

CHILDREN'S CONDITIONAL REASONING
PART II: TOWARDS A RELIABLE TEST OF
CONDITIONAL REASONING ABILITY

1. INTRODUCTION

This is the second in a sequence of three papers describing an investigation of fifth graders' ability to learn to distinguish between valid and fallacious inferences from simple conditional premises.

The first paper presented the experimental materials. The present paper discusses the measuring instrument used to assess the effect of the experimental materials. This paper also focuses on several problems and difficulties encountered in the development of reliable tests in conditional reasoning. A fairly comprehensive search of the relevant literature failed to reveal the consideration of these problems in previous studies of children's conditional reasoning ability.

2. THE MEASURING INSTRUMENT

In order to examine the effect of the teaching-learning process on students' conditional reasoning ability, a 32-item test was developed. The development process was carried out through several field trials, item analysis and revisions.

Each test item starts with two premises; the first premise is a conditional sentence* formulated with a reasonable hypothetical content designed to make sense to fifth graders in the selected population. In each test item the conditional sentence is followed by a second premise which is its antecedent, its consequent, or one of these denied, thus providing four logical forms for the pair of premises. The test consists of four 8-item sets, each set in one of the following four logical forms:

AA: Affirmation of the Antecedent, (Modus Ponendo Ponens).

DC: Denial of the Consequent, (Modus Tollendo Tollens).

AC: Affirming the Consequent.

DA: Denying the Antecedent.†

* A conditional sentence is one of the form: 'If a , then b '; ' a ' is the antecedent and ' b ' is the consequent.

† For a review of the logical form see Table 1.

Each of these 8-item subtests consists of four pairs of items. The conditional premise in each pair is stated in one of the following four negation modes:

PP: No negation occurs in either the antecedent or the consequent.

PN: Negation occurs in the consequent only.

NP: Negation occurs in the antecedent only.

NN: Negation occurs in both the antecedent and the consequent.

Items in the various logical forms and negation modes were interspread. Each item called for a choice of one out of three answers. All items were presented in written form, each item on a separate half-size sheet, all assembled to form a puzzle booklet. The investigators administered the test as a group test. (See paragraph no. 7: Administration, below).

3. PROBLEMS ENCOUNTERED IN DEVELOPING THE INSTRUMENT

3.1 *Logical form.* Let a , b denote two declarative English statements. Table 1 gives the four patterns of inferences of conditional logic.

Following the four logical patterns of inference from conditional premises given in Table 1, it is easy to build up four test-items from any given conditional sentence: e.g., the AA example given in Table 1 would become a test item stating:

If Jane is Jack's sister, then Jane lives in Washington St.

Jane is Jack's sister.

Does Jane live in Washington St.?

Clearly the answer is in the affirmative.

The model presented in Table 1 is within the realm of sentential logic. This is a very limited portion of deductive reasoning. There are many examples of intuitively correct deductions that cannot adequately be mirrored by this model. The case most relevant to this study is exemplified by the following:

If a positive integer has more than two divisors, then it is not a prime number.

26 is a positive integer which has more than two divisors.

Therefore 26 is not a prime number.

Even though the above deduction very closely resembles the AA pattern of inference, in fact it is not a pure application of AA. Observe that the second given sentence is not exactly the antecedent of the given conditional sentence,

TABLE 1: RULES OF INFERENCE AND THEIR VALIDITY

Premises ($a \rightarrow b$ denotes if a , then b)	Conclu- sion(s)	Name of rule of Inference (abbreviated)	Example
$\begin{cases} a \rightarrow b \\ a \end{cases}$	b valid	Modus Ponendo Ponens (AA for Affirmation of the Antecedent)	If Jane is Jack's sister (a), then Jane lives in Washington St. (b) Jane is Jack's sister. (a) \therefore Jane lives in Washington St. (b)
$\begin{cases} a \rightarrow b \\ \neg b \end{cases}$	$\neg a$ valid	Modus Tollendo Tollens (DC for Denial of the Consequent)	If Jane is Jack's sister (a), then Jane lives in Washington St. (b) Jane does not live in Washington St. ($\neg b$) \therefore Jane is not Jack's sister ($\neg a$)
$\begin{cases} a \rightarrow b \\ b \end{cases}$	a ; $\neg a$ neither is valid	Affirming the Consequent (AC)	If Jane is Jack's sister (a), then Jane lives in Washington St. (b) Jane lives in Washington St. (b) \therefore Jane is Jack's sister. (a) or Jane is not Jack's sister. ($\neg a$)
$\begin{cases} a \rightarrow b \\ \neg a \end{cases}$	b ; $\neg b$ neither is valid	Denying the Antecedent (DA)	If Jane is Jack's sister (a), then Jane lives in Washington St. (b) Jane is not Jack's sister. ($\neg a$) \therefore Jane does not live in Washington St. ($\neg b$) or Jane lives in Washington St. (b)

