In *Educational Studies in Mathematics*, (2002) 51, 3, 193-223.

Students' Understandings of Logical Implication

Celia Hoyles and Dietmar Küchemann

School of Mathematics, Science and Technology,

Institute of Education,

University of London,

Bedford Way,

London WC1H 0AL, UK.

Telephone: +44 (0)20 7612 6659   Fax: +44 (0)20 7612 6686

E-mail: c.hoyles@ioe.ac.uk

Abstract

We report results from an analysis of responses to a written question in which high-attaining students in English schools, who formed part of a longitudinal nation-wide survey on proof conceptions, were asked to assess the equivalence of two statements about elementary number theory, one a logical implication and the other its converse, to evaluate the truth of the statements and to justify their conclusions. We present an overview of responses at the end of Year 8 (age 13 years) and an analysis of the approaches taken, and follow this with an analysis of the data collected from students who answered the question again in Year 9 (age 14 years) in order to distinguish learning trajectories. From these analyses, we distinguished three strategies, empirical, focussed-empirical and focussed-deductive, that represent shifts in attention from an inductive to a deductive approach. We noted some progress from

Year 8 to Year 9 in the use of the focussed strategies but this was modest at best. The most marked progress was in recognition of the logical necessity of a conclusion of an implication when the antecedent was assumed to be true.

Finally we present some theoretical categories to capture different types of meanings students assign to logical implication and the rationale underpinning these meanings. The categories distinguish responses where a statement of logical implication is (or is not) interpreted as equivalent to its converse, where the antecedent and consequent are (or are not) seen as interchangeable, and where conclusions are (or are not) influenced by specific data.

KEY WORDS

proof, logic, logical implication, deduction

In *Educational Studies in Mathematics*, (2002) 51, 3, 193-223.

Introduction

A fundamental objective of mathematics education must be to help students recognise and construct mathematical arguments, that is to engage in the process of proving. Hanna (2000) lists eight different functions of proof and proving, of which she regards verification and explanation as the most fundamental. When it comes to the mathematics classroom, Hanna places particular emphasis on explanation, which is surely right. However, verification and explanation clearly interact and our interest in this paper is on the former, and specifically the role in verification played by logical implication. Rodd (2000) argues that logical implication in the form of modus ponens reasoning ($p \Rightarrow q$, p so q), is one of the most basic structures for establishing a mathematical truth. Logical implication is also central to school mathematics, and to what Sowder and Harel (1998) call analytic proof schemes.

Deductive proofs of this type have the potential to be transparent, since "the validity of the conclusion flows from the proof itself, not from any external authority" (Hanna, 1995, p.46). Yet it is well known in mathematics education that most school students do not find the deductive process straightforward and tend to use inductive reasoning[i] to validate conjectures in mathematics rather than to prove them deductively (e.g., Bell, 1976; Van Dormolen, 1977; Balacheff, 1988). Even when students seem to understand the function of proof in the mathematics classroom (e.g., Hanna, 1989; de Villiers, 1990; Godino & Recio, 1997) and to recognise that proofs must be general, they still frequently fail to employ an accepted method of proving to convince themselves of the truth of a new conjecture, preferring instead to rely on pragmatic methods and more data (e.g., Fischbein, 1982; Vinner, 1983; Coe & Ruthven, 1994; Rodd, 2000; Simon, 2000). In the Fischbein study, for example, it was reported that only a minority of students judged that empirical checks would not increase their confidence in a proof that had already been accepted as valid and general. Fischbein argued that this seemingly contradictory behaviour was due to the fact that while students were being asked about a mathematical proof, their experience was mostly with empirical proof: "This (mathematical) way of thinking, knowing and proving, basically contradicts the practical adaptive way of knowing which is permanently in search of additional confirmation" (ibid, p.17). How students learn to move between mathematical ways of proving and those that are rooted in everyday thinking is at the heart of our study.

Since the 1980s in the U.K., there has been a shift of emphasis in the curriculum from geometry to simple number theory (see Hoyles, 1997). This shift has had two consequences. First, rather little attention has been paid to the mathematical relationships and structures in the number patterns generated and investigated in the classroom (see for example, Hewitt, 1992); and second, students have had few opportunities to engage with the deductive process in general and with the structure of a logical implication in particular [ii].

In a recent study of student proof conceptions (Healy & Hoyles, 2000), it was reported that, in line with previous research, most U.K. students were aware that a valid proof must be general. Additionally, when asked to assume a theorem was true, and then to consider whether the theorem would hold in a more restricted domain (e.g., for square numbers, when it was assumed to hold for whole numbers), most students did not feel that they needed more data, a finding which to some extent contrasts with that from the Fischbein study, mentioned earlier[iii]. Healy & Hoyles also found that this recognition of the general applicability and logical necessity of the conclusion of a theorem was a significant predictor of students' overall performance on their proof test.

Another strand of research in the area of proving in school mathematics has shed light on how the processes of explanation, justification and even logical necessity can be fostered by teachers (see Kieran, Forman & Sfard, 2001; Yackel & Cobb, 1996). Yackel (2001) in particular took Toulmin's scheme (Toulmin, 1958) comprising conclusion, data, warrant and backing, as elaborated for mathematics education by Krummheuer (1995), to analyse interactive argumentation. Though a written question is far removed from the kind of interaction reported by Yackel, we also have adopted Toulmin's scheme as part of our analysis.

As background to the research reported here, we first briefly turn to review studies into students' appreciation of logical relations and investigations of the logical properties of students' arguments.

Background

In many areas of mathematics education there has been considerable research into how students come to understand fundamental mathematical structures and relationships, for example function and proportion. By contrast, there has been rather little research into students' understandings of the structure of logical relations, and in particular of logical implication. Research into children's understanding of logical reasoning has largely been undertaken in the field of developmental psychology, originating with the seminal work of Inhelder and Piaget (1958), who used propositional calculus as a basis for their analysis of children's responses: for example, they argued that understanding an implication $p \Rightarrow q$ required an appreciation of its equivalence to the appropriate four combinations of truth values of the two propositions p and q, where p is the antecedent and q the consequent.

In considering this formal structure of implication, it is important to recognise that the significance of combinations of truth values differs according to whether one is using an implication to draw conclusions (e.g., O'Brien, Shapiro and Reali, 1971), or trying to determine whether an implication is true (e.g., Wason, 1960). Studies into children's understanding of implication have usually focussed on the former situation, and have generally found that children (and indeed adults) have difficulty in understanding that implication includes a consideration of the case when the antecedent is false. A further, perhaps more pertinent, result from this corpus of research is that children tend to treat a conditional statement and its converse as equivalent, a phenomenon described as 'child logic' (O'Brien et al 1971)[iv].

