no diag)	%	2	3	3
32D	Number	170	204	374
Correct, diagram showing 'absolute' counter example	%	13	16	14
41N	Number	7	7	14
Correct, analytic, dynamic explanation (no diagram)	%	1	1	1
41D	Number	3	9	12
Correct, analytic, dynamic explanation with diagram	%	0	1	0
42N	Number	14	14	28
Correct, analytic, non-dynamic explanation (no diagram)	%	1	1	1
42D	Number	16	13	29
Correct, analytic, non-dynamic explanation with diagram	%	1	1	1

Table 6G1.3 Y8G1 correct answer, use of diagram

Y8G1		Girls	Boys	All
21	Number	68	61	129
Correct, weak explanation (no diagram)	%	5	5	5
	Mean Baseline Score	15.4	15.4	15.4
22	Number	30	40	70
Correct, weak diagram	%	2	3	3
	Mean Baseline Score	14.2	15.7	15.0
31	Number	262	283	545
Correct, diagram showing clear counter example	%	19	22	20
	Mean Baseline Score	16.5	16.7	16.6
32N	Number	28	44	72
Correct, description of 'absolute' counter example (no diagram)	%	2	3	3
	Mean Baseline Score	16.1	17.7	17.0
32D	Number	170	204	374
Correct, diagram showing 'absolute' counter example	%	13	16	14
	Mean Baseline Score	16.8	17.4	17.1
41N	Number	7	7	14
Correct, analytic, dynamic explanation (no diagram)	%	1	1	1
	Mean Baseline Score	15.4	16.9	16.1
41D	Number	3	9	12
Correct, analytic, dynamic explanation with diagram	%	0	1	0
	Mean Baseline Score	17.0	17.6	17.4
42N	Number	14	14	28
Correct, analytic, non-dynamic explanation (no diagram)	%	1	1	1
	Mean Baseline Score	16.9	16.9	16.9
42D	Number	16	13	29
Correct, analytic, non-dynamic explanation with diagram	%	1	1	1
	Mean Baseline Score	19.9	18.5	19.3

Table 6G1.4 Y8G1 correct answer, use of diagram/baseline

Y8G1		Girls	Boys	All
Correct, no diagram	Number	117	126	243
(c21,32N,41N,42N	%	9	10	9
Correct; diagram	Number	481	549	1030
(c22,31,32D,41D,42D)	%	35	42	39
Total sample	Number	1360	1303	2663

Table 6G1.5 Y8G1 correct answer, use of diagram, summary

Y8G1		Girls	Boys	All
Correct, no diagram (c21,32N,41N,42N	Number	117	126	243
	%	9	10	9
	Mean Baseline Score	15.8	16.4	16.1
Correct; diagram (c22,31,32D,41D,42D)	Number	481	549	1030
	%	35	42	39
	Mean Baseline Score	16.5	16.9	16.7
Total sample	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G1.6 Y8G1 correct answer, use of diagram, summary/baseline

Y8G1		Girls	Boys	All
Incorrect decision (code 11)	Number	670	496	1166
	%	49	38	44
Correct decision but no explanation (code 12)	Number	63	100	163
	%	5	8	6
Correct decision; weak explanation or counter example (code 2)	Number	98	101	199
	%	7	8	7
Correct decision; explicit counter example (code 3)	Number	460	531	991
	%	34	41	37
Correct decision; correct analytic reason (code 4)	Number	40	43	83
	%	3	3	3
Other incorrect (code 9)	Number	29	32	61
	%	2	2	2
Total	Number	1360	1303	2663

Table 6G1.7 Y8G1 broad-code frequencies

Y8G1		Girls	Boys	All
Incorrect decision (code 11)	Number	670	496	1166
	%	49	38	44
	Mean Baseline Score	13.7	14.4	14.0
Correct decision but no explanation (code 12)	Number	63	100	163
	%	5	8	6
	Mean Baseline Score	13.2	15.5	14.6
Correct decision; weak explanation or counter example (code 2)	Number	98	101	199
	%	7	8	7
	Mean Baseline Score	15.0	15.5	15.3
Correct decision; explicit counter example (code 3)	Number	460	531	991
	%	34	41	37
	Mean Baseline Score	16.5	17.0	16.8
Correct decision; correct analytic reason (code 4)	Number	40	43	83
	%	3	3	3
	Mean Baseline Score	17.9	17.5	17.7
Other incorrect (code 9)	Number	29	32	61
	%	2	2	2
	Mean Baseline Score	14.3	14.9	14.6
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G1.8 Y8G1 broad-code frequencies/baseline

Y8G1		Girls	Boys	All
Incorrect decision, or correct but no explanation (code 1)	Number	733	596	1329
	%	54	46	50
Correct decision; weak explanation or counter example (code 2)	Number	98	101	199
	%	7	8	7
Correct decision; explicit counter example (code 3)	Number	460	531	991
	%	34	41	37
Correct decision; correct analytic reason (code 4)	Number	40	43	83
	%	3	3	3
Other incorrect (code 9)	Number	29	32	61
	%	2	2	2
Total	Number	1360	1303	2663

Table 6G1.9 Y8G1 main-code frequencies

Y8G1		Girls	Boys	All
Incorrect decision, or correct but no explanation (code 1)	Number	733	596	1329
	%	54	46	50
	Mean Baseline Score	13.6	14.6	14.1
Correct decision; weak explanation or counter example (code 2)	Number	98	101	199
	%	7	8	7
	Mean Baseline Score	15.0	15.5	15.3
Correct decision; explicit counter example (code 3)	Number	460	531	991
	%	34	41	37
	Mean Baseline Score	16.5	17.0	16.8
Correct decision; correct analytic reason (code 4)	Number	40	43	83
	%	3	3	3
	Mean Baseline Score	17.9	17.5	17.7
Other incorrect (code 9)	Number	29	32	61
	%	2	2	2
	Mean Baseline Score	14.3	14.9	14.6
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G1.10 Y8G1 main-code frequencies/baseline

Y8G1		Girls	Boys	All
0 (code 11) Incorrect decision	Number	670	496	1166
	%	49	38	44
1 (code 12) Correct answer; no or incorrect explanation	Number	63	100	163
	%	5	8	6
2 (code 21,22) Correct answer; weak explanation or weak diagram	Number	98	101	199
	%	7	8	7
2.5 (code 31) Correct answer; clear counter example	Number	262	283	545
	%	19	22	20
3 (code 32,41,42) Correct answer; 'absolute' counter example or analytic reason	Number	238	291	529
	%	18	22	20
0 (code 91,92,93) Other incorrect answers	Number	29	32	61
	%	2	2	2
Total	Number	1360	1303	2663
	Mean Total Score	1.20	1.44	1.32

Table 6G1.11 Y8G1 scores

Y8G1		Girls	Boys	All
0 (code 11) Incorrect decision	Number	670	496	1166
	%	49	38	44
	Mean Baseline Score	13.7	14.4	14.0
1 (code 12) Correct answer; no or incorrect explanation	Number	63	100	163
	%	5	8	6
	Mean Baseline Score	13.2	15.5	14.6
2 (code 21,22) Correct answer; weak explanation or weak diagram	Number	98	101	199
	%	7	8	7
	Mean Baseline Score	15.0	15.5	15.3
2.5 (code 31) Correct answer; clear counter example	Number	262	283	545
	%	19	22	20
	Mean Baseline Score	16.5	16.7	16.6
3 (code 32,41,42) Correct answer; 'absolute' counter example or analytic reason	Number	238	291	529
	%	18	22	20
	Mean Baseline Score	16.9	17.5	17.2
0 (code 91,92,93) Other incorrect answers	Number	29	32	61
	%	2	2	2
	Mean Baseline Score	14.3	14.9	14.6
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3
	Mean Total Score	1.20	1.44	1.32

Table 6G1.12 Y8G1 scores/baseline

Frequencies For Item Y8G2a

Y8G2a		Girls	Boys	All
11	Number	493	428	921
	%	36	33	35
12	Number	220	214	434
	%	16	16	16
12A	Number	0	4	4
	%	0	0	0
13	Number	10	5	15
	%	1	0	1
20	Number	41	58	99
	%	3	4	4
20A	Number	5	6	11
	%	0	0	0
31	Number	347	363	710
	%	26	28	27
31A	Number	21	26	47
	%	2	2	2
32	Number	155	119	274
	%	11	9	10
32A	Number	14	11	25
	%	1	1	1
32L	Number	4	2	6
	%	0	0	0
91	Number	33	47	80
	%	2	4	3
92	Number	1	4	5
	%	0	0	0
93	Number	16	16	32
	%	1	1	1
Total	Number	1360	1303	2663

Table 6G2a.1 Y8G2a code frequencies (number and percent)

Y8G2a		Girls	Boys	All
11	Number	493	428	921
	Mean Baseline Score	14.1	15.0	14.5
12	Number	220	214	434
	Mean Baseline Score	13.4	14.7	14.0
12A	Number	0	4	4
	Mean Baseline Score	x	16.0	16.0
13	Number	10	5	15
	Mean Baseline Score	17.3	15.6	16.7
20	Number	41	58	99
	Mean Baseline Score	16.4	16.6	16.5
20A	Number	5	6	11
	Mean Baseline Score	18.4	16.2	17.2
31	Number	347	363	710
	Mean Baseline Score	15.7	16.9	16.3
31A	Number	21	26	47
	Mean Baseline Score	16.1	16.0	16.1
32	Number	155	119	274
	Mean Baseline Score	16.5	17.2	16.8
32A	Number	14	11	25
	Mean Baseline Score	14.7	17.5	15.9
32L	Number	4	2	6
	Mean Baseline Score	19.0	20.5	19.5
91	Number	33	47	80
	Mean Baseline Score	13.9	14.2	14.1
92	Number	1	4	5
	Mean Baseline Score	11.0	17.3	16.0
93	Number	16	16	32
	Mean Baseline Score	15.1	14.4	14.8
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G2a.2 Y8G2a mean baseline score per code

Y8G2a		Girls	Boys	All
Incorrect decision (code 11)	Number	493	428	921
	%	36	33	35
Correct decision, no logical argument (code 12,13)	Number	230	223	453
	%	17	17	17
Correct decision; reference to specific example (code 2)	Number	46	64	110
	%	3	5	4
Correct decision; general logical argument (code 3)	Number	541	521	1062
	%	40	40	40
Other incorrect (code 9)	Number	50	67	117
	%	4	5	4
Total	Number	1360	1303	2663

Table 6G2a.3 Y8G2a broad-code frequencies

Y8G2a		Girls	Boys	All
Incorrect decision (code 11)	Number	493	428	921
	%	36	33	35
	Mean Baseline Score	14.1	15.0	14.5
Correct decision, no logical argument (code 12,13)	Number	230	223	453
	%	17	17	17
	Mean Baseline Score	13.6	14.7	14.1
Correct decision; reference to specific example (code 2)	Number	46	64	110
	%	3	5	4
	Mean Baseline Score	16.6	16.5	16.6
Correct decision; general logical argument (code 3)	Number	541	521	1062
	%	40	40	40
	Mean Baseline Score	16.0	16.9	16.4
Other incorrect (code 9)	Number	50	67	117
	%	4	5	4
	Mean Baseline Score	14.2	14.4	14.4
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G2a.4 Y8G2a broad-code frequencies/baseline

Y8G2a		Girls	Boys	All
Incorrect or correct decision, no logical argument (code 1)	Number	723	651	1374
	%	53	50	52
Correct decision; reference to specific example (code 2)	Number	46	64	110
	%	3	5	4
Correct decision; general logical argument (code 3)	Number	541	521	1062
	%	40	40	40
Other incorrect (code 9)	Number	50	67	117
	%	4	5	4
Total	Number	1360	1303	2663