The items used in our study are closer to Wason's selection task (see for example, Wason & Shapiro, 1971) than to the items used by O'Brien, since we ask students to determine whether or not various given implications are true. However, structurally our items are simpler than Wason's classic task as well as being in a different (and non-arbitrary) domain. Additionally, since the 1970s, evidence has accumulated that children rarely argue solely on the basis of universal formal laws of logic or domain-independent abstract rules. There are different interpretations as to why this might be the case: that children call up pragmatic reasoning structures derived from experience in context (Cheng & Holyoak, 1985), or that their arguments are derived from knowledge and the way it is structured (Ceci, 1990). Anderson, Chinn, Chang, Waggoner and Yi (1997) found that in naturally occurring arguments children tended to omit parts of the logic of their deductions, to be cryptic when mentioning the known or obvious, and elliptical when expressing their position, giving no more information than was necessary. Nonetheless, Anderson et al concluded that the children's arguments were logically complete: that is, the framework they used was not inconsistent with formal logic.

Anderson et al's analysis focussed on what they termed informal deduction. We follow this approach and, in contrast to Wason and to Inhelder and Piaget, we are not concerned with the strict requirements of formal logic. Rather we draw on an important distinction made for example by Mitchell (1962), between *material implication*, represented by "p⇒q", which is part of propositional logic, and *hypothetical proposition,* represented by "if p, q" that "asserts only what is the case if its antecedent is realised" (Mitchell, 1962, p64)[v]. We claim that when studying reasoning in school mathematics, the latter (hypothetical proposition) is a more appropriate interpretation of logical implication than the former (material implication), since in school mathematics, students have to appreciate the consequence of an implication when the antecedent is taken to be true[vi] (see also Deloustal-Jorrand, 2002).

There is some evidence from cross-sectional studies of logical thinking that the use of child logic decreases with age (O'Brien et al, 1971). Similarly, some researchers have postulated a developmental hierarchy of mathematical justification, "from inductive (empirical) reasoning toward deductive reasoning and toward a greater level of generality" (Simon, 2000), though clear evidence for this seems to be lacking. More specifically, there has not been any systematic study of students' understandings of logical implication in the context of school mathematics, or how these understandings might develop and change over time – two surprising gaps in the corpus of research knowledge given the importance of deduction in school mathematics. The study reported aims to throw light on both of these under-researched areas.

Methodology

The findings reported in this paper arise from The Longitudinal Proof Project, a study that aims to describe students' learning trajectories in mathematical reasoning over time. The particular focus of this paper is an investigation of the following three (interrelated) research questions concerning students' meanings of logical implication where the reference knowledge is simple number theory.

1. How do students who are not taught about the structural meaning of logical implication determine whether a statement of logical implication is true or not and do their approaches change over time?
2. Are students aware of the general applicability and logical necessity of the consequent in a statement of logical implication if both the statement and the antecedent are true?
3. How do students conceptualise the relationship of logical implication and its converse and does this conceptualisation change over time?

Sample and research instruments

In the Longitudinal Proof Project, data are collected through annual surveying of students in the highest attaining class (or classes) of randomly selected schools within nine geographically diverse English regions. The survey instruments were specially devised written proof tests comprising a range of questions designed to probe different aspects of proving in the domains of number/algebra and geometry. The first written test (the Yr 8 test) was administered to 2663 students in 63 schools in

June 2000 when the students were approaching the end of Year 8 (age 13 years). The same students (with some inevitable dropout) were tested again in the summer term of 2001 using a new instrument (the Yr 9 test) that was designed to trace developments in mathematical reasoning as well as to build cross-sectional profiles of student understandings at a particular age and place of schooling. The Yr 9 test includes some questions that are identical to those in the Yr 8 test, some that are slightly modified and some that are new. The Yr 8 and Yr 9 tests were both administered in parallel versions, A and B, with the same questions presented in a different order. The same students have been tested again in June 2002 with the similar aim of testing both understandings and development.

All the questions in the proof tests were designed in collaboration with mathematics teachers in five design schools and extensively piloted in these schools. During the pilot stage of each proof test, a detailed set of codes was drawn up and validated, partly on the basis of theoretical distinctions in proving to which we wished to pay attention, such as generalisations in algebra, visual argument in geometry, and partly on the basis of the student responses. All the scripts were coded by one of the authors and one other coder and the consistency and reliability of the coding outcomes regularly checked[vii]. The code frequencies for both Year 8 and Year 9 responses have been analysed using descriptive statistics[viii] and a small number of students (and their teachers) interviewed to probe responses in more detail and to assist in our interpretation of the analyses.

In this paper, we present analyses of students' responses to one question, L1, involving logical implication, which was one of nine in the 50-minute Year 8 proof test, and appeared again as one of ten in the 55-minute Year 9 test (with the names of Joe and Fred changed to Pam and Viv).

A question like L1 would not be familiar to students in England, although they would have come across odd and even numbers. L1 has several parts that aim to find out how students conceive of the structure of logical implication, how they attempt to determine the truth of a statement that comprises a logical implication, and how the latter process might interact with the former. Analyses of cross-sectional responses to the question as a whole provide a snapshot of students' reasoning at a given point in time. Additionally, analyses of the same students' responses over time provide a picture of how reasoning develops, when implication is used as part of mathematical argumentation in classrooms but its structure not made explicit.

The students are presented with two statements[ix], an implication and its converse, the first of which is false and the other true (see Figure 1). Students are asked to decide whether the two statements are saying the same thing, to make a conclusion assuming one of the statements is true, and to evaluate the truth of each statement in turn. Regarding these evaluations, we note that students are asked to determine whether a given statement (which is expressed as an implication, $p \Rightarrow q$) is true. Given our earlier discussion of deductive reasoning in school mathematics, the crucial combinations under investigation are those where the antecedent is true, and the consequent either true or false[x]. It is anticipated that how students approach these evaluations will shed light on our other areas of interest, namely, the prior knowledge and the evidence generated within the question that students use to decide whether a statement and its converse are saying the same thing, and second, the extent to which students are able and willing to use a result to deduce a consequence.

Insert Fig 1 about here

Fig 1: Question L1, which concerns a statement of logical implication and its converse

Results

We present the results, first through an analysis of Year 8 student responses to each part of question L1, and second through comparisons with the responses of some of the same students to the same question one year later. For the Year 8 analysis we report on the students (N = 2663) who took both the Yr 8 test and a 'baseline maths test' administered a few weeks earlier[xi].

Year 8 students' responses to L1a

We decided to ask students at the outset whether they thought Joe's and Fred's claims were the same ("... saying the same thing"). Clearly, responses to this might not be carefully thought out, and we expected that some students would change their answer after working through the later parts of the question, where there is further opportunity to think about the truth of each statement (Joe's in L1c, Fred's in L1d). In coding students' answers to L1a, therefore, we distinguished between a straightforward 'No' and an answer that had clearly been changed from 'Yes' to 'No'.

The frequency distribution of responses is given in Table I.

Item	Response code	Year 8 responses Frequency %	Number
L1a	code 10: Incorrect (Yes)	71	1897

code 31: Correct (Yes changed to No)	15	387
code 32: Correct (No)	13	355
code 9: Miscellaneous incorrect (including no response)	1	24

Table I: Distribution of Year 8 responses to L1a (N = 2663)

The table shows that 13% of students correctly stated from the outset that Joe's and Fred's statements were not saying the same thing, with a further 15% changing their answer from 'Yes' to 'No' at some stage. The vast majority of students, 71%, stated that the statements were saying the same thing.

Year 8 students' responses to L1b

To investigate how far students are aware of the general applicability and logical necessity associated with a statement of logical implication, students were asked in L1b to assume that one of the statements (Fred's) is true, and then to decide whether they could deduce a result from this assumption, or whether the result could still only be verified by induction (i.e., on the basis of further data)[xii].

In L1b, students were told that the product of two whole numbers is 1271. This large and rather obscure number was chosen to discourage students from trying to find the possible values for the whole numbers (that is, 31 and 41, and 1 and 1271). Just under half the students in the sample (47%) chose the correct option (that the sum must be even) with another 47% choosing the empirical option (you can't be sure until you know what the numbers are), as shown in Table II. Clearly, some students may have come to the correct conclusion by reference to properties of data rather than by a direct and general deduction (for example reasoning that two numbers whose product is 1271 must both be odd so their sum is even, or, indeed, that the numbers could be 1 and 1271 in which case the particular sum, 1272, is even). Nonetheless, it seems safe to assume that roughly half of the sample could play the mathematical game of supposing a statement is true, whether it is or not, and then making a correct deduction on that basis.

Item	Response code	Year 8 responses Frequency %	Number
L1b	code 10: Empirical ("Can't be sure until you know what the numbers are")	47	1241
	code 30: Deduction ("Sum is even")	47	1249
	code 9: Miscellaneous incorrect (including no response)	6	173

Table II: Distribution of Year 8 responses to L1b (N = 2663)

Year 8 students' responses to L1c: (Sum-even $\Rightarrow$ Product-odd)

L1c asks whether Joe's statement is in fact true. The question can be answered correctly (namely that the statement is false) by means of a counter example. As Table III shows, 26% of the total Year 8 sample gave an invalid response or no justification (code 1) and another 22 % gave responses that were only partially correct (code 2). Of these code 2 responses, most (18% of the total sample) stated that Joe's statement was true and justified this with data that did indeed confirm the statement by picking a pair (or several pairs) of odd numbers (e.g., "Yes, Joe is right because 5+5=10 (even), 5×5=25 (odd)"). Given the prevalence of these responses, they were given a separate code of 22.

Thirty six percent of students correctly stated Joe's statement was false and supported this by reference to the existence of a counter example. Of this 36%, most (28% of the total sample) gave specific counter examples (code 3), and usually just one (e.g., "No, e.g., 2+4=6, but 2×4=8 which is even, so they are both even"). Some students however (8% of the total sample) described the counter example in general terms (code 4), as consisting of even numbers (e.g., "No: even + even = even, but even × even = even also"). A general explanation of this sort is interesting for two reasons. First, it represents a shift from simply looking at data to considering underlying structure: it goes beyond showing *that* the statement is false, though this is all that the item requires, to giving some insight into *why* it is false. Second, a general counter example *might* have been found by deduction rather induction, by considering the mathematical relationships involved in the statement.

Item	Response code	Year 8 responses Frequency %	Number
L1c	code 1: Correct or incorrect decision; invalid or no justification	26	684
	code 2[a]: Correct or incorrect decision; flawed or incomplete justification	22	589
	code 3: Correct decision; valid justification, specific	28	753
	code 4: Correct decision; valid justification, general	8	216
	code 9: Miscellaneous incorrect (including terms not understood and no response)	16	420

Table III: Distribution of Year 8 responses to L1c (Sum-even ⇒ Product-odd) (N = 2663)
[a]The code 2 responses were divided into the following sub codes:
code 21/23: Correct decision; flawed or incomplete justification (4%, 100 students)
code 22: Wrong decision; justified by reference only to confirming cases (18%, 489 students)

Year 8 students' responses to L1d: (Product-odd ⇒ Sum-even)

L1d asks about the truth of Fred's statement. In contrast to L1c, this can only be answered decisively by a deductive argument, for example, "If the product is odd, then the two numbers have to be odd and so the sum will always be even", which would be accepted as appropriate for students of this age.[xiii].

Table IV shows that 24% of students correctly stated that Fred's statement was true and supported this with correct but only empirical examples (code 2). Such responses usually again consisted of just one example (e.g., "Yes, 3×3=9, product odd; 3+3=6, sum even"). A further 9% of students supported their correct evaluation of Fred's statement with a general description of the starting numbers, namely that they must both be odd (code 4). Such explanations were often cryptic so while all code 4 responses moved beyond reference to specific odd starting numbers, it was frequently difficult to distinguish responses that expressed *the necessity* of the starting numbers being odd and of the sum hence being even (i.e. a deductive approach), from responses where the need for odd numbers might have been the result of an inductive generalisation: (e.g., “I could only find odd numbers that made the product odd and sum even”). Finally some students were clearly influenced by their answer to L1c and wrote ‘same as Joe’, (11%, 294 students). Thus care has to be taken in comparing response frequencies between L1c and L1d as some frequencies might well have been different if the order had been reversed.

Item	Response code	Year 8 responses	
		Frequency %	Number
L1d	code 1: Correct or incorrect decision; no valid justification	51	1353
	code 2[b]: Correct decision; incomplete justification, empirical	24	643
	code 4: Correct decision; valid justification, general	9	242
	code 9: Miscellaneous incorrect (including terms not understood and no response)	16	422

Table IV: Distribution of Year 8 responses to L1d (Product-odd ⇒ Sum-even) (N = 2663)

[b]The code 2 responses were divided into the following sub codes:

code 21: Correct decision; confirmation by one empirical example (15%, 391 students)

code 22: Correct decision; confirmation by several empirical examples (6%, 163 students)

code 23: Correct decision; confirmation by one or several empirical examples and awareness that this is not conclusive (0%, 5 students)

code 24: Correct decision; crucial experiment (confirmation by one or several empirical examples with at least one starting number greater than 10) (3%, 84 students)

Comparison of Year 8 and Year 9 Students’ Responses

As mentioned earlier, the same question (with the names of Joe and Fred changed to Pam and Viv) was used again in the Year 9 survey administered in June 2001. Table V shows the code frequencies in Year 8 and Year 9, and their differences, for students' responses to each part of question L1. (These frequencies are based on a smaller number of students (N = 1078) than the frequencies for Tables I to IV (N = 2663)[xiv].