Table 6G2a.5 Y8G2a main-code frequencies

Y8G2a		Girls	Boys	All
Incorrect or correct decision, no logical argument (code 1)	Number	723	651	1374
	%	53	50	52
	Mean Baseline Score	13.9	14.9	14.4
Correct decision; reference to specific example (code 2)	Number	46	64	110
	%	3	5	4
	Mean Baseline Score	16.6	16.5	16.6
Correct decision; general logical argument (code 3)	Number	541	521	1062
	%	40	40	40
	Mean Baseline Score	16.0	16.9	16.4
Other incorrect (code 9)	Number	50	67	117
	%	4	5	4
	Mean Baseline Score	14.2	14.4	14.4
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G2a.6 Y8G2a main-code frequencies/baseline

Y8G2a		Girls	Boys	All
0 (code 11) Incorrect decision	Number	493	428	921
	%	36	33	35
1 (code 12,13) Correct answer; no logical argument	Number	230	223	453
	%	17	17	17
2 (code 20) Correct answer; reference to specific example	Number	46	64	110
	%	3	5	4
2.5 (code 31) Correct answer; logical argument but not explicit	Number	368	389	757
	%	27	30	28
3 (code 32) Correct answer; explicit logical argument	Number	173	132	305
	%	13	10	11
0 (code 91,92,93) Other incorrect answers	Number	50	67	117
	%	4	5	4
Total	Number	1360	1303	2663
	Mean Total Score	1.29	1.32	1.31

Table 6G2a.7 Y8G2a scores

Y8G2a		Girls	Boys	All
0 (code 11) Incorrect decision	Number	493	428	921
	%	36	33	35
	Mean Baseline Score	14.1	15.0	14.5
1 (code 12,13) Correct answer; no logical argument	Number	230	223	453
	%	17	17	17
	Mean Baseline Score	13.6	14.7	14.1
2 (code 20) Correct answer; reference to specific example	Number	46	64	110
	%	3	5	4
	Mean Baseline Score	16.6	16.5	16.6
2.5 (code 31) Correct answer; logical argument but not explicit	Number	368	389	757
	%	27	30	28
	Mean Baseline Score	15.8	16.8	16.3
3 (code 32) Correct answer; explicit logical argument	Number	173	132	305
	%	13	10	11
	Mean Baseline Score	16.4	17.2	16.8
0 (code 91,92,93) Other incorrect answers	Number	50	67	117
	%	4	5	4
	Mean Baseline Score	14.2	14.4	14.4
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3
	Mean Total Score	1.29	1.32	1.31

Table 6G2a.8 Y8G2a scores/baseline

Frequencies For Item Y8G2b

Y8G2b		Girls	Boys	All
11	Number	665	705	1365
	%			
12	Number	195	241	436
	%	14	18	16
12A	Number	2	4	6
	%	0	0	0
13	Number	14	5	19
	%	1	0	1
13E	Number	6	1	7
	%	0	0	0
20	Number	272	264	536
	%	20	20	20
20A	Number	61	98	159
	%	4	8	6
31N	Number	128	99	227
	%	9	8	9
31NA	Number	17	24	41
	%	1	2	2
31D	Number	158	156	314
	%	12	12	12
31DA	Number	33	40	73
	%	2	3	3
32N	Number	74	47	121
	%	5	4	5
32D	Number	115	71	186
	%	8	5	7
40N	Number	11	13	24
	%	1	1	1
40D	Number	57	32	89
	%	4	2	3
91	Number	82	86	168
	%	6	7	6
92	Number	11	4	15
	%	1	0	1
93	Number	58	48	106
	%	4	4	4
Total	Number	1360	1303	2663

Table 6G2b.1 Y8G2b code frequencies (number and percent)

Y8G2b		Girls	Boys	All
11	Number	66	70	136
	Mean Baseline Score	12.3	14.7	13.6
12	Number	195	241	436
	Mean Baseline Score	14.4	14.8	14.6
12A	Number	2	4	6
	Mean Baseline Score	13.0	16.5	15.3
13	Number	14	5	19
	Mean Baseline Score	14.6	12.8	14.2
13E	Number	6	1	7
	Mean Baseline Score	15.8	13.0	15.4
20	Number	272	264	536
	Mean Baseline Score	15.2	16.4	15.8
20A	Number	61	98	159
	Mean Baseline Score	15.7	16.7	16.3
31N	Number	128	99	227
	Mean Baseline Score	15.6	16.6	16.0
31NA	Number	17	24	41
	Mean Baseline Score	18.6	17.2	17.8
31D	Number	158	156	314
	Mean Baseline Score	15.2	16.3	15.7
31DA	Number	33	40	73
	Mean Baseline Score	17.8	17.4	17.6
32N	Number	74	47	121
	Mean Baseline Score	14.6	15.4	14.9
32D	Number	115	71	186
	Mean Baseline Score	15.4	15.9	15.6
40N	Number	11	13	24
	Mean Baseline Score	16.7	17.1	16.9
40D	Number	57	32	89
	Mean Baseline Score	15.5	16.4	15.9
91	Number	82	86	168
	Mean Baseline Score	12.4	14.0	13.2
92	Number	11	4	15
	Mean Baseline Score	12.6	17.3	13.9
93	Number	58	48	106
	Mean Baseline Score	14.8	15.8	15.3
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G2b.2 Y8G2b mean baseline score per code

Y8G2b		Girls	Boys	All
31N,31NA	Number	145	123	268
Correct decision; squares in same orientation, no diagram	%	11	9	10
31D,31DA	Number	191	196	387
Correct decision; considers squares in same orientation, diagram	%	14	15	15
32N	Number	74	47	121
Correct decision; relates to four identical overlaps, no diagram	%	5	4	5
32D	Number	115	71	186
Correct decision; relates to four identical overlaps, diagram	%	8	5	7
40N	Number	11	13	24
Correct decision; uses compensation argument, no diagram	%	1	1	1
40D	Number	57	32	89
Correct decision; uses compensation argument, diagram	%	4	2	3
Total sample	Number	1360	1303	2663

Table 6G2b.3 Y8G2b correct answer, use of diagram

Y8G2b		Girls	Boys	All
31N,31NA	Number	145	123	268
Correct decision; considers squares in same orientation,	%	11	9	10
no diagram	Mean Baseline Score	16.0	16.7	16.3
31D,31DA	Number	191	196	387
Correct decision; considers squares in same orientation,	%	14	15	15
diagram	Mean Baseline Score	15.7	16.5	16.1
32N	Number	74	47	121
Correct decision; relates to four identical overlaps, no	%	5	4	5
diagram	Mean Baseline Score	14.6	15.4	14.9
32D	Number	115	71	186
Correct decision; relates to four identical overlaps,	%	8	5	7
diagram	Mean Baseline Score	15.4	15.9	15.6
40N	Number	11	13	24
Correct decision; uses compensation argument, no	%	1	1	1
diagram	Mean Baseline Score	16.7	17.1	16.9
40D	Number	57	32	89
Correct decision; uses compensation argument, diagram	%	4	2	3
	Mean Baseline Score	15.5	16.4	15.9
Total sample	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G2b.4 Y8G2b correct answer, use of diagram/baseline

Y8G2b		Girls	Boys	All
31N,31NA,32N,40N	Number	230	183	413
Correct decision; structural explanation, no diagram	%	17	14	16
31D,31DA,32D,40D	Number	363	299	662
Correct decision; structural explanation, diagram	%	27	23	25
Total sample	Number	1360	1303	2663

Table 6G2b.5 Y8G2b correct answer, use of diagram, summary

Y8G2b		Girls	Boys	All
31N,31NA,32N,40N	Number	230	183	413
Correct decision; structural explanation, no diagram	%	17	14	16
	Mean Baseline Score	15.5	16.4	15.9
31D,31DA,32D,40D	Number	363	299	662
Correct decision; structural explanation, diagram	%	27	23	25
	Mean Baseline Score	15.6	16.3	15.9
Total sample	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G2b.6 Y8G2b correct answer, use of diagram, summary/baseline

Y8G2b		Girls	Boys	All
Correct or close estimate; no structural explanation	Number	283	321	604
(code 1)	%	21	25	23
Correct answer; reason not fully explicit	Number	333	362	695
(code 20)	%	24	28	26
Correct answer; considers squares in same orientation	Number	336	319	655
(code 31)	%	25	24	25
Correct answer; relates to four identical overlaps	Number	189	118	307
(code 32)	%	14	9	12
Correct answer; uses compensation argument	Number	68	45	113
(code 40)	%	5	3	4
Other incorrect	Number	151	138	289
(code 9)	%	11	11	11
Total	Number	1360	1303	2663

Table 6G2b.7 Y8G2b broad-code frequencies

Y8G2b		Girls	Boys	All
Correct or close estimate; no structural explanation (code 1)	Number	283	321	604
	%	21	25	23
	Mean Baseline Score	14.0	14.8	14.4
Correct answer; reason not fully explicit (code 20)	Number	333	362	695
	%	24	28	26
	Mean Baseline Score	15.3	16.5	15.9
Correct answer; considers squares in same orientation (code 31)	Number	336	319	655
	%	25	24	25
	Mean Baseline Score	15.8	16.6	16.2
Correct answer; relates to four identical overlaps (code 32)	Number	189	118	307
	%	14	9	12
	Mean Baseline Score	15.1	15.7	15.3
Correct answer; uses compensation argument (code 40)	Number	68	45	113
	%	5	3	4
	Mean Baseline Score	15.7	16.6	16.1
Other incorrect (code 9)	Number	151	138	289
	%	11	11	11
	Mean Baseline Score	12.7	14.1	13.4
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G2b.8 Y8G2b broad-code frequencies/baseline

Y8G2b		Girls	Boys	All
0 (code 11) Close but wrong estimate;	Number	66	70	136
	%	5	5	5
1 (code 12,13) Correct answer; no structural explanation	Number	217	251	468
	%	16	19	18
2 (code 20) Correct answer; reason not fully explicit	Number	333	362	695
	%	24	28	26
3 (code 31,32,40) Correct answer; explicit logical argument	Number	593	482	1075
	%	44	37	40
0 (code 91,92,93) Other incorrect answers	Number	151	138	289
	%	11	11	11
Total	Number	1360	1303	2663
	Mean Total Score	1.96	1.86	1.91

Table 6G2b.9 Y8G2b scores

Y8G2b		Girls	Boys	All
0 (code 11) Close but wrong estimate;	Number	66	70	136
	%	5	5	5
	Mean Baseline Score	12.3	14.7	13.6
1 (code 12,13) Correct answer; no structural explanation	Number	217	251	468
	%	16	19	18
	Mean Baseline Score	14.5	14.8	14.6
2 (code 20) Correct answer; reason not fully explicit	Number	333	362	695
	%	24	28	26
	Mean Baseline Score	15.3	16.5	15.9
3 (code 31,32,40) Correct answer; explicit logical argument	Number	593	482	1075
	%	44	37	40
	Mean Baseline Score	15.6	16.4	15.9
0 (code 91,92,93) Other incorrect answers	Number	151	138	289
	%	11	11	11
	Mean Baseline Score	12.7	14.1	13.4
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3
	Mean Total Score	1.96	1.86	1.91

Table 6G2b.10 Y8G2b scores/baseline

Frequencies For Item Y8G3

Y8G3a (own approach)		Girls	Boys	All
Choice A	Number	617	442	1059
	%	45	34	40
Choice B	Number	471	448	919
	%	35	34	35
Choice C	Number	102	172	274
	%	8	13	10
Choice D	Number	120	184	304
	%	9	14	11
c9	Number	50	57	107
	%	4	4	4
Total	Number	1360	1303	2663

Table 6G3.1 Y8G3a choice frequencies (number and percent)

Y8G3a (own approach)		Girls	Boys	All
Choice A	Number	617	442	1059
	Mean Baseline Score	15.2	16.0	15.5
Choice B	Number	471	448	919
	Mean Baseline Score	14.4	15.5	14.9
Choice C	Number	102	172	274
	Mean Baseline Score	15.0	15.8	15.5
Choice D	Number	120	184	304
	Mean Baseline Score	15.0	16.0	15.6
c9	Number	50	57	107
	Mean Baseline Score	13.9	15.3	14.7
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G3.2 Y8G3a mean baseline score per choice

Y8G3a (own approach)		Girls	Boys	All
Choice A	Number	617	442	1059
	%	45	34	40
	Mean Baseline Score	15.2	16.0	15.5
Choice B	Number	471	448	919
	%	35	34	35
	Mean Baseline Score	14.4	15.5	14.9
Choice C	Number	102	172	274
	%	8	13	10
	Mean Baseline Score	15.0	15.8	15.5
Choice D	Number	120	184	304
	%	9	14	11
	Mean Baseline Score	15.0	16.0	15.6
c9	Number	50	57	107
	%	4	4	4
	Mean Baseline Score	13.9	15.3	14.7
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G3.3 Y8G3a choice frequencies/baseline

Y8G3b (best mark)		Girls	Boys	All
Choice A	Number	172	150	322
	%	13	12	12
Choice B	Number	343	236	579
	%	25	18	22
Choice C	Number	632	688	1320
	%	46	53	50
Choice D	Number	131	159	290
	%	10	12	11
c9	Number	82	70	152
	%	6	5	6
Total	Number	1360	1303	2663

Table 6G3.4 Y8G3b choice frequencies (number and percent)

Y8G3b (best mark)		Girls	Boys	All
Choice A	Number	172	150	322
	Mean Baseline Score	14.2	15.9	15.0
Choice B	Number	343	236	579
	Mean Baseline Score	14.8	15.3	15.0
Choice C	Number	632	688	1320
	Mean Baseline Score	15.1	15.9	15.5
Choice D	Number	131	159	290
	Mean Baseline Score	15.1	15.6	15.4
c9	Number	82	70	152
	Mean Baseline Score	14.2	15.6	14.9
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G3.5 Y8G3b mean baseline score per choice

Y8G3b (best mark)		Girls	Boys	All
Choice A	Number	172	150	322
	%	13	12	12
	Mean Baseline Score	14.2	15.9	15.0
Choice B	Number	343	236	579
	%	25	18	22
	Mean Baseline Score	14.8	15.3	15.0
Choice C	Number	632	688	1320
	%	46	53	50
	Mean Baseline Score	15.1	15.9	15.5
Choice D	Number	131	159	290
	%	10	12	11
	Mean Baseline Score	15.1	15.6	15.4
c9	Number	82	70	152
	%	6	5	6
	Mean Baseline Score	14.2	15.6	14.9
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G3.6 Y8G3b choice frequencies/baseline

Y8G3a/G3b		G3a: Own approach					
		Choice A	Choice B	Choice C	Choice D	other	Total
G3b: Best mark	Choice A	8	3	1	0	0	12
	Choice B	9	10	1	1	0	22
	Choice C	18	17	8	5	0	50
	Choice D	3	3	0	4	0	11
	other	1	1	0	0	3	6
	Total	40	35	10	11	4	100

Table 6G3.7 Y8G3a/G3b choice frequencies (percent)

Y8G3		Own	Best
Choice A	Number	2	7
Choice B	Number	10	8
Choice C	Number	75	74
Choice D	Number	1	0
c9	Number	16	15
Total	Number	104	104

Table 6G3.8 Y8G3 teachers' choice frequencies (number)

Y8G3		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 6G3.9 Y8G3 teacher rankings of choices: frequencies (number)

Y8G3	Students who selected					
	Choice A for own approach N = 1059	Choice B for own approach N = 919	Choice C for own approach N = 274	Choice D for own approach N = 304	other N = 107	Total N = 2663
Mean validity rating of choice A	0.96	0.95	1.06	1.22	0.34	0.97
Mean validity rating of choice B	0.91	0.68	0.99	1.16	0.28	0.84
Mean validity rating of choice C	0.60	0.65	1.03	0.46	0.18	0.63
Mean validity rating of choice D	0.25	0.25	0.32	0.07	0.08	0.23
Total mean validity rating	2.72	2.53	3.41	2.91	0.88	2.67

Note: Maximum possible validity rating for choice A, B, C, D is 2, 2, 2, 1 respectively

Table 6G3.10 Y8G3 mean validity rating of each choice, for students who chose A, B, C or D for own approach

Y8G3	Students who selected					
	Choice A for best mark N = 322	Choice B for best mark N = 579	Choice C for best mark N = 1320	Choice D for best mark N = 290	other N = 152	Total N = 2663
Mean validity rating of choice A	0.75	0.95	1.05	1.13	0.58	0.97
Mean validity rating of choice B	0.84	0.59	0.95	1.04	0.47	0.84
Mean validity rating of choice C	0.50	0.52	0.81	0.34	0.24	0.63
Mean validity rating of choice D	0.28	0.27	0.25	0.08	0.11	0.23
Total mean validity rating	2.36	2.33	3.07	2.60	1.39	2.67

Note: Maximum possible validity rating for choice A, B, C, D is 2, 2, 2, 1 respectively

Table 6G3.11 Y8G3 mean validity rating of each choice, for students who chose A, B, C or D for best mark

Y8G3 Mean explanatory power ([mean of N(agree) - mean of N(disagree)]	Students who selected					
	Choice A for own approach N = 1059	Choice B for own approach N = 919	Choice C for own approach N = 274	Choice D for own approach N = 304	other N = 107	Total N = 2663
Mean explanatory power of A	0.31	0.16	0.02	-0.06	0.00	0.17
Mean explanatory power of B	0.32	0.51	0.09	0.01	0.07	0.32
Mean explanatory power of C	0.46	0.46	0.69	0.17	0.09	0.43
Mean explanatory power of D	0.12	0.11	-0.02	0.56	0.01	0.15

Table 6G3.12 Y8G3 mean explanatory power of each choice, for students who chose A, B, C or D for own approach

Y8G3 Mean explanatory power ([mean of N(agree) - mean of N(disagree)]	Students who selected					
	Choice A for best mark N = 322	Choice B for best mark N = 579	Choice C for best mark N = 1320	Choice D for best mark N = 290	other N = 152	Total N = 2663
Mean explanatory power of A	0.46	0.22	0.13	-0.02	0.13	0.17
Mean explanatory power of B	0.31	0.54	0.28	0.16	0.16	0.32
Mean explanatory power of C	0.26	0.32	0.62	0.19	0.13	0.43
Mean explanatory power of D	0.21	0.03	0.10	0.55	0.10	0.15

Table 6G3.13 Y8G3 mean explanatory power of each choice, for students who chose A, B, C or D for best mark

Frequencies For Item Y8G4a

Y8G4a		Girls	Boys	All
20	Number	343	263	606
	%	25	20	23
30	Number	665	657	1322
	%	49	50	50
30E	Number	19	11	30
	%	1	1	1
91	Number	63	63	126
	%	5	5	5
92	Number	7	7	14
	%	1	1	1
93	Number	263	302	565
	%	19	23	21
Total	Number	1360	1303	2663

Table 6G4a.1 Y8G4a code frequencies (number and percent)

Y8G4a		Girls	Boys	All
20	Number	343	263	606
	Mean Baseline Score	14.4	15.5	14.9
30	Number	665	657	1322
	Mean Baseline Score	16.1	17.0	16.5
30E	Number	19	11	30
	Mean Baseline Score	15.4	15.6	15.5
91	Number	63	63	126
	Mean Baseline Score	12.8	13.9	13.3
92	Number	7	7	14
	Mean Baseline Score	13.3	14.1	13.7
93	Number	263	302	565
	Mean Baseline Score	12.7	13.9	13.3
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G4a.2 Y8G4a mean baseline score per code

Y8G4a		Girls	Boys	All
Correct first step of calculation (code 20)	Number	343	263	606
	%	25	20	23
Correct calculation (code 30,30E)	Number	684	668	1352
	%	50	51	51
Other incorrect (code 9)	Number	333	372	705
	%	24	29	26
Total	Number	1360	1303	2663

Table 6G4a.3 Y8G4a broad-code frequencies

Y8G4a		Girls	Boys	All
Correct first step of calculation (code 20)	Number	343	263	606
	%	25	20	23
	Mean Baseline Score	14.4	15.5	14.9
Correct calculation (code 30,30E)	Number	684	668	1352
	%	50	51	51
	Mean Baseline Score	16.1	16.9	16.5
Other incorrect (code 9)	Number	333	372	705
	%	24	29	26
	Mean Baseline Score	12.7	13.9	13.3
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G4a.4 Y8G4a broad-code frequencies/baseline

Y8G4a		Girls	Boys	All
1 (code 20) Correct first step of calculation	Number	343	263	606
	%	25	20	23
2 (code 30,30E) Correct calculation	Number	684	668	1352
	%	50	51	51
0 (code 91,92,93) Other incorrect answers	Number	333	372	705
	%	24	29	26
Total	Number	1360	1303	2663
	Mean Total Score	1.26	1.23	1.24

Table 6G4a.5 Y8G4a scores

Y8G4a		Girls	Boys	All
1 (code 20) Correct first step of calculation	Number	343	263	606
	%	25	20	23
	Mean Baseline Score	14.4	15.5	14.9
2 (code 30,30E) Correct calculation	Number	684	668	1352
	%	50	51	51
	Mean Baseline Score	16.1	16.9	16.5
0 (code 91,92,93) Other incorrect answers	Number	333	372	705
	%	24	29	26
	Mean Baseline Score	12.7	13.9	13.3
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3
	Mean Total Score	1.26	1.23	1.24

Table 6G4a.6 Y8G4a scores/baseline

Frequencies For Item Y8G4b

Y8G4b		Girls	Boys	All
30	Number	939	809	1748
	%	69	62	66
91	Number	60	75	135
	%	4	6	5
92	Number	8	9	17
	%	1	1	1
93	Number	353	410	763
	%	26	31	29
Total	Number	1360	1303	2663

Table 6G4b.1 Y8G4b code frequencies (number and percent)

Y8G4b		Girls	Boys	All
30	Number	939	809	1748
	Mean Baseline Score	15.7	16.6	16.1
91	Number	60	75	135
	Mean Baseline Score	12.8	13.7	13.3
92	Number	8	9	17
	Mean Baseline Score	12.6	13.8	13.2
93	Number	353	410	763
	Mean Baseline Score	13.1	14.6	13.9
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G4b.2 Y8G4b mean baseline score per code

Y8G4b		Girls	Boys	All
Correct order of calculations	Number	939	809	1748
(code 30)	%	69	62	66
Incorrect	Number	421	494	915
(code 9)	%	31	38	34
Total	Number	1360	1303	2663

Table 6G4b.3 Y8G4b broad-code frequencies

Y8G4b		Girls	Boys	All
Correct order of calculations	Number	939	809	1748
(code 30)	%	69	62	66
	Mean Baseline Score	15.7	16.6	16.1
Incorrect	Number	421	494	915
(code 9)	%	31	38	34
	Mean Baseline Score	13.0	14.4	13.8
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G4b.4 Y8G4b broad-code frequencies/baseline

Y8G4b		Girls	Boys	All
2 (code 30) Correct order of calculations	Number	939	809	1748
	%	69	62	66
0 (code 9) Incorrect	Number	421	494	915
	%	31	38	34
Total	Number	1360	1303	2663
	Mean Total Score	1.38	1.24	1.31

Table 6G4b.5 Y8G4b scores

Y8G4b		Girls	Boys	All
2 (code 30) Correct order of calculations	Number	939	809	1748
	%	69	62	66
	Mean Baseline Score	15.7	16.6	16.1
0 (code 9) Incorrect	Number	421	494	915
	%	31	38	34
	Mean Baseline Score	13.0	14.4	13.8
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3
	Mean Total Score	1.38	1.24	1.31

Table 6G4b.6 Y8G4b scores/baseline

Frequencies For Item Y8G4c

Y8G4c		Girls	Boys	All
30	Number	605	529	1134
	%	44	41	43
91	Number	75	84	159
	%	6	6	6
92	Number	7	12	19
	%	1	1	1
93	Number	673	678	1351
	%	49	52	51
Total	Number	1360	1303	2663

Table 6G4c.1 Y8G4c code frequencies (number and percent)