Item	Code	Frequency (%) Yr 8	Yr 9	Yr9 – Yr8[c]
L1a	code 10: Incorrect (Yes)	69	62	-7
	code 31: Correct (Yes changed to No)	16	19	3
	code 32: Correct (No)	14	18	4
	code 9: Miscellaneous incorrect (including no response)	1	1	0
L1b	code 10: Incorrect, empirical ("Can't be sure")	42	23	-19
	code 30: Correct, deduction ("Sum is even")	53	72	19
	code 9: Miscellaneous incorrect (including no response)	5	5	0
L1c	code 1: Correct or incorrect decision; invalid or no justification	25	20	-5
	code 2: Correct or incorrect decision; flawed or incomplete justification	21	19	-2
	code 21/23: Correct decision; flawed justification	3	3	0
	code 22: Wrong decision; flawed justification	18	16	-1
	code 3: Correct decision; valid justification, specific	32	37	5
	code 4: Correct decision; valid justification, general	9	11	2
	code 9: Miscellaneous incorrect (including terms not understood and no response)	13	12	-1
L1d	code 1: Correct or incorrect decision; no valid justification	51	45	-5
	code 2: Correct decision; incomplete justification, empirical	24	27	3
	code 21: Correct decision; one empirical example	14	13	-02
	code 22: Correct decision; several empirical examples	07	11	04
	code 23: Correct decision; aware examples not conclusive	00	00	00
	code 24: Correct decision; crucial experiment	03	03	00
	code 4: Correct decision; valid justification, general	10	14	4
	code 9: Miscellaneous incorrect (including terms not understood and no response)	15	13	-1

Table V: Changes in response frequencies for L1a, L1b, L1c and L1d from Year 8 to Year 9 (N = 1078)
[c] Frequencies are rounded to the nearest integer

For each item in Table V, the codes are ordered from 1 to 4, and it can be argued that codes with a higher leading digit represent a higher level of response (apart from code 9). On this basis, there is discernible but modest progress from Year 8 to Year 9 in students' responses to each item; fewer students made an incorrect response, more students gave valid justifications and more students recognised that the implication and its converse presented in the question were not equivalent.

The clearest progress can be seen in L1b, where there is a net migration of nearly 20% of the total sample of students from the incorrect code 10 to the correct code 30 response. Progress is less marked for the other items, being well under 10% in each case. Thus, for example, in item L1a there is a net migration of about 7% of students from the code 10 response to a code 31 or 32 response. There is a similar net percentage move from code 1 or 2 responses to code 3 or 4 responses in L1c and from code 1 responses to code 2 or 4 responses in L1d.

In the case of L1a, though the net percentage progress was small, it is interesting to note that despite the complexity of the question (and its written test nature) a substantial minority of students (16% in Year 8, 19% in Year 9, N = 1078) were willing to re-assess their initial evaluation on the basis of their own proofs. Interestingly, these students who changed their answer from ‘Yes’ to ‘No’ on the basis of evidence achieved a somewhat higher mean score on the other numerical/algebraic items in the survey and a higher baseline maths test score than those who stayed with ‘Yes’ or ‘No’ from the outset. However, over 60% of students in each year still maintained that the two statements were ‘saying the same thing’.

Since the study is longitudinal, as well as considering net progress, we are able to examine the percentages of *actual* students who progressed, regressed or gave the same kind of response to each item from Year 8 to Year 9. This is shown in Table VI, (where, as before, progress is defined in terms of the order of the leading digit, except for 9).

	Frequency (% of total sample)			
Item	Progress from Y8 to Y9	Regress from Y8 to Y9	Stay the same in Y8 and Y9	Code 9 response in Y8 and/or Y9
L1a	25	16	56	2
L1b	26	9	54	11
L1c	28	20	29	23
L1d	22	15	37	26

Table VI: Percentage of total sample of students whose responses progress, regress or stay the same for items L1a, L1b, L1c and L1d from Year 8 to Year 9 (N = 1078)

Responses to L1b again stand out as indicated in Table VI by the difference in the Progress and Regress frequencies, which is higher for L1b than the other parts of the question. Additionally, if one considers the ratio of these frequencies, rather than the difference, this is also far greater for L1b (at roughly 3:1 compared to roughly 3:2 for each of the other items).

Discussion of Results

In this section, we distinguish different patterns in student responses to separate parts of question L1 and attempt to interpret these patterns in terms of student strategies and meanings. We are also interested in the interactions between student responses to the various parts of question L1, and set out to trace interactions that led to students' eventual evaluations as to whether an implication and its converse are equivalent. However we should note from the outset that inconsistency in students' responses was not uncommon: for example, a student might appear to recognise the importance of satisfying the antecedent in a logical implication in answer to one part of the question but not in answer to another part. These inconsistencies led us to wonder how far students' justifications of their conclusions (either in writing in this study or verbally during an interactive episode as reported in other studies) can be assumed to be based on an appreciation that if a justification is to be mathematical, it must be applied consistently. Further analysis of the strategies used by students reveal however that apparent inconsistent responses may sometimes be the result of a consistent use of strategies that happen to be limited in their effectiveness, as we discuss later.

We begin by presenting what we think is a useful first approximation of the strategies students use to solve L1c and L1d (ignoring possible interactions between these parts) and how use of the strategies changed over time. Second, on the basis of responses to L1b, we discuss students' awareness of the general applicability and logical necessity of a statement assumed to be true. Finally we analyse students' responses to question L1 as a whole using a version of Toulmin's scheme, in order to come up with a typology of students' meanings for logical implication and the rationale underlying these meanings.

Student strategies to determine the truth of a statement of logical implication

We present first a model of student attempts to determine the truth or otherwise of Joe's and Fred's statements. The model posits three types of strategy, which we call empirical (X), focussed-empirical (Y) and focussed-deductive (Z). The first of these is the least likely to produce a valid conclusion.

X. In the *empirical strategy*, students start by more or less randomly choosing starting numbers, which are not restricted to those that fit the antecedent in the given statement, to generate data about sums and products. Subsequently, they try to relate the data to the statement they are trying to evaluate. Students are likely to use specific starting numbers, although some might reduce the amount of data by treating the numbers in a more general way, as simply 'odd' or 'even', and they might then also adopt an exhaustive approach for combining odds and evens. Both these additional approaches are likely to improve the chances of using the empirical strategy successfully.

Y. In the *focussed-empirical strategy*, students start by *only* selecting starting numbers that fit the antecedent and then go on to produce sums or products in order to determine the truth-value of the consequent and hence of the statement as a whole. Again, some students might also adopt a general approach and perhaps an exhaustive approach as well.

Z. In the *focussed-deductive strategy*, students start by asking themselves what *types* of starting number would fit the antecedent i.e., they use their insight or knowledge about the first proposition to *deduce* the starting numbers or starting number types. They then go on to produce sums or products in order to determine the truth-value of the consequent and hence of the statement as a whole, as with strategy Y. They then seek to explain their conclusion by reference to the original number type. Effectively, this strategy subsumes a general and exhaustive approach.