Y8G4c		Girls	Boys	All
30	Number	605	529	1134
	Mean Baseline Score	15.9	16.9	16.4
91	Number	75	84	159
	Mean Baseline Score	12.7	14.2	13.5
92	Number	7	12	19
	Mean Baseline Score	12.9	13.6	13.3
93	Number	673	678	1351
	Mean Baseline Score	14.2	15.1	14.6
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G4c.2 Y8G4c mean baseline score per code

Y8G4c		Girls	Boys	All
Correct order of reasons	Number	605	529	1134
(code 30)	%	44	41	43
Incorrect	Number	755	774	1529
(code 9)	%	56	59	57
Total	Number	1360	1303	2663

Table 6G4c.3 Y8G4c broad-code frequencies

Y8G4c		Girls	Boys	All
Correct order of reasons	Number	605	529	1134
(code 30)	%	44	41	43
	Mean Baseline Score	15.9	16.9	16.4
Incorrect	Number	755	774	1529
(code 9)	%	56	59	57
	Mean Baseline Score	14.0	15.0	14.5
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6G4c.4 Y8G4c broad-code frequencies/baseline

Y8G4c		Girls	Boys	All
2 (code 30) Correct order of	Number	605	529	1134
reasons	%	44	41	43
0 (code 9) Incorrect	Number	755	774	1529
	%	56	59	57
Total	Number	1360	1303	2663
	Mean Total Score	0.89	0.81	0.85

Table 6G4c.5 Y8G4c scores

Y8G4c		Girls	Boys	All
2 (code 30) Correct order of reasons	Number	605	529	1134
	%	44	41	43
	Mean Baseline Score	15.9	16.9	16.4
0 (code 9) Incorrect	Number	755	774	1529
	%	56	59	57
	Mean Baseline Score	14.0	15.0	14.5
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3
	Mean Total Score	0.89	0.81	0.85

Table 6G4c.6 Y8G4c scores/baseline

Frequencies For Item Y8L1a

Y8L1a		Girls	Boys	All
10	Number	1011	886	1897
	%	74	68	71
31	Number	184	203	387
	%	14	16	15
32	Number	154	201	355
	%	11	15	13
91	Number	7	12	19
	%	1	1	1
92	Number	0	0	0
	%	0	0	0
93	Number	4	1	5
	%	0	0	0
Total	Number	1360	1303	2663

Table 6L1a.1 Y8L1a code frequencies (number and percent)

Y8L1a		Girls	Boys	All
10	Number	1011	886	1897
	Mean Baseline Score	14.5	15.5	15.0
31	Number	184	203	387
	Mean Baseline Score	16.4	16.7	16.6
32	Number	154	201	355
	Mean Baseline Score	15.3	16.0	15.7
91	Number	7	12	19
	Mean Baseline Score	15.9	14.6	15.1
92	Number	0	0	0
	Mean Baseline Score	x	x	x
93	Number	4	1	5
	Mean Baseline Score	10.5	18.0	12.0
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6L1a.2 Y8L1a mean baseline score per code

Y8L1a		Girls	Boys	All
"Yes (statements are the same)" (code 10)	Number	1011	886	1897
	%	74	68	71
"No (statements are not the same)" (code 31,32)	Number	338	404	742
	%	25	31	28
Other incorrect (code 9)	Number	11	13	24
	%	1	1	1
Total	Number	1360	1303	2663

Table 6L1a.3 Y8L1a broad-code frequencies

Y8L1a		Girls	Boys	All
"Yes (statements are the same)" (code 10)	Number	1011	886	1897
	%	74	68	71
	Mean Baseline Score	14.5	15.5	15.0
"No (statements are not the same)" (code 31,32)	Number	338	404	742
	%	25	31	28
	Mean Baseline Score	15.9	16.3	16.2
Other incorrect (code 9)	Number	11	13	24
	%	1	1	1
	Mean Baseline Score	13.9	14.8	14.4
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6L1a.4 Y8L1a broad-code frequencies/baseline

Y8L1a		Girls	Boys	All
0 (code 10) "Yes (statements are the same)"	Number	1011	886	1897
	%	74	68	71
1 (code 31) "Yes" changed to "No"	Number	184	203	387
	%	14	16	15
2 (code 32) "No (statements are not the same)"	Number	154	201	355
	%	11	15	13
0 (code 9) Incorrect	Number	11	13	24
	%	1	1	1
Total	Number	1360	1303	2663
	Mean Total Score	0.36	0.46	0.41

Table 6L1a.5 Y8L1a scores

Y8L1a		Girls	Boys	All
0 (code 10) "Yes (statements are the same)"	Number	1011	886	1897
	%	74	68	71
	Mean Baseline Score	14.5	15.5	15.0
1 (code 31) "Yes" changed to "No"	Number	184	203	387
	%	14	16	15
	Mean Baseline Score	16.4	16.7	16.6
2 (code 32) "No (statements are not the same)"	Number	154	201	355
	%	11	15	13
	Mean Baseline Score	15.3	16.0	15.7
0 (code 9) Incorrect	Number	11	13	24
	%	1	1	1
	Mean Baseline Score	13.9	14.8	14.4
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3
	Mean Total Score	0.36	0.46	0.41

Table 6L1a.6 Y8L1a scores/baseline

Frequencies For Item Y8L1b

Y8L1b		Girls	Boys	All
10	Number	629	612	1241
	%	46	47	47
30	Number	654	595	1249
	%	48	46	47
91	Number	13	16	29
	%	1	1	1
92	Number	0	0	0
	%	0	0	0
93	Number	64	80	144
	%	5	6	5
Total	Number	1360	1303	2663

Table 6L1b.1 Y8L1b code frequencies (number and percent)

Y8L1b		Girls	Boys	All
10	Number	629	612	1241
	Mean Baseline Score	13.8	15.1	14.4
30	Number	654	595	1249
	Mean Baseline Score	16.0	16.7	16.4
91	Number	13	16	29
	Mean Baseline Score	14.2	13.9	14.1
92	Number	0	0	0
	Mean Baseline Score	x	x	x
93	Number	64	80	144
	Mean Baseline Score	13.5	14.4	14.0
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6L1b.2 Y8L1b mean baseline score per code

Y8L1b		Girls	Boys	All
"Can't be sure" (code 10)	Number	629	612	1241
	%	46	47	47
"Sum is EVEN" (code 30)	Number	654	595	1249
	%	48	46	47
Other incorrect including "Sum is ODD" and choice of more than one option (code 9)	Number	77	96	173
	%	6	7	6
Total	Number	1360	1303	2663

Table 6L1b.3 Y8L1b broad-code frequencies

Y8L1b		Girls	Boys	All
"Can't be sure" (code 10)	Number	629	612	1241
	%	46	47	47
	Mean Baseline Score	13.8	15.1	14.4
"Sum is EVEN" (code 30)	Number	654	595	1249
	%	48	46	47
	Mean Baseline Score	16.0	16.7	16.4
Other incorrect including "Sum is ODD" and choice of more than one option (code 9)	Number	77	96	173
	%	6	7	6
	Mean Baseline Score	13.6	14.4	14.0
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6L1b.4 Y8L1b broad-code frequencies/baseline

Y8L1b		Girls	Boys	All
0 (code 10) "Can't be sure"	Number	629	612	1241
	%	46	47	47
2 (code 30) "Sum is EVEN"	Number	654	595	1249
	%	48	46	47
0 (code 9) Other incorrect including "Sum is ODD" and choice of more than one option	Number	77	96	173
	%	6	7	6
Total	Number	1360	1303	2663
	Mean Total Score	0.96	0.91	0.94

Table 6L1b.5 Y8L1b scores

Y8L1b		Girls	Boys	All
0 (code 10) "Can't be sure"	Number	629	612	1241
	%	46	47	47
	Mean Baseline Score	13.8	15.1	14.4
2 (code 30) "Sum is EVEN"	Number	654	595	1249
	%	48	46	47
	Mean Baseline Score	16.0	16.7	16.4
0 (code 9) Other incorrect including "Sum is ODD" and choice of more than one option	Number	77	96	173
	%	6	7	6
	Mean Baseline Score	13.6	14.4	14.0
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3
	Mean Total Score	0.96	0.91	0.94

Table 6L1b.6 Y8L1b scores/baseline

Frequencies For Item Y8L1c

Y8L1c		Girls	Boys	All
11	Number	150	171	321
	%	11	13	12
11F	Number	47	33	80
	%	3	3	3
13	Number	41	64	105
	%	3	5	4
13F	Number	4	4	8
	%	0	0	0
14	Number	87	83	170
	%	6	6	6
21	Number	20	10	30
	%	1	1	1
22	Number	262	207	469
	%	19	16	18
22S	Number	7	2	9
	%	1	0	0
22F	Number	11	1	12
	%	1	0	0
23	Number	34	36	70
	%	3	3	3
31	Number	69	118	187
	%	5	9	7
31S	Number	2	2	4
	%	0	0	0
32	Number	291	264	555
	%	21	20	21
32S	Number	5	2	7
	%	0	0	0
41	Number	14	13	27
	%	1	1	1
41S	Number	2	1	3
	%	0	0	0
42	Number	10	24	34
	%	1	2	1
42L	Number	0	1	1
	%	0	0	0
43	Number	21	22	43
	%	2	2	2
44	Number	59	47	106
	%	4	4	4
44L	Number	0	2	2
	%	0	0	0
50	Number	0	0	0
	%	0	0	0
50L	Number	0	0	0
	%	0	0	0
91	Number	56	73	129
	%	4	6	5
92	Number	1	2	3
	%	0	0	0
93	Number	162	118	280
	%	12	9	11
93F	Number	5	3	8
	%	0	0	0
Total	Number	1360	1303	2663

Table 6L1c.1 Y8L1c code frequencies (number and percent)

Y8L1c		Girls	Boys	All
11	Number	150	171	321
	Mean Baseline Score	14.5	14.3	14.4
11F	Number	47	33	80
	Mean Baseline Score	14.0	15.2	14.5
13	Number	41	64	105
	Mean Baseline Score	13.5	14.6	14.2
13F	Number	4	4	8
	Mean Baseline Score	13.3	17.3	15.3
14	Number	87	83	170
	Mean Baseline Score	13.9	15.9	14.8
21	Number	20	10	30
	Mean Baseline Score	17.1	17.0	17.0
22	Number	262	207	469
	Mean Baseline Score	14.7	15.5	15.0
22S	Number	7	2	9
	Mean Baseline Score	14.1	16.0	14.6
22F	Number	11	1	12
	Mean Baseline Score	14.3	14.0	14.3
23	Number	34	36	70
	Mean Baseline Score	13.7	14.4	14.1
31	Number	69	118	187
	Mean Baseline Score	16.0	17.2	16.7
31S	Number	2	2	4
	Mean Baseline Score	17.5	17.5	17.5
32	Number	291	264	555
	Mean Baseline Score	16.1	17.2	16.6
32S	Number	5	2	7
	Mean Baseline Score	16.0	14.5	15.6
41	Number	14	13	27
	Mean Baseline Score	17.6	17.4	17.5
41S	Number	2	1	3
	Mean Baseline Score	14.0	18.0	15.3
42	Number	10	24	34
	Mean Baseline Score	13.3	15.3	14.7
42L	Number	0	1	1
	Mean Baseline Score	x	11.0	11.0
43	Number	21	22	43
	Mean Baseline Score	15.4	17.2	16.3
44	Number	59	47	106
	Mean Baseline Score	17.4	17.4	17.4
44L	Number	0	2	2
	Mean Baseline Score	x	19.5	19.5
50	Number	0	0	0
	Mean Baseline Score	x	x	x
50L	Number	0	0	0
	Mean Baseline Score	x	x	x
91	Number	56	73	129
	Mean Baseline Score	13.2	14.2	13.8
92	Number	1	2	3
	Mean Baseline Score	12.0	17.0	15.3
93	Number	162	118	280
	Mean Baseline Score	13.4	14.9	14.0
93F	Number	5	3	8
	Mean Baseline Score	15.8	12.3	14.5
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6L1c.2 Y8L1c mean baseline score per code

Y8L1c		Girls	Boys	All
11	Number	197	204	401
	%	14	16	15
13	Number	45	68	113
	%	3	5	4
14	Number	87	83	170
	%	6	6	6
21	Number	20	10	30
	%	1	1	1
22	Number	280	210	489
	%	21	16	18
23	Number	34	36	70
	%	3	3	3
31	Number	71	120	191
	%	5	9	7
32	Number	296	266	562
	%	22	20	21
41	Number	16	14	30
	%	1	1	1
42	Number	10	25	35
	%	1	2	1
43	Number	21	22	43
	%	2	2	2
44	Number	59	49	108
	%	4	4	4
50	Number	0	0	0
	%	0	0	0
91	Number	56	73	129
	%	4	6	5
92	Number	1	2	3
	%	0	0	0
93	Number	167	121	288
	%	12	9	11
Total	Number	1360	1303	2663

Table 6L1c.3 Y8L1c main-code frequencies (number and percent)

Y8L1c		Girls	Boys	All
11	Number	197	204	401
	Mean Baseline Score	14.3	14.4	14.4
13	Number	45	68	113
	Mean Baseline Score	13.4	14.8	14.2
14	Number	87	83	170
	Mean Baseline Score	13.9	15.9	14.8
21	Number	20	10	30
	Mean Baseline Score	17.1	17.0	17.0
22	Number	280	210	489
	Mean Baseline Score	14.7	15.5	15.0
23	Number	34	36	70
	Mean Baseline Score	13.7	14.4	14.1
31	Number	71	120	191
	Mean Baseline Score	16.1	17.2	16.8
32	Number	296	266	562
	Mean Baseline Score	16.1	17.1	16.6
41	Number	16	14	30
	Mean Baseline Score	17.1	17.4	17.3
42	Number	10	25	35
	Mean Baseline Score	13.3	15.1	14.6
43	Number	21	22	43
	Mean Baseline Score	15.4	17.2	16.3
44	Number	59	49	108
	Mean Baseline Score	17.4	17.5	17.4
50	Number	0	0	0
	Mean Baseline Score	x	x	x
91	Number	56	73	129
	Mean Baseline Score	13.2	14.2	13.8
92	Number	1	2	3
	Mean Baseline Score	12.0	17.0	15.3
93	Number	167	121	288
	Mean Baseline Score	13.4	14.8	14.0
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6L1c.4 Y8L1c main-code frequencies/baseline

Y8L1c		Girls	Boys	All
Correct or incorrect decision; no valid justification (code 1)	Number	329	355	684
	%	24	27	26
Correct or incorrect decision; incomplete or flawed justification (code 2)	Number	334	256	590
	%	25	20	22
Correct decision; valid justification, specific (code 3)	Number	367	386	753
	%	27	30	28
Correct decision; valid justification, general (code 4)	Number	106	110	216
	%	8	8	8
Correct decision; valid justification, general, plus explanation of why justification true (code 5)	Number	0	0	0
	%	0	0	0
Other incorrect (code 9)	Number	224	196	420
	%	16	15	16
Total	Number	1360	1303	2663

Table 6L1c.5 Y8L1c broad-code frequencies

Y8L1c		Girls	Boys	All
Correct or incorrect decision; no valid justification (code 1)	Number	329	355	684
	%	24	27	26
	Mean Baseline Score	14.1	14.8	14.5
Correct or incorrect decision; incomplete or flawed justification (code 2)	Number	334	256	590
	%	25	20	22
	Mean Baseline Score	14.7	15.4	15.0
Correct decision; valid justification, specific (code 3)	Number	367	386	753
	%	27	30	28
	Mean Baseline Score	16.1	17.2	16.6
Correct decision; valid justification, general (code 4)	Number	106	110	216
	%	8	8	8
	Mean Baseline Score	16.6	16.9	16.7
Correct decision; valid justification, general, plus explanation of why justification true (code 5)	Number	0	0	0
	%	0	0	0
	Mean Baseline Score	x	x	x
Other incorrect (code 9)	Number	224	196	420
	%	16	15	16
	Mean Baseline Score	13.4	14.6	13.9
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6L1c.6 Y8L1c broad-code frequencies/baseline

Y8L1c		Girls	Boys	All
0 (code 1) Correct or incorrect decision; no valid justification	Number	329	355	684
	%	24	27	26
2 (code 2) Correct or incorrect decision; incomplete or flawed justification	Number	334	256	590
	%	25	20	22
2.5 (code 31) Correct decision; valid justification, specific and not explicit	Number	71	120	191
	%	5	9	7
3 (code 32,4) Correct decision; valid justification, general and/or explicit	Number	402	376	778
	%	30	29	29
0 (code 9) Other incorrect	Number	224	196	420
	%	16	15	16
Total	Number	1360	1303	2663
	Mean Total Score	1.51	1.49	1.50

Table 6L1c.7 Y8L1c scores

Y8L1c		Girls	Boys	All
0 (code 1) Correct or incorrect decision; no valid justification	Number	329	355	684
	%	24	27	26
	Mean Baseline Score	14.1	14.8	14.5
2 (code 2) Correct or incorrect decision; incomplete or flawed justification	Number	334	256	590
	%	25	20	22
	Mean Baseline Score	14.1	14.8	14.5
2.5 (code 31) Correct decision; valid justification, specific and not explicit	Number	71	120	191
	%	5	9	7
	Mean Baseline Score	16.1	17.2	16.8
3 (code 32,4) Correct decision; valid justification, general and/or explicit	Number	402	376	778
	%	30	29	29
	Mean Baseline Score	16.2	17.1	16.6
0 (code 9) Other incorrect	Number	224	196	420
	%	16	15	16
	Mean Baseline Score	13.4	14.6	13.9
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3
	Mean Total Score	1.51	1.49	1.50

Table 6L1c.8 Y8L1c scores/baseline

Frequencies For Item Y8L1d

Y8L1d		Girls	Boys	All
11	Number	210	250	460
	%	15	19	17
11J	Number	168	126	294
	%	12	10	11
13	Number	174	210	384
	%	13	16	14
13J	Number	120	95	215
	%	9	7	8
21	Number	149	164	313
	%	11	13	12
21J	Number	45	33	78
	%	3	3	3
22	Number	82	53	135
	%	6	4	5
22J	Number	19	9	28
	%	1	1	1
23	Number	3	1	4
	%	0	0	0
23J	Number	0	1	1
	%	0	0	0
24	Number	31	38	69
	%	2	3	3
24J	Number	6	9	15
	%	0	1	1
41	Number	11	3	14
	%	1	0	1
41J	Number	0	2	2
	%	0	0	0
42	Number	5	6	11
	%	0	0	0
42J	Number	2	0	2
	%	0	0	0
43	Number	24	27	51
	%	2	2	2
43J	Number	3	3	6
	%	0	0	0
43L	Number	0	1	1
	%	0	0	0
44	Number	59	67	126
	%	4	5	5
44J	Number	17	11	28
	%	1	1	1
44L	Number	0	1	1
	%	0	0	0
50	Number	2	1	3
	%	0	0	0
91	Number	92	98	190
	%	7	8	7
92	Number	2	3	5
	%	0	0	0
93	Number	106	71	177
	%	8	5	7
93J	Number	30	20	50
	%	2	2	2
Total	Number	1360	1303	2663

Table 6L1d.1 Y8L1d code frequencies (number and percent)

Y8L1d		Girls	Boys	All
11	Number	210	250	460
	Mean Baseline Score	14.6	15.3	15.0
11J	Number	168	126	294
	Mean Baseline Score	15.2	16.6	15.8
13	Number	174	210	384
	Mean Baseline Score	14.6	15.7	15.2
13J	Number	120	95	215
	Mean Baseline Score	14.2	14.6	14.4
21	Number	149	164	313
	Mean Baseline Score	14.7	15.9	15.3
21J	Number	45	33	78
	Mean Baseline Score	14.2	15.8	14.8
22	Number	82	53	135
	Mean Baseline Score	15.9	17.0	16.3
22J	Number	19	9	28
	Mean Baseline Score	16.7	16.1	16.5
23	Number	3	1	4
	Mean Baseline Score	16.7	13.0	15.8
23J	Number	0	1	1
	Mean Baseline Score	x	17.0	17.0
24	Number	31	38	69
	Mean Baseline Score	16.3	16.1	16.2
24J	Number	6	9	15
	Mean Baseline Score	16.3	15.6	15.9
41	Number	11	3	14
	Mean Baseline Score	15.5	18.0	16.1
41J	Number	0	2	2
	Mean Baseline Score	x	16.5	16.5
42	Number	5	6	11
	Mean Baseline Score	17.0	17.2	17.1
42J	Number	2	0	2
	Mean Baseline Score	15.0	x	15.0
43	Number	24	27	51
	Mean Baseline Score	13.8	15.5	14.7
43J	Number	3	3	6
	Mean Baseline Score	10.0	13.3	11.7
43L	Number	0	1	1
	Mean Baseline Score	x	11.0	11.0
44	Number	59	67	126
	Mean Baseline Score	19.0	19.2	19.1
44J	Number	17	11	28
	Mean Baseline Score	15.8	15.9	15.8
44L	Number	0	1	1
	Mean Baseline Score	x	21.0	21.0
50	Number	2	1	3
	Mean Baseline Score	18.0	18.0	18.0
91	Number	92	98	190
	Mean Baseline Score	13.4	14.3	13.9
92	Number	2	3	5
	Mean Baseline Score	11.5	16.3	14.4
93	Number	106	71	177
	Mean Baseline Score	13.9	15.4	14.5
93J	Number	30	20	50
	Mean Baseline Score	14.6	14.7	14.6
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6L1d.2 Y8L1d mean baseline score per code

Y8L1d		Girls	Boys	All
11	Number	378	376	754
	%	28	29	28
13	Number	294	305	599
	%	22	23	22
21	Number	194	197	391
	%	14	15	15
22	Number	101	62	163
	%	7	5	6
23	Number	3	2	5
	%	0	0	0
24	Number	37	47	84
	%	3	4	3
41	Number	11	5	16
	%	1	0	1
42	Number	7	6	13
	%	1	0	0
43	Number	27	31	58
	%	2	2	2
44	Number	76	79	155
	%	6	6	6
50	Number	2	1	3
	%	0	0	0
91	Number	92	98	190
	%	7	8	7
92	Number	2	3	5
	%	0	0	0
93	Number	136	91	227
	%	10	7	9
Total	Number	1360	1303	2663

Table 6L1d.3 Y8L1d main-code frequencies (number and percent)

Y8L1d		Girls	Boys	All
11	Number	378	376	754
	Mean Baseline Score	14.9	15.7	15.3
13	Number	294	305	599
	Mean Baseline Score	14.4	15.3	14.9
21	Number	194	197	391
	Mean Baseline Score	14.6	15.9	15.2
22	Number	101	62	163
	Mean Baseline Score	16.0	16.9	16.3
23	Number	3	2	5
	Mean Baseline Score	16.7	15.0	16.0
24	Number	37	47	84
	Mean Baseline Score	16.3	16.0	16.1
41	Number	11	5	16
	Mean Baseline Score	15.5	17.4	16.1
42	Number	7	6	13
	Mean Baseline Score	16.4	17.2	16.8
43	Number	27	31	58
	Mean Baseline Score	13.3	15.2	14.3
44	Number	76	79	155
	Mean Baseline Score	18.3	18.8	18.5
50	Number	2	1	3
	Mean Baseline Score	18.0	18.0	18.0
91	Number	92	98	190
	Mean Baseline Score	13.4	14.3	13.9
92	Number	2	3	5
	Mean Baseline Score	11.5	16.3	14.4
93	Number	136	91	227
	Mean Baseline Score	14.0	15.2	14.5
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6L1d.4 Y8L1d main-code frequencies/baseline

Y8L1d		Girls	Boys	All
Correct or incorrect decision; no valid justification (code 1)	Number	672	681	1353
	%	49	52	51
Correct decision; incomplete justification (code 2)	Number	335	308	643
	%	25	24	24
Correct decision; valid, general justification (code 4)	Number	121	121	242
	%	9	9	9
Correct decision; valid justification, general, plus explanation of why justification true (code 5)	Number	2	1	3
	%	0	0	0
Other incorrect (code 9)	Number	230	192	422
	%	17	15	16
Total	Number	1360	1303	2663