The distinction between Y and Z is subtle and not always discernible from students' scripts. Nonetheless, the distinction is useful from a theoretical perspective for two reasons. First, it denotes a change of viewpoint, from an empirical, data-driven approach to an approach that is deductive and based on structural relationships, i.e., to the kind of approach that is involved in mathematical proof. Second, the distinction brings out an important difference in how statements that turn out to be true, such as Fred's, can be evaluated: strategies Y and Z, used accurately, both lead to the correct conclusion, that Fred's statement is true, but whereas in Z it is *necessarily* true, this is not the case in Y – at least, not until other kinds of starting numbers have been exhaustively eliminated (i.e., odd-even and even-even pairs). In strategy X, the likelihood of coming up with a correct and valid conclusion will depend, at least in part, on the data that students happen to generate. Students using strategy X might well choose an odd and an even starting number (say 2 and 3), which produces data that is irrelevant, since it does not fit the antecedent for Joe (2+3 is not even) or for Fred (2×3 is not odd), but from which they might well be tempted to draw a conclusion which, even if it turns out to be correct, can not be valid. With Y and Z, on the other hand, the data is more focussed and under the students' control.

It is tempting to see the three strategies as ordered, both in terms of effectiveness and in terms of shifts in thinking, from an inductive approach to one that is deductive and concerned with structure that is appropriate for proof. Thus, for example, we noted a tendency for students who generated a lot of data (and who thus perhaps were using strategy X) to make wrong evaluations, particularly when the data did not fit the antecedent. However, this view has to be treated with caution, especially as the approach of treating the starting numbers only in terms of their odd/even property and the use of an exhaustive approach for generating data might be present. Other influences might also apply. For example, some students found data that fitted the antecedent in the implication (and thus perhaps used strategy Y), but took $p\Rightarrow q$ to be true even though they had other data that would serve as counter examples (perhaps in the belief that 'a statement is true if it is *sometimes* true' by analogy with 'a statement is false if it is sometimes false').

Many students' produced empirical explanations, that is they involved one (or sometimes more) specific numerical example rather just referring to odds and evens, and often these comprised only small starting numbers (e.g., 2), and commonly identical starting numbers (i.e., both 2). Such numbers are far removed from numbers that would provide a 'crucial experiment' (Balacheff, 1988). However, it is possible that for some students these numbers served as generic examples. An obvious appeal is their economy, in terms of the arithmetic involved but also in terms of *choice*: having chosen a number of a certain *type* (for example, one that is even), the simplest way of choosing a second number of this type is to choose the same number again!

In as much as Strategies Y and Z seem more likely to be successful than strategy X, it is possible that some of the (albeit modest) net progress (of about 10%) on L1c and L1d from Year 8 to Year 9 is due to a shift from using strategy X to using strategies Y or Z. Distinguishing these strategies also provides a possible explanation for the apparent inconsistency of many students' responses to parts L1c and L1d, within a given year or between years. Thus, for strategies X and Y in particular, some inconsistencies could be due not to students switching strategies but to the fact that finding the crucial combinations for evaluating Joe's and Fred's statements is to some extent a matter of chance. This could also help to explain why a relatively high proportion of students seem to regress on L1c and L1d, compared to L1b.

Students' appreciation of the logical necessity of an implication

This issue was investigated largely through responses to question L1b. Here, students are asked to suppose that Fred's statement (Product odd $\Rightarrow$ sum even) is true and whether, on that basis, it is then possible to draw a conclusion about the sum of two whole numbers whose product is 1271, or whether the value of the numbers has first to be determined. In contrast to Fischbein's study reviewed earlier and in a similar way to the earlier Healy and Hoyles study, students were not given a proof of the statement and nor did they have to consider whether the statement was actually true.

The idea was to test whether students were able to free themselves from the empirical and consider the implications of a statement at a structural level.

In the event, a high proportion of our students operated successfully at this general level on item L1b, a proportion far higher than for those who seemed to operate at a general level on L1c and, in particular, L1d, where such an approach was needed for a conclusive argument. Thus responses to L1b provide useful evidence that many students (at least in our high-attaining sample) can appreciate the logical necessity of an implication when the antecedent is true and reason at this level in certain circumstances. Also, since students' performance on L1b was found to be a significant predictor in the multilevel analysis of their score on the proof test as a whole, which replicates a finding from the Healy and Hoyles (2000) study, it would seem that considering the implications of a statement at a purely hypothetical level is an important component of learning to cope with mathematical proving.

Students' meanings of logical implication and its converse

In contrast to the preceding sections, our interest here is not so much on the way students evaluate the individual statements (Joe's and Fred's), but on how they decide on whether the two statements are 'saying the same thing'. We present a typology of student responses to L1 *as a whole*, which attempts to capture the different ways students conceptualise the relationship of logical implication and its converse and the rationale underpinning these different meanings, in particular how far they are influenced (or not) by data. To come up with different types of argument, we use Toulmin's scheme comprising conclusion, data, warrant and backing. The 'conclusion' is the claim about whether the statement $p \Rightarrow q$ is saying the same thing as the statement $q \Rightarrow p$; the 'data', which provide a foundation for the claim, is the information about the form of the two statements $p \Rightarrow q$ or $q \Rightarrow p$ in this particular case or about their truth values for the specific propositions stated by Joe and Fred; the 'warrant' is a general principle by which the student seeks to justify why the data support the conclusion, and which is given legitimacy by the 'backing', which might involve further data and warrants.

We distinguish four theoretical categories of student meanings for logical implication, referred to as Types A to D, which capture students' conclusions about the equivalence (or not) of logical implication and its converse and the basis for their conclusions after answering the question as a whole [xv]. The typology was devised partly from theory: one might expect some students to equate a conditional statement with its converse whilst other students might, from the outset, regard them as different (perhaps, in this particular context, through an awareness of the structure of odd and even numbers, or perhaps more generally on the basis of a 'container' scheme as suggested by Rodd, 2000). The different categories in this typology were also however derived from close inspection of the students' scripts and the analysis we have presented earlier[xvi].

We describe each category or type by reference to the basic schematic diagram used by Krummheuer, 1995, based on Toulmin's categories (see Figure 2). Since the backing will not call upon formal logic, the schematic diagram is repeated for some types to indicate that the backing is the conclusion from more data.

Insert Fig 2 about here

Fig 2: Schematic diagram for analysing arguments

In all that follows, proposition p is 'sum is even' and proposition q is 'product is odd'.

Type A: Logical implication and its converse are the same (from the outset without reference to data)

Type A responses make no reference to data concerning the truth values of antecedent and consequent, and simply regard antecedent and consequent as interchangeable.

A typical Type A response profile can be summarised as follows. Students answer 'Yes' to L1a, and use as initial data the fact that in statements $p \Rightarrow q$ and $q \Rightarrow p$, the particular propositions p and q are 'just the other way round'. Students justify this with the general warrant that reversing the order of any propositions makes no difference to a statement of logical implication, and so $p \Rightarrow q$ is the same as $q \Rightarrow p$ for all p and q. Backing is given to the warrant by the claim that the particular statements (Joe's and Fred's), are either both true or both false.

The crucial characteristic of a Type A response is that students do not make proper use of specific data to come to their conclusion; that is, they do not test the truth values of $p \Rightarrow q$ and $q \Rightarrow p$ independently, to demonstrate their equivalence. Rather, they test the truth value of L1c by reference to pairs of numbers (data for this statement) but then simply announce that the truth value of L1d is the same without reference to numbers.

An example from student interviews illustrates a typical Type A response. Student S, in her written response to the Yr 8 test, had decided that Joe's and Fred's statements were saying the same thing, and she presented evidence that Joe's statement was false. She then simply stated that Fred's statement was false also. S was asked why she had written that the statements were saying the same thing:

S: Because Joe says that if the sum of two whole numbers is even, then the product is odd and Fred is saying the same thing, but he's saying that if the product of two whole numbers is odd, then the sum is even. So I think that he's basically saying that they're showing that Joe works out sum of the numbers first and Fred works out the product first, but basically they think the same thing.

The schematic diagram summarising Type A responses, in the case when both statements are evaluated as true, is given in Figure 3.

Insert Figure 3 about here

Fig3: Schematic diagram for a Type A response

Type B: Logical Implication and its converse are the same by reference to data

Type B responses refer to data concerning the truth values of antecedent and consequent, while seeing antecedent and consequent as interchangeable.

A typical Type B response profile can be summarised as follows. Students answer 'Yes' to L1a, and use as initial data the fact that in statements $p \Rightarrow q$ and $q \Rightarrow p$, the propositions p and q are 'just the other way round'. Students then use data about whether the particular statements $p \Rightarrow q$ and $q \Rightarrow p$ are true (that is, Joe's and Fred's statements) to back up their justification for the general equivalence of the two statements. In order to decide about the truth values of the statements $p \Rightarrow q$ and $q \Rightarrow p$, the same pairs of numbers are used to produce the data, indicating that the antecedent and the consequent are seen as interchangeable.

The clearest case of a Type B response is where the same pair of even numbers is used for both statements to show that one proposition (p) is true and the other (q) is false, which is taken to mean that both statements are false, thus backing the warrant asserting their general equivalence. We illustrate this response by an extract from an interview with student T. T felt the statements were saying the same thing and that the statements were both false since she was "...thinking just of the words... what they're actually saying and whether it's the same thing in English terms". To explain this further, she switched to numbers:

> T: ...there is no way if you are talking in maths terms of deciding whether 4 times 2 *caused* (our emphasis) 4 plus 2 to be what it is or vice versa... for this particular thing there's no distinction because 4 plus 2 is going to equal 6 no matter what 4 times 2 is… so um which ever way you put these round doesn't matter because they don't actually *cause* (our emphasis) each other to occur the way that they do… .

T was expressing quite explicitly that she was seeing the propositions as interchangeable, i.e. one is not contingent on the other. This is illustrated in the schematic diagram in Figure 4.

Insert Figure 4 about here

Fig4: Schematic Diagram for a Type B response

Another case that clearly fits Type B is when propositions p and q are both declared false, by using one odd and one even starting number, and where this in turn is used as evidence that both statements are false, and therefore equivalent. Such a response exhibits rather little understanding of how to evaluate the truth of an implication in school mathematics, since for neither Joe's nor Fred's statement is the antecedent true. Where students responded with the same pair of *odd* numbers, to produce data to show that both propositions and therefore both statements were true, it is not possible to decide whether they are distinguishing antecedent from consequent, and therefore whether the response is truly of Type B, though it fits the profile.

Type C: Logical implication and its converse are not equivalent by reference to data

Type C responses regard logical implication and its converse as not equivalent after testing with data the truth values of each specific statement.

A typical Type C response profile can be summarised as follows. Students answer ‘Yes’ to L1a initially, but this is changed to ‘No’ after finding that L1c is false and L1d is true. Thus in Type C responses, it is asserted that a statement $p \Rightarrow q$ is not the same as a statement $q \Rightarrow p$ because $p \Rightarrow q$ is false (since in this case there are data that make p true but q false), and $q \Rightarrow p$ is true (since there are data that make q true and p true or, more powerfully, because there are also no data that make q true but p false). Thus, students whose responses fit Type C start with what appears to be a Type A or Type B response and then, after selecting appropriate pairs of numbers to find the truth values of L1c and L1d independently (and thus obtaining False, True), they change their answer to L1a from ‘Yes’ to ‘No’.

We illustrate this response by a further extract from the interview with student S. Following her Type A response, the interviewer asked S to think about Fred's statement without referring to Joe's, as she had done in her written work. After first re-asserting that it would be false, S started to test this conjecture, and as she did so she became uncertain:

> S: I think it’s probably false. Because I think if you have 10 and 8 the product would be – *oh hold on, they’d be even* (our emphasis). If you have… maybe he is right, I don’t know! I’m not sure about it now....

Later, having been asked why 10 and 8 could not be used for Fred's statement, S gave a reply that suggests that she was now beginning to see Fred's statement as a hypothetical proposition, in that data that did not satisfy the antecedent could not be used: "Because then the product would be even, and it's supposed to be odd". S was now convinced that Fred's statement was true, and having already shown that Joe's statement was false, S changed her answer to L1a to ‘No’.

This type of response is summarised by the schematic diagram shown in Figure 5.

Insert Figure 5 about here

Fig 5: Schematic Diagram for a Type C response

Type D: Logical Implication and its converse are simply not equivalent

In Type D responses, logical implication and its converse are seen as not equivalent; data concerning the truth values of antecedent and consequent are used to confirm rather than justify this conclusion.

A typical Type D response can be summarised as follows. Students answer ‘No’ to L1a. They assert that a statement $p \Rightarrow q$ is not the same as statement $q \Rightarrow p$, because the order matters, and they confirm this with subsequent data which shows that $p \Rightarrow q$ is false and $q \Rightarrow p$ is true. Not surprisingly, students are unlikely to give very clear reasons for their assertion of the non-equivalence. They just somehow ‘feel’ it is the case, as illustrated in the following extract from an interview with student W:

W: I don’t think they’re saying the same thing.

I: Could you say why you don’t think they’re saying the same thing?

W: Because, that’s not a simple reversal. Adding up and timesing is different. It depends what different numbers you have.

I: How do you know it’s not a simple reversal?

W: Because, there are different ways of doing things.

I: Right….

W: If you times something by one then you get the same but if you add a one then it makes a difference…

W: Right……

W: And so, it changes and also divide and times would be a pair and I'd have to think about that more if it was something like that but because there are different ways, processes of doing something.

Thus W's conviction that the two statements were not the same, though it might have been based on rather nebulous analogies with other mathematical experiences, did not rely on specific data.

In Type D responses, students distinguish antecedent from consequent both at a general level and in their choice of pairs of numbers to determine (or demonstrate) the truth values of Joe's and Fred's statements. The schematic diagram summarising Type D responses is given in Figure 6.

Insert Figure 6 about here

Fig 6: Schematic Diagram for a Type D response

Concluding Remarks

We start by recognising the limitations of findings that attempt to describe student understandings based on their response to written tests that can only give a partial view. Nonetheless we found some quite marked patterns and trends in the responses, which gives us some confidence in our conclusions. Additionally, the follow-up interviews with selected students lend support to our theoretical categories, although the interviews also showed how students can change their responses during interactions with a researcher.