Table 6L1d.5 Y8L1d broad-code frequencies

Y8L1d		Girls	Boys	All
Correct or incorrect decision; no valid justification (code 1)	Number	672	681	1353
	%	49	52	51
	Mean Baseline Score	14.7	15.5	15.1
Correct decision; incomplete justification (code 2)	Number	335	308	643
	%	25	24	24
	Mean Baseline Score	15.2	16.1	15.6
Correct decision; valid, general justification (code 4)	Number	121	121	242
	%	9	9	9
	Mean Baseline Score	16.8	17.7	17.3
Correct decision; valid justification, general, plus explanation of why justification true (code 5)	Number	2	1	3
	%	0	0	0
	Mean Baseline Score	18.0	18.0	18.0
Other incorrect (code 9)	Number	230	192	422
	%	17	15	16
	Mean Baseline Score	13.8	14.8	14.2
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3

Table 6L1d.6 Y8L1d broad-code frequencies/baseline

Y8L1d		Girls	Boys	All
0 (code 1) Correct or incorrect decision; no valid justification	Number	672	681	1353
	%	49	52	51
2 (code 2) Correct decision; incomplete justification	Number	335	308	643
	%	25	24	24
3 (code 4,50) Correct decision; valid, general justification	Number	123	122	245
	%	9	9	9
0 (code 9) Other incorrect	Number	230	192	422
	%	17	15	16
Total	Number	1360	1303	2663
	Mean Total Score	0.76	0.75	0.76

Table 6L1d.7 Y8L1d scores

Y8L1d		Girls	Boys	All
0 (code 1) Correct or incorrect decision; no valid justification	Number	672	681	1353
	%	49	52	51
	Mean Baseline Score	14.7	15.5	15.1
2 (code 2) Correct decision; incomplete justification	Number	335	308	643
	%	25	24	24
	Mean Baseline Score	15.2	16.1	15.6
3 (code 4,50) Correct decision; valid, general justification	Number	123	122	245
	%	9	9	9
	Mean Baseline Score	16.9	17.7	17.3
0 (code 9) Other incorrect	Number	230	192	422
	%	17	15	16
	Mean Baseline Score	13.8	14.8	14.2
Total	Number	1360	1303	2663
	Mean Baseline Score	14.8	15.8	15.3
	Mean Total Score	0.76	0.75	0.76

Table 6L1d.8 Y8L1d scores/baseline

7. TEACHER SURVEY

The Year 8 sample of students came from a total of 114 mathematics classes (including the students in mixed attainment groups) and they were taught by a total of 113 teachers. 105 teachers returned the Teacher Questionnaire, but since we do not have the gender of one of these teachers, the tables below are compiled from the remaining 104 questionnaires. The questionnaire is shown in Appendix C.

Table 7.1 shows the number of years of teaching of the teachers in our sample. The sample would seem to be highly experienced with well over half the teachers having taught for more than 10 years. On the other hand, a sizeable minority (over a quarter) have taught for 5 years or less, although Table 7.2 suggests that many of these did not go directly into teaching. Indeed, our sample seems to be quite old. As can also be seen, the sample is quite evenly split between women and men.

Number of Teachers	Years of Teaching							
	0-5	6-10	11-15	16-20	21-25	26-30	Over 30	Grand Total
Women	14	10	6	5	12	3	0	50
Men	16	6	2	5	4	9	7	49
Grand Total	30	16	8	10	16	12	7	99

Table 7.1 Years of Teaching Experience (5 teachers gave no information)

Age of Teachers	Number of Teachers		
	Women	Men	Grand Total
Under 25	3	1	4
25-29	8	5	13
30-39	14	10	24
40-49	16	16	32
50-59	11	18	29
Grand Total	52	50	102

Table 7.2 Age of Teachers (2 teachers gave no information)

Table 7.3 suggests there is quite an even split between women and men for the various categories of school responsibility.

School Responsibility	Number of Teachers		
	Women	Men	Grand Total
Head of Maths	8	9	17
Responsible for KS3	3	2	5
Advanced Skills Teacher	2	4	6
Mentor for Beginning Teachers		1	1
Other	21	20	41
Blank	18	15	33
Grand Total	52	51	103

Table 7.3 School Responsibility (1 teacher gave no information)
('Other' includes: Head of Upper School, Head of Lower School, Head of Year, Deputy Head of Year,, Numeracy Co-ordinator, ICT Coordinator, Liaison with Feeder Schools, Y8 Maths Assessment, Exams Officer)

Tables 7.4 and 7.5 show that 87 teachers in the sample have a BEd or other first degree, with 52 (i.e. about half the sample) majoring in mathematics.

Teaching Qualification	Number of Teachers		
	Women	Men	Grand Total
Maths Degree	23	13	36
Other Degree	12	17	29
No Degree	16	20	36
Grand Total	51	50	101

Table 7.4 Teaching Qualification - Degree other than BEd (3 teachers gave no information)

BEd	Number of Teachers		
	Women	Men	Grand Total
Maths	7	9	16
Other	3	3	6
No BEd	42	39	81
Grand Total	52	51	103

Table 7.5 Teaching Qualification - BEd (1 teacher gave no information)

Tables 7.6 and 7.7 show that about one third of the sample have a PGCE (in most cases in mathematics) and about one third have a Cert Ed (in half the cases in mathematics). None of the sample had a PGCE *and* a Cert Ed.

PGCE	Number of Teachers		
	Women	Men	Grand Total
Maths	18	18	36
Other	3	2	5
No PGCE	31	31	62
Grand Total	52	51	103

Table 7.6 Teaching Qualification - PGCE (1 teacher gave no information)

Cert Ed	Number of Teachers		
	Women	Men	Grand Total
Maths	8	8	16
Other	6	10	16
No Cert Ed	38	33	71
Grand Total	52	51	103

Table 7.7 Teaching Qualification - Cert Ed (1 teacher gave no information)

Nine of the sample had a Diploma, 8 a Masters degree (Tables 7.8 and 7.9). None had a PhD.

Diploma	Number of Teachers		
	Women	Men	Grand Total
Maths	2	2	4
Other	2	3	5
No Diploma	48	46	94
Grand Total	52	51	103

Table 7.8 Higher Education - Diploma (1 teacher gave no information)

Masters	Number of Teachers		
	Women	Men	Grand Total
Maths	2		2
Other	1	5	6
No Higher Degree	49	46	95
Grand Total	52	51	103

Table 7.9 Higher Education - Masters (No PhD) (1 teacher gave no information)

Tables 7.10 and 7.11 show that during 1999/2000 about half the teachers attended the LEA 2-day KS3 National Numeracy Strategy Conference, but that only about one sixth attended NOF ICT training.

National Numeracy Training	Number of Teachers		
	Women	Men	Grand Total
Yes	15	14	29
Blank	37	37	74
Grand Total	52	51	103

Table 7.10 Attendance at LEA's 2-day KS3 National Numeracy Training during 1999/2000
(1 teacher gave no information)

Attendance at New Opportunities Fund	Number of Teachers		
	Women	Men	Grand Total
Yes	11	7	18
No	41	44	85
Grand Total	52	51	103

Table 7.11 Attendance at New Opportunities Fund ICT Training during 1999/2000 (1 teacher gave no information)

Table 7.12 shows that over four fifths of the teachers were involved in some CPD or INSET in mathematics education during a year of their choosing since September 1995 (this can be put more simply as over four fifths of the teachers were involved in some CPD or INSET in mathematics education since September 1995). Tables 7.13 and 7.14 indicate that during the chosen year, slightly more of the sessions were school based than LEA based, although nearly one third of teachers attended 6 or more school based sessions and a similar proportion attended 6 or more LEA based sessions (where a session is a morning, afternoon, twilight or evening).

Attendance at CDP or INSET in Mathematics Education	Number of Teachers		
	Women	Men	Grand Total
Yes	44	41	85
No	8	9	17
Grand Total	52	50	102

Table 7.12 Attendance at CDP or INSET in Mathematics Education during any one year since Sept 1995 (2 teachers gave no information))

Attendance at School based CPD or INSET in Mathematics Education	Number of Teachers		
	Women	Men	Grand Total
0 Sessions	13	12	25
1-5 Sessions	23	19	42
6-10 Sessions	8	11	19
Over 10 Sessions	6	8	12
Grand Total	51	50	101

Table 7.13 Attendance at School based CPD or INSET in Mathematics Education during any one year since Sept 1995 (3 teachers gave no information)

Attendance at LEA based CPD or INSET in Mathematics Education	Number of Teachers		
	Women	Men	Grand Total
0 Sessions	18	14	32
1-5 Sessions	21	20	41
6-10 Sessions	11	13	24
Over 10 Sessions	1	2	3
Over 20 sessions		1	1
Grand Total	51	50	101

Table 7.14 Attendance at LEA based CPD or INSET in Mathematics Education during any one year since Sept 1995 (3 teachers gave no information)

Teachers were also asked about CDP or INSET that was college based, or provided through distance learning, or provided by other means. 26 women and 26 men were involved in none of these other forms of CPD or INSET, another 10 women and 6 men very involved in very little other CPD or INSET (1 - 5 sessions), 12 women and 16 men were involved in a 'moderate' amount of other CPD or INSET and just 4 women and 2 men were involved in a high amount of other CPD (more than 10 sessions in one location).

Table 7.15 shows that about one third of the teachers belong to one (or more) professional association. Specifically, teachers were asked whether they belonged to the ATM, MA or IMA. This information is shown in Table 7.16, where it can be seen that the ATM has 15 members amongst our sample of teachers, the MA has 6 and the IMA one (note that these numbers include teachers that belong to more than one association).

Current Membership of Any Professional Association	Number of Teachers		
	Women	Men	Grand Total
Yes	18	14	32
No	34	36	70
Grand Total	52	50	102

Table 7.15 Current Membership of Any Professional Association (2 teachers refused to answer)

Current Membership of a Professional Association	Number of Teachers		
	Women	Men	Grand Total
ATM	7	3	10
MA	1	1	2
IMA	0	0	0
ATM and MA	2	2	4
ATM and IMA	1	0	1
Other	7	8	15
None	34	36	70
Grand Total	52	50	102

Table 7.16 Current Membership of Specific Professional Associations (2 teachers refused to answer)

Teachers were asked whether they were involved in various extra-curricular mathematics activities during 1999/2000 (in particular, organising a mathematics club, organising students for master classes or the UK Maths Challenge, or taking students to mathematics events). The responses are summarised in Table 7.17, which shows that two third of the teachers are involved in extra-curricular mathematics activities of some kind.

Extra-curricular Mathematics Activities?	Number of Teachers		
	Women	Men	Grand Total
Yes	38	31	69
No	14	21	35
Grand Total	52	52	104

Table 7.17 Involvement in Extra-curricular Mathematics Activities during 1999/2000

Table 7.18 shows that nearly two thirds of teachers used at least one software package with students during the 1999/2000. Teachers were asked specifically about the use of LOGO, dynamic geometry, spreadsheets, databases and the integrated learning system. Spreadsheets were by far the most popular kind of software and were used by about half the teachers, followed by LOGO and databases (Table 7.19). Interestingly, just 5 teachers said they used a dynamic geometry package.

Software Used With Students	Number of Teachers		
	Women	Men	Grand Total
2 or more packages	25	18	43
1 package	12	13	25
None	15	21	36
Grand Total	52	52	104

Table 7.18 Number of Software Packages Used With Students in Current School Year

Specific Software Used With Students	Number of Teachers		
	Women	Men	Grand Total
LOGO	14	15	29
Dynamic Geometry	4	1	5
Spreadsheet	27	25	52
Database	13	8	21
Integ' Learning System	1	4	5
Other	16	7	23
None	15	21	36

Table 7.19 Specific Software Packages Used With Students in Current School Year
(note that some teachers used more than one package)
'Other' includes: SMILE, Omnigraph

Teachers were asked to evaluate, in various ways, the choices presented to students in Questions A3 and G3 from the Proof Survey. Table 7.20 shows the responses to the question, 'Whose answer would your students say would get the best mark from you?'. As can be seen, the vast majority (about 70 %) predicted that students would choose Answer B, which was also overwhelmingly the answer that teachers selected as being closest to what they would have done (Table 7.21), although sizeable minorities chose A for best mark (18 %) and for own approach (16 %).

Question A3: Best Mark	Number of Teachers		
	Women	Men	Grand Total
Answer A	13	6	19
Answer B	32	38	70
Answer C		1	1
Answer A or B	2		2
93	1	1	2
Blank	4	6	7
Grand Total	52	52	104

Table 7.20 Best Mark for Answers to A3

Question A3: Closest Answer to Teacher's	Number of Teachers		
	Women	Men	Grand Total
Answer A	10	7	17
Answer B	36	41	77
Answers A or B	3		3
Blank	3	4	7
Grand Total	52	52	104

Table 7.21 A3 Answer Closest to Teacher's

Teachers were asked to give each answer a mark out of 10. The resulting rankings for all four answers are shown in Table 7.22 (there were no substantial differences between the choices for women and men, so these are not shown separately).

Question A3: Ranking of Answers A, B, C, D	Number of Teachers (Women and Men)			
	A	B	C	D
First	7	97	1	0
Second	93	3	2	3
Third	1	1	10	94
Fourth	0	0	88	4
Blank	3	3	3	3
Grand Total	104	104	104	104

Table 7.22 A3 - Ranking of Answers A, B, C, D

Table 7.23 shows the responses for Question G3 to the question, 'Whose answer would your students say would get the best mark from you?'. As can be seen, the vast majority (about 70 %) predicted that students would choose Answer C, which was also overwhelmingly the answer that teachers selected as being closest to what they would have done (Table 7.24).

Question G3: Best Mark	Number of Teachers		
	Women	Men	Grand Total
Answer A	6	1	7
Answer B	5	3	8
Answer C	35	39	74
Answers A/B	1		1
Other		1	1
Blank	5	8	13
Grand Total	52	52	104

Table 7.23 Best Mark for Answers to A3 (Other:- Comment made by teacher although answer not graded)

Question G3: Closest Answer to Teacher's	Number of Teachers		
	Women	Men	Grand Total
Answer A	2		2
Answer B	5	5	10
Answer C	38	37	75
Answer D		1	1
Other		1	1
None	2	1	3
Blank	5	7	12
Grand Total	52	52	104

Table 7.24 A3 Answer Closest to Teacher's (Other:- Comment made by teacher although answer not graded)

The teachers' rankings for all four answers are shown in Table 7.25 (there were no substantial differences between the choices for women and men, so these are not shown separately).

Question G3: Ranking of Answers A, B, C, D	Number of Teachers (Women and Men)			
	A	B	C	D
First	4	7	96	1
Second	30	81	0	5
Third	61	11	3	7
Fourth	4	0	0	85
Other	1	1	0	1
Blank	4	4	5	5
Grand Total	104	104	104	104

Table 7.25 A3 - Ranking of Answers A, B, C, D

8. SCHOOL SURVEY

The schools in the survey were within 10 LEAs in 9 geographical regions across England. Care was taken to ensure that the LEAs were such as to include rural, semi-rural and urban schools. Within each LEA, schools were selected from randomly ordered lists of maintained secondary schools, stratified for age-range. To avoid choosing very small schools, most of whose students would be likely to have difficulty with the Proof Survey questions, we only considered schools with over 120 fifteen year olds (according to the DfES website). In all, 63 schools participated in the Year 8 survey. The tables below show data taken from the DfES website or from the School Questionnaire that was completed for us by the head of mathematics in each school. The School Questionnaire is shown in Appendix D.

Table 8.1 show that most of our schools are community schools. About three fifths are 11 - 18 schools and two fifths 11 - 16 schools (with one 12 - 18 and one 12 - 16 school). Ten of the schools are specialist schools (Table 8.2).

Type of School	Number of schools		
	11or12 -16	11or12 -18	Grand Total
City Technology College or City College for the Technology of Arts		1	1
Community School	19	28	47
Foundation School	4	1	5
Voluntary Aided School	2	6	8
Voluntary Controlled School	1	1	2
Grand Total	26	37	63

Table 8.1 Type of school (Type A in DfES Classification)
(The sample contains one 12 - 16 school and one 12 - 18 school. The other 61 schools start at age 11)

Special Designation	Age Range of Students		
	11or12 -16	11or12 -18	Grand Total
Designated as an Arts College under the Specialist School Programme		1	1
Designated as a Language College under the Specialist School Programme		2	2
Designated as a Sports College under the Specialist School Programme		1	1
Designated as a Technology College under the Specialist School Programme	1	5	6
Blank	25	28	53
Grand Total	26	37	63

Table 8.2 Special Designation Schools showing Age Range of Students

According to the DfES, one of the 63 schools is classed as selective and all the others are comprehensive, although one of the comprehensive schools claims to have some academic selection, according to the School Questionnaire. Four of the schools are girls' schools (including the selective school) and all the other 59 schools are mixed.

Table 8.3 shows the numbers of schools for various percentage ranges of Year 11 students gaining at least five A* to C grades at GCSE in 1999. The National average is about 50 %, while the median for our sample of schools is 52 %.

		Number of Schools										
% 5 or more A* - C		<10	10-<20	20-<30	30-<40	40-<50	50-<60	60-<70	70-<80	80-<90	90-100	Grand Total
School Age Range	11or12-16	0	1	1	3	5	5	9	2	0	0	26
	11or12-18	0	1	3	3	10	10	6	1	2	1	37
Total		0	2	4	6	16	15	14	3	2	1	63

Table 8.3 Percentage of Students who gained five or more GCSEs at Grades A* to C in 1999

In line with the previous project, schools were asked whether they regarded themselves as urban, rural or suburban. As can be seen from Table 8.4, the three types are fairly evenly represented, with perhaps a slight under-representation of urban schools.

Area	Number of Schools		
	11or12 -16	11or12 -18	Grand Total
Urban	5	10	15
Rural	8	10	18
Suburban	8	10	18
Other*	1	2	3
Blank	4	5	9
Grand Total	26	37	63

Table 8.4 Type of School by Area (*Two schools classed themselves as urban/rural and one as urban/rural/suburban)

Schools were asked, in an open-response format, whether they had special status, such as being designated a technology school or beacon school. The responses are shown in Table 8.5.

Special Status of School	Number of Schools		
	11or12 -16	11or12 -18	Grand Total
Beacon	2	2	4
EAZ	1		1
Gifted & talented		1	1
Home	1		1
Language		1	1
Language College		1	1
Performing Arts School		1	1
Sports and Community College		1	1
Technology		2	2
Technology College		3	3
Technology School	1	1	2
Voluntary Aided		1	1
Blank	21	23	44
Grand Total	26	37	63

Table 8.5 Special Status of School

Schools were asked for the approximate number of Year 8 students. As can be seen from Table 8.6, the most common range was 150 - 200 students, which also contains the median school. One school had only 112 Year 8 students, even though only schools with more than 120 fifteen year olds were considered for the sample. (Interestingly, this was our one selective school, and it is possible other selective schools might have been excluded by the imposition of this size restriction.)

		Number of Schools						
Number of Y8 Students		100 - 150	>150 - 200	>200 - 250	>250 - 300	>300 - 350	Blank	Grand Total
School Age Range	11or12 -16	4	11	8	1		2	26
	11or12 -18	8	14	8	4	3		37
Grand Total		12	25	16	5	3	2	63

Table 8.6 Distribution of total number of Y8 students across schools

The School Questionniare asked whether the Y8 mathematics classes were set, banded, or organised in mixed ability groups. The results are shown in Table 8.7. These responses are similar to, but do not match exactly, the information gathered, mostly in telephone conversations with the head of mathematics, when the schools were first invited to participate in the project. In particular, the table suggests that only 3 schools have mixed ability groups, whereas we had thought there were 6.

Organisation of Y8 Classes	School Age Range		
	11or12 -16	11or12 -18	Grand Total
Set	22	23	46
Banded	3	9	12
Mixed Ability		3	3
Set & Banded	1	1	2
Blank		1	1
Grand Total	26	37	63

Table 8.7 Organisation of Current Y8 Classes

For the students taking the Proof Survey, the School Questionnaire asked for an estimate of the approximate percentage who would be taking the Level 6 - 8 Key Stage 3 test in the following year (ie in May 2001). As can be seen there is a wide variation between schools, with a distribution that is far from even. One might have expected a greater uniformity, given that only high attaining students from each school were participating in the survey.

		Number of Schools											
% Y8 in L6-8 KS3 Test		0-<10	10-<20	20-<30	30-<40	40-<50	50-<60	60-<70	70-<80	80-<90	90-100	Blank	Grand Total
School Age Range	11or12 -16	6	2	4	1	1	3	2	0	0	6	1	26
	11or12 -18	7	4	9	2	3	2	0	1	0	6	3	37
Grand Total		13	6	13	3	4	5	2	1	0	12	4	63

Table 8.8 Approximate Percentage of participating Y8 Students Predicted to take Level 6 - 8 KS3 Test

Table 8.9 shows which GCSE examination syllabus the school is currently using. As can be seen, the use of OCR/SMP is relatively low, and it would be interesting to find out whether this is a stable phenomenon, or whether it represents a decline in line with the recent fall in the number of schools using SMP.

GCSE Examination Syllabus	Number of Schools		
	11or12 -16	11or12 -18	Grand Total
EdExcel	10	18	28
OCR/SMP	6	5	11
SEG/NEAB	10	14	24
Grand Total	26	37	63

Table 8.9 GCSE Examination Syllabus currently used in the school

Table 8.10 shows the scheme or textbook used in the schools in Year 8. What is striking is the success, in terms of uptake, of Key Maths.

Textbook Used in Year 8	Number of Schools		
	11or12 -16	11or12 -18	Grand Total
Key Maths 8	15	16	31
SMP 11-16 Booklets	2	7	9
Vickers New National Curriculum Maths	4	1	5
STP 8A		3	3
Other*	5	10	15
Grand Total	26	37	63

Table 8.10 Textbook Used in Year 8

(*Other includes: Maths In Action - Nelson Blackie / Heinemann Maths / Understanding Maths / SMILE Independent Learning Scheme / Nelson - Essential Skills / 'In-house Materials' / 'Variety or Sources' / Based on SMILE / 'Non Textbook Driven - Key Maths only a source'/ 'Various including NMP Maths for Secondary School')

Table 8.11 shows that the majority of schools have 180 minutes of mathematics lessons per week, which is also the median time. This quite closely matches the recent DfES guidelines suggesting that schools should have 4 lots of 45 minute or 5 lots of 40 minute lessons of mathematics per week.

		Number of Schools							
Number of minutes of maths lessons per week		140	150	175	180	200	210	225	Grand Total
School Age Range	11or12 -18		4	3	15	2	2		26
	11or12 -18	1	4	2	19	6	3	2	37
Grand Total		1	8	5	34	8	5	2	63

Table 8.11 Total Duration (in minutes) of Y8 Mathematics Lessons per Week

Table 8.12 shows that in two thirds of the schools at least some Year 8 students were involved in extra-curricular mathematics activities, be this in school or elsewhere.

Are any Y8 Students involved in Extra-curricular Activities?	Number of Schools		
	11or12 -16	11or12 -18	Grand Total
Yes	20	24	44
No	6	13	19
Grand Total	26	37	63

Table 8.12 Y8 Students Engaged in Extra-curricular Activities

9. MULTILEVEL ANALYSIS

The summary chapter of the Y8 multilevel analysis is given below, the complete report is given in Appendix Z.