Overall, the research has shown that even high-attaining students often fail to appreciate how data can properly be used to support a conclusion as to whether $p \Rightarrow q$ is true or not, and there is only modest progress from age 13 to age 14[xvii]. Thus our study provides some but rather limited support to the assumption that students move from empirical to more deductive approaches with increasing maturity. It does however indicate the complexity underlying learning to prove deductively in mathematics and that progress is not likely therefore to be smooth and trouble free.

In comparing responses across different parts of question L1, one striking feature is that the proportion of students who seemed to call upon general deductive reasoning is far higher in L1b, that is in recognition of the logical necessity of a conclusion of an implication when the antecedent was assumed to be true, than was evident in parts L1c and L1d. Additionally there was most marked and sustained progress in response to this item, suggesting that the higher level responses to L1b are generally more 'secure' than for the other three items; or put another way, there is perhaps a greater element of chance and uncertainty in some of the higher level responses to L1a, L1c and L1d.

A more detailed examination of individual students' scripts served to illustrate how students go about determining the truth of a statement of logical implication in this domain, and we distinguished three strategies from our analysis; empirical, focussed-empirical, and focussed-deductive. The first two strategies in particular may help to explain the inconsistency in students' responses to L1c and L1d within and between years 8 and 9, since even if students use the strategies consistently, their success will depend on the particular data they happen to choose. The first strategy is likely to be the least successful, though the three strategies are not strictly hierarchical as they may (or may not) be used in conjunction with generic, general and exhaustive approaches.

Our study identified a range of meanings students assign to the conditional relationship and its converse. Most assert that a conditional statement and its converse are equivalent and do not test this assertion against evidence of the truth values of the statements, and there is little if any progress over time. This argument could of course suggest superficial engagement with the question and an absence of reflection. However, it could also indicate a rather deep appreciation of mathematical structure in that a structural equivalence is declared so that there is no need to present further data to test it. By more detailed examination of the students' responses to the question under study as whole, we distinguished four theoretical categories where logical implication is (or is not) interpreted as equivalent to its converse, where the antecedent and consequent are (or are not) seen as interchangeable, and where conclusions are (or are not) influenced by specific data.

This analysis opens up questions as to the meanings students assign to their 'reasons' or verifications. It may not just be that reasons are not general, or not concerned with structure – they might simply be conjunctions of evidence. This calls into question what students mean when they present even a correct counter example. Is it in fact showing a recognition that the statement is not true or that the statement is *sometimes* not true? As we have noted, the majority of written explanations were limited to empirical reasoning but in considering empirical reasoning from this different vantage point, it is evident that many students used data in ways that did not distinguish the antecedent from the consequent, or put another way the students did not appreciate the temporal nature of if-then in the present context[xviii]. One might speculate that students who recognise that the antecedent and consequent are not interchangeable would be able to cope with the general deduction required in part L1b. It is also likely that they would be able to answer L1c competently using a counter example obtained by a focussed-empirical or focussed-deductive approach and that they would be able to assess the truth value of L1d if not deductively then by finding supporting empirical evidence. It is also apparent that students who adopted only empirical reasoning did not always realise that it was not 'the mathematical game' to choose only confirming examples. This tendency, to want only to confirm a hypothesis, has long been recognised in experimental studies (Wason, 1960) and noted by Fischbein (1982) as a pragmatic and intuitive approach to proving.

This research has led us to question the value of using formal logic as a model for studying deduction in secondary school mathematics, that is to follow the type of analysis used by Inhelder and Piaget (1958) in their study of hypothetic-deductive reasoning among adolescents or by Wason in his selection task. We have argued that in the context of school mathematics, logical implication is more usefully seen as being primarily concerned with situations in which the antecedent is true (rather than material implication where one is equally concerned with the antecedent being false). Related to this, we suggest that students are more likely to be successful in determining the truth of an implication if they use a focussed-empirical or focussed-deductive approach on data that fit the antecedent than if they start by generating data 'randomly' and then attempt to assess the truth of the statement.

Although we have shown some general, but rather modest, progress in responses as students moved on a year. At the same time, we can point to remarkable shifts in reasoning in some individuals. Why these shifts occurred is almost impossible to say from our study, although it may be the case that significant events in the classroom had impinged on a student's reasoning[xix]. Most Year 9 students who we interviewed were unable to recall *why* they had answered as they did in Year 9 let alone in Year 8, and, disconcertingly, when shown their Year 8 responses were often quite ready to switch back to their earlier (incorrect) argument. Clearly this is an area where understanding can be fragile! Perhaps, given that our findings suggest that there is some connection in this particular mathematical context between understanding the structure of implication and successful deductive reasoning, a way forward would be to make the strategies and different profiles of responses explicit in the classroom. The next step would then be to design activities that focus on developing meanings for the structural properties of logical implication in mathematics, and teaching that sustains and develops these meanings over time.

Acknowledgements

We gratefully acknowledge the support of the Economic and Social Research Council (ESRC), Project number R000237777. We also wish to thank Einar Jahr of Hedmark University College, Norway, for his detailed comments on the paper from the perspective of a logician.

[i] *Deductive* and *inductive* are being used very broadly here, to distinguish between reasoning based on properties or structural relationships and reasoning based on empirical data. This is not to suggest that these kinds of reasoning are entirely distinct and that, for example, broadly inductive reasoning may not contain deductive sequences, as the analysis by Reid (2002) clearly shows.

[ii] In the past when students were taught Euclidean proofs, they would have come across conditional relationships when learning about a proof and its converse, although it is unlikely that the formal connection between them would have been made explicit.

[iii] The questions asked of the students in the two studies were rather different in focus, which may also account for the different result. The focus of the Fischbein work was on whether a particular theorem was true or not.

[iv] This description is somewhat misleading, as the phenomenon is also widely observed among adults. The reason for this converse error in everyday life might be that conversational discourse often (though not always) carries an unstated second meaning: for example, if we say, ‘If it is raining, then I carry my umbrella’, then it is reasonable to assume the converse statement as well as the inverse i.e., ‘If it is not raining, then I do not carry my umbrella’, else you might as well say ‘I carry my umbrella all the time’. Clearly this is not always the case. O'Brien in fact found that the difficulties he identified could be considerably reduced by putting the implication in a realistic context, such as "If it's Jim's car, it is red", since experience indicates that here the converse statement is clearly not true.

[v] Quine makes a similar distinction between what he calls the *material conditional* and our everyday attitude to the conditional, where, "if …. the antecedent turns out to be false, our conditional affirmation (of the consequent) is as if it had never been made" (Quine, 1974, p19).

[vi] The particular statements of logical implication used in our study involve quantification, so if one *were* to view them in terms of formal logical, then strictly speaking this would involve predicate rather than propositional logic.

[vii] Children's utterances tend to be cryptic and elliptical, with only as much information revealed as is deemed necessary to be understood at a given moment (Anderson et al, 1997). The same phenomenon was apparent in many of our students' written explanations, even though the communication here was to a remote audience and non-interactive, where one might have expected students to be more explicit. When coding such explanations it is tempting to read between the lines and to assign underlying strategies to them, a temptation we sought to resist.