Summary

Total scores

In sections 2 and 3 we modelled total scores for Geometry and Algebra constructive proof, and for Geometry and Algebra validity rating (G3c and A3c, respectively), as multivariate responses. The scores for the 'Logic' question , L1, were included in the total for Algebra constructive proof.
The student's score on the baseline test was a statistically significant predictor for all four scores. For constructive proof, we found (Model 0) that without adjustment for baseline score there was no statistically significant effect of gender. When baseline score was included in the model (as it was in all other models), there was in addition a statistically significant gender effect, in favour of girls, except in the case of Geometry VR score (see Model 6). This gender effect, where present, was not itself dependent on the student's baseline score, in other words, there was no interaction between gender and baseline score. Girls tended to perform less well than boys on the baseline test: for example, the median baseline score for girls is 15 and for boys is 16. The gender effect, where present, was comparable in size and of opposite sign to the effect of this one-point difference in baseline score. In Geometry VR score, where no statistically significant gender effect was found, the effect of a one-point difference in baseline score was at most one-tenth of a raw-score point. See the table on page 45 for more detail.
We may summarise this finding by saying that girls tended to perform better on these proof tests, relatively to their performance on the baseline test, than boys.
Also in Model 0, we found slight but statistically significant variation in constructive proof scores between classes within schools, amounting to some 3 to 4 per cent of the total variance (between-school variance accounted for five times as much as this). Once baseline score was adjusted for, variation between classes within schools ceased to be statistically significant. Not could we find any teacher-level variables with statistically significant effects.

The only variables not already part of the proof test that were found to have a statistically significant effect on total scores, for either constructive proof or validity rating, were:

School level
- %A*-C
- Use of Textbook 2

Class level
- (none)

Student level
- Gender
- Baseline score

In particular, whether a school was 11–18 or girls-only had no statistically significant effect. Textbook 2 was found to be beneficial for Algebra constructive scores and for Geometry VR scores. The school's %A*-C had a positive effect on all four scores.

After adjustment for student gender, baseline score, and the school's %A*-C, we found that schools varied in their effectiveness, except in the case of Geometry VR score, for which no school-level variation was detected. Schools' residual performance in Algebra constructive proof was significantly more variable for boys than for girls once the school's %A*-C was adjusted for. Correlations at school level between girls' and boys' residual performance were high ($r \cup 0.9$) within Algebra and Geometry constructive proof and appreciable ($r > 0.7$) for Algebra VR, but correlations across these subjects were more modest (generally, $r < 0.6$). Thus, in a school whose girls did better than predicted in Algebra the boys tended to do better in Algebra also; but neither the girls nor the boys in that school would be especially likely, as a group, to do better than predicted in Geometry. This is demonstrated by the rankings of schools for Algebra and Geometry in Models 1 and 2.

We found no evidence of differences in school effectiveness for students with different baseline scores.

Turning to the relative sizes of the different effects, we need to be careful not to over-interpret apparent differences. All the estimates are subject to error. With this caveat, we found from Model 7 that the gender effect on total Algebra constructive score was about half the effect of using textbook 2. The effect on this score of the school's %A*-C, compared to the average, was less than the gender effect for more than 70% of students. The remaining 30% of students attended schools whose %A*-C departed so far from the mean that its effect equalled or exceeded the gender effect. By contrast, the effect of baseline score for the top 15% and the bottom 15% of students was of a higher order of magnitude than the gender effect. Residual school effects on the total Algebra constructive score for girls were generally smaller in size than the gender effect itself, though schools at the extremes exceeded this. For boys, as we have noted, the school effects tended to be larger. See the tables on pages 43 and 45 for more detail and for the effect sizes for the other scores.

Individual scores for constructive proof

These were modelled in section 4, and no attempt was made to rank schools on these very limited outcomes. The main purpose was to study correlations between scores on different questions and whether any additional statistically significant effects could be found. We consider the Geometry questions first.

Student baseline score and school %A*-C were significant predictors for all individual Geometry scores. Gender was significant for all except G2a, but *negative* for G1. Geometry VR score was significant for all except G2b. The response coded 30 to question L1b was interpreted as an indicator that the student appreciated proof as general. This indicator had a statistically significant positive effect on the score for G4 (only). Use of textbook 2 was found to have a statistically significantly positive effect on G1 score (only).

We could find no statistically significant school-level residual variation in the scores for G2a or G2b. There was high residual correlation at school level between girls' and boys' scores for G1 and between girls' and boys' scores for G4; also between boys' scores for G1 and G4, but not between girls' scores for the separate questions. Correlations at student level were all low. An individual student who scores highly on one of these questions, compared to expectation and after adjusting for any school residual effect, is not especially likely to score highly on another.

We now consider the separate Algebra questions A1, A2, and A4, together with the Logic question, L1. Student baseline score and school %A*-C were significant predictors for all scores. Gender was significant, and positive, for A2 and A4 only. Algebra VR score was significant for A4 and L1 only. 'Proof-as-general' was significant for A2 and A4. Use of textbook 2 was significant for A1 and A2 only.

With these fixed effects in the model, statistically significant residual correlation at school level between girls' and boys' scores was present for question A4 only. Within gender, there appeared to be high school-level correlation between A1 and A2 scores, and between A2 and L1 scores. At student level, correlations were once again low.

Where these findings differ from those for the total constructive scores they are not easy to interpret. And all should be treated with caution in view of the distributional characteristics of the individual scores. See pages 48, 53, and 54.

Choice of own approach in multiple choice questions

Probabilities of different choices for the student's own approach in these two questions, as indicated by their answers to G3a and A3a, were analysed in sections 5 and 6. No class-level residual variation was found, and no statistically significant class-level or teacher effects. In particular, the teacher's choice for their own approach was not significant.

The following variables, not already part of the proof test, were found to have a statistically significant effect on at least one of these probabilities:

School level

More than 3 hours of maths per week

Student level

Gender

Baseline score

Note that neither school %A*-C nor use of a particular textbook nor GCSE syllabus was significant. While the amount of maths taught per week was significant, the existence of a maths club was not.

Models of two types were considered. The first type was for school comparisons, and excluded all variables apart from those tabulated above. The second type was designed to explore how the outcome probabilities were associated with other aspects of the student's response to the proof test.

In a model to compare school effects, the number of hours of maths per week should not be included: although significant, it is, like choice of textbook, within the school's control. This leaves only gender and baseline score.

We consider G3 first. Of the four possible choices, A was termed 'empirical', and was used as the base category with which to compare the probabilities of each of the other three choices. Choice C was a correct formal proof and was objectively the best approach. Interestingly, baseline score had *no* statistically significant effect on the probabilities of choices C or D: both were significantly less likely than choice A. Baseline score had a significant negative effect on the probability of choice B: thus, this choice became less likely, the higher the baseline score. Given a baseline score, girls were less likely than boys to make choice B, and they were always less likely than boys to choose C or D.

We found no significant school-level residual variance in the gender effect. There was very high correlation ($r > 0.9$) at school level between choices B and C, in other words, a school with a higher than predicted proportion of students opting for choice B would tend also to have a higher than predicted proportion opting for choice C. Correlations for the other pairs of choices (B,D and C,D) were only moderate ($r < 0.6$).

Following Healy and Hoyles (1999), we then explored the effects on the probabilities of choosing B, C, and D of the student's view of that proof's explanatory power and validity, and whether it would gain the best mark from the teacher. We also allowed total Geometry VR score and 'more than 3 hours of maths per week' as possible predictors. We found no significant effect of scores in Geometry constructive proof. For each choice, B, C, or D, we found that thinking a proof would gain the best mark predisposed a student to choose it for their own approach, as did thinking it explained *why* the result was true (or false). A correct assessment of the proof's validity (as shown by the student's response to question G3c) made choices B and D significantly less likely. (These proofs are not valid.) A correct assessment of validity for choice C (the best proof) made it more likely, but more significant for the probability of this choice was the total VR score in Geometry (i.e. the total score for G3c). Being in a school that offered more than three hours of maths a week also had a statistically significant positive effect on the probability of choosing option C. This effect of numbers of hours of teaching, which was confined to the probability of option C, was present in other models not described in this report, and was the only school-level fixed effect found to be statistically significant for this outcome.

We now consider question A3 – the Algebra multiple choice question. Again, there were four possible choices, A to D, and we termed choice A 'empirical' and used it as the base. The best choice (in the sense of most complete and valid) was option B. Baseline score had no statistically significant effect on the likelihood of choosing this proof, which boys were as likely to choose as option A. Girls, whatever their baseline score, were significantly less likely than boys to choose option B. The likelihood of choices C and D diminished with increasing baseline score and was less for girls than for boys with a given baseline score.

After adjustment for gender and baseline score, there was little residual variation at school level, and this only in the probability of choosing option B. Only one school – number 62 – was significantly above expectation in the proportion of its students choosing this option.

A correct assessment of the validity of option B increased the probability of the student's choosing it for their own approach. Correct assessments of validity for options C and D (which are invalid) reduced the probability of their choice. Thinking that a proof, whether B, C, or D, would gain the best mark from the teacher increased the likelihood of its choice as the student's own approach. The likelihood of choosing option B appeared to be unaffected by the student's opinion of that proof's explanatory power, while a positive view of the explanatory power of either proof C or proof D increased the likelihood that the student would choose it for their own approach. As was the case with the objectively best choice for Geometry, we found that more than three hours of maths per week in the school had a statistically significant and positive effect on choice B for Algebra. A high total VR score in Algebra also was associated with increased likelihood of this choice as the student's own approach. In the presence of these rational effects, the effects of Algebra constructive scores proved to be difficult to interpret. These were, therefore, omitted.

Choice for best mark in multiple choice questions

Probabilities of different choices for best mark for these two questions, as indicated by students' answers to G3b and A3b, were analysed in sections 7 and 8. As in the previous analyses, choice A was in each case the base. As before, no class-level residual variation was found, and no statistically significant teacher effects. In particular, the teacher's choice for their own approach was not significant for student choices for best mark. Nor was teacher's opinion of what students would choose for best mark.

The variables that were not already part of the proof test and were found to have a statistically significant effect on at least one of the probabilities in this group were:

School level

Use of Textbook 2

Student level

Gender.

Thus, there was no effect of baseline score.

We were not concerned to compare schools on these outcome probabilities, so we proceeded directly to explore associations with other student responses to the proof test, as well as the variables above. We excluded student choice for own approach as a predictor, as the previous analyses had already demonstrated that the two probabilities were related, and it is more reasonable to suppose that the student's view of what would gain the best mark should predict the student's choice for their own approach, rather than the other way round.

Considering Geometry first, we found that the probability of choosing proof B, C, or D in preference to A for best mark was increased if the student considered that proof to explain the result. A correct assessment of the validity of the proof increased the likelihood of its choice for proof C (the best proof) and decreased it for proofs B and D. A high total VR score for Geometry further increased the probability of choosing proof C for best mark. Interestingly, use of textbook 2 also increased the probability of choosing proof C. In fact, there was a preference for option C over option A for best mark even among those who did not correctly judge its validity and had poor overall VR scores. Girls were less likely than boys to choose proofs C or D, conditionally on the other variables.

All of these effects (apart from the gender effect) are relatively easy to understand. By contrast, the effect of constructive proof scores in Geometry was not easy to interpret. High constructive scores were predicted to make each of choices B, C, and D more likely for best mark. The effect on choice C, however, only just reached statistical significance and had the smallest magnitude. It seems most reasonable to ignore these apparent effects.

In Algebra, proof D was by far the least likely to be chosen for best mark, and its likelihood was further decreased by a correct assessment of its validity. Girls were less likely than boys to make this choice, as they were also to make choice C. Proof B, objectively the best, was generally most likely to be chosen for best mark, and this probability was further increased if the student correctly judged its validity or felt it had explanatory power. A high total VR score in Algebra further increased the probability of this choice for best mark. Proof C also was made more likely if the student felt it had explanatory power, but less likely if the student correctly judged its validity. The effect of constructive score in Algebra was not of interest: a high score was predicted to make option D even less likely, but the effect was otherwise non-significant.

References

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

Healy, L. and Hoyles, C. (1999) Justifying and Proving in School Mathematics: Technical report on the nationwide survey (London: Institute of Education, June)