[viii] Multilevel statistical analysis using student data in conjunction with teacher and school data has also been completed for both year groups, though not the focus of this paper.

[ix] Strictly speaking from a perspective of logic, p and q are not statements but open sentences whose truth values vary according to two variables. We do however adopt the term statement here as this is used in classrooms.

[x] The other combinations may shed light on whether the statement under consideration actually is an implication, but of course only if students can process the combinations effectively.

[xi] This test comprised selected items from TIMSS, none of which involved deductive reasoning.

[xii] A similar question was used in the previous Healy and Hoyles study and was found to have significant outcomes.

[xiii] In our sample, only a handful of students attempted to 'unpack' this deduction, by trying to explain *why* only odd starting numbers produce an odd product or why their sum must be even.

[xiv] In version B of the Yr 9 test, L1 appeared at the end of the test and it was noticeable that the number of students who made little or no attempt at this question was considerably higher than in version A (218 compared to 127 students). 1984 students took the Yr 8 *and* Yr 9 tests, but to make what we believe to be more accurate comparisons, we present here only the responses of the 1078 students who completed version A of the Yr 9 test (and either version of the Yr 8 test).

[xv] Clearly some students were unable to sustain this multiple level of engagement with the question, as revealed by some of the inconsistencies in their responses. We have not attempted to model these inconsistent responses, nor those responses where students seemed confused about some of the basic terms of odd, even, sum and product.

[xvi] In order to select a range of scripts for this further analysis, students' total scores on the question in Year 8 and Year 9 were used to calculate the percentage increase in the average score for each class; a sample of classes was then chosen which included classes where the percentage increase was very high (greater than 40%) and where it was close to (including just below) zero. All the student responses to L1 as a whole in these classes were looked at again.

[xvii] A very similar pattern of response frequencies is apparent in preliminary analysis of the Year 10 (age 15 years) data.

[xviii] Although it has to be said this might in part be due to an artefact of question L1, namely that the question is framed in terms of starting numbers that affect the sum and product, and hence the antecedent and consequent, 'simultaneously'.

[xix] For example, one of our students, when interviewed about his written response, explained that he had used large starting numbers, such as 9 and 17, to test Fred's statement, "Because I remember doing an experiment before this, which it worked up to a point, then didn't work over the point. We've done one this year as well, it worked up to number seven or something like that".

References

Anderson, R. C., Chinn, C., Chang, J., Waggoner, M. and Yi, H.: 1997, 'On the logical integrity of children's arguments', *Cognition and Instruction*, 15, 2, 135-167.

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

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

Ceci, Stephen J (1990). *On Intelligence…more or less. A bio-ecological Treatise of Intellectual Development.* Century Psychology Series, Prentice Hall, USA.

Cheng, P. W. & Holyoak, K. J.: 1985, 'Pragmatic reasoning schemas', *Cognitive Psychology*, 17, 391-416.

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

Deloustal-Jorrand, V. (2002). 'Implication and mathematical reasoning', *Proceedings of the 26th Conference of the International Group for the Psychology of Mathematics Education* 26, 2, Norwich, England, 281-288.

De Villiers, M.: 1990, 'The role and function of proof in mathematics', *Pythagoras*, 24, 17-24.

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

Godino, J. D. and Recio, A. M.: 1997, 'Meaning of proofs in mathematics education', *Proceedings of the 21st Conference of the International Group for the Psychology of Mathematics Education*, 2, Lahti, Finland, 313-320.

Hanna, G.: 1989, 'Proofs that prove and proofs that explain', *Proceedings of the 13th Conference of the International Group for the Psychology of Mathematics Education*, G.R. Didactique CNRS Paris, 45-51.

Hanna, G.: 1995, 'Challenges to the importance of proof', *For the Learning of Mathematics*, 15, 3, 42-49.

Hanna, G.: 2000, Proof, Explanation and Exploration: an Overview, *Educational Studies in Mathematics*, 44, 5-23

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

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

Hoyles, C.: 1997, 'The curricular shaping of students' approaches to proof', *For the Learning of Mathematics*, 17, 1, 7-16.

Inhelder, B. & Piaget, J.: 1958, *The Growth of Logical Thinking*, Routledge and Kegan Paul, London.

Kieran C, Forman, E. Sfard A. (eds.): 2001, Special Issue "Bridging the Individual and the Social: Discursive Approaches to Research in Mathematics Education", *Educational Studies in Mathematics*, 46, 1/3,

Krummheuer, G.: 1995, 'The ethnology of argumentation', in P. Cobb & H. Bauersfeld (eds.), *The Emergence of Mathematical Meaning: Interaction in Classroom Cultures*, Erlbaum, Hillsdale, NJ, pp 229-269.

Mitchell, D.: 1962, *An Introduction to Logic*. Hutchinson, London.

O'Brien, T. C., Shapiro, B. J., and Reali, N. C.: 1971, 'Logical thinking – language and context', *Educational Studies in Mathematics*, 4, 201-219.

Quine, W. V.: 1974, *Methods of Logic*. (Third edition; first edition published in 1950.) Routledge & Kegan Paul, London.

Reid, D. A.: 2002, ‘Conjectures and refutations in Grade 5 mathematics.’ *Journal for Research in Mathematics Education*, 33, 1, 5-29.

Rodd, M.: 2000, ‘On mathematical warrants’, *Mathematical Thinking and Learning*, 3, 22-244.

Simon, M. A.: 2000, ‘Reconsidering mathematical validation in the classroom’, *Proceedings of the 24th Conference of the International Group for the Psychology of Mathematics Education*, 4, Hiroshima, Japan, 161-168.

Sowder, L. and Harel, G.: 1998, ‘Types of students' justifications’, *The Mathematics Teacher*, 91, 8, 670–675.

Toulmin, S.: 1958, *The Uses of Argument*, Cambridge University Press, Cambridge, UK.

Van Dormolen, J.: 1977, ‘Learning to understand what a proof really means’, *Educational Studies in Mathematics*, 8, 27-34.

Vinner, S.: 1983, ‘The notion of proof – some aspects of students’ views at the senior high school level’, *Proceedings of the 7th Conference of the International Group for the Psychology of Mathematics Education*, 289-294.

Wason, P. C.: 1960, ‘On the failure to eliminate hypotheses in a conceptual task’, *Quarterly Journal of Experimental Psychology*, 12, 129-140.

Wason, P. C. and Shapiro, D.: 1971, ‘Natural and contrived experience in a reasoning task’, *Quarterly Journal of Experimental Psychology*, 23, 63-71.

Yackel, E.: 2001, ‘Explanation, justification and argumentation in mathematics classrooms’, *Proceedings of the 25th Conference of the International Group for the*

Psychology of Mathematics Education, 1, Freudenthal Institute, Utrecht, Holland, pp 9-23.

Yackel, E. & Cobb, P.: 1996, 'Sociomath norms, argumentation, and autonomy in mathematics', *Journal for Research in Mathematics Education*, 27, 458-477.