but is rather an instantiation of the antecedent. Second, the conditional sentence itself is not really built up from two declarative sentences a , b connected by 'if . . . then', since it refers to an arbitrary (unspecified) number. Its antecedent subject, 'a positive integer', allows for a wide range of numbers. some of which have more than two divisors, others exactly two, still others fewer than two. This lack of a well defined truth value for the antecedent excludes it from the set of interpretations which any well-formed formula of

sentential logic can take. The phrase 'A positive integer', within the antecedent, is an example of a free variable, i.e., a variable which is not controlled by a quantifier. The lack of definite truth value for the antecedent may be remedied in various ways, for example, by stating: 'There exists a positive integer which has more than two divisors.' This sentence is true. Alternatively, one could say: 'All positive integers have more than two divisors,' which is a false sentence. However, these remediations change the assertion expressed in the original conditional sentence. The conditional sentence 'if a positive integer has more than two divisors, then it is not a prime' says in fact: 'For *all* x , if x is a positive number with more than two divisors, then x is not a prime.' (As such it belongs to another part of logic known as *predicate* logic). In the indicated sentence x is no longer a free variable since it is controlled by the quantifier 'for all'. The truth or falsity of such a sentence is no longer determined by the truth or falsity of its antecedent and consequent parts. Rather its truth value is determined by a check of all possible interpretations of the sentence. This is done, in principle, by putting every positive integer to the tests of having more than two divisors, and of being a prime. Since no positive integer with more than two divisors is a prime, this sentence is true, and our sentence about the number 26 is validly deduced from it.

English expressions are sometimes misleading as in the example above, when they *seem* to include a free variable whereas in fact they contain an implicit controlling quantifier which is not explicitly mentioned. This, in fact, happened twice in our example, once in the antecedent as discussed, and again in the consequent where the word 'it' was used.

In the test developed for the present study, examples similar to the above appear very often, and are not usually distinguished from pure sentential ones. It is however important to be aware of the differences between the two logical models underlying inferences from conditional premises. These differences may explain some of the difficulties children encounter in drawing conclusions.

3.2. Negation mode. Let p , q denote two declarative English sentences, neither of which includes negation. Four conditional sentences of the form 'if a then b ' may be constructed by substituting either p or not- p for a , and either q or not- q for b . From each conditional sentence we thus obtain four sentences, one in each negation mode: PP, PN, NP, NN.

By repeating the procedure of item construction described in Section 3.1 above, we obtain four items for each negation mode, one in each logical form; thus, altogether sixteen items are obtained from each conditional sentence. Table 3 clarifies this structure.

TABLE 2: RULES OF INFERENCE EXTENDED

x denotes a variable; a, b are one-place predicate symbols; t is a constant symbol; and "V" means for all.*

Premises	Conclusion	Rule of Inference	Example
$\{ \begin{array}{l} \forall x[a(x) \rightarrow b(x)] \\ a(t) \end{array} \}$	b(t) valid	AA	If someone plays too much football [a(x)], then he does not do enough homework. [b(x)]. John plays too much football. [a(t)] \therefore John does not do enough homework. [b(t)]
$\{ \begin{array}{l} \forall x[a(x) \rightarrow b(x)] \\ \neg b(t) \end{array} \}$	$\neg a(t)$ valid	DC	If someone plays too much football [a(x)], then he does not do enough homework. [b(x)]. John does enough homework. [¬b(t)] \therefore John does not play too much football. [¬a(t)]
$\{ \begin{array}{l} \forall x[a(x) \rightarrow b(x)] \\ b(t) \end{array} \}$	a(t); ¬a(t) neither is valid	AC	If someone plays too much football [a(x)], then he does not do enough homework. [b(x)]. John does not do enough homework. [b(t)]. \therefore John plays too much football. [a(t)] or John does not play too much football. [¬a(t)]
$\{ \begin{array}{l} \forall x[a(x) \rightarrow b(x)] \\ \neg a(t) \end{array} \}$	¬b(t); b(t) neither is valid	DA	If someone plays too much football [a(x)], then he does not do enough homework. [b(x)] John does not play too much football. [¬a(t)] \therefore John does not do enough homework. [¬b(t)] or John does enough homework. [b(t)]

* The existential quantifier is not introduced because it rarely appears relevant to this study.

O'Brien (1972) used three basic conditional sentences to create a 48-item test of conditional reasoning. From each basic sentence, 16 items were formed according to the structure given in Table 3. O'Brien applied a full factorial design to study the effects of logical form, negation mode and their interaction.

TABLE 3: FOUR LOGICAL FORMS IN EACH OF THE FOUR NEGATION MODES

Logical Form	Negation Mode			
	PP	PN	NP	NN
AA	$p \rightarrow q$ p	$p \rightarrow \neg q$ p	$\neg p \rightarrow q$ $\neg p$	$\neg p \rightarrow \neg q$ $\neg p$
DC	$p \rightarrow q$ $\neg q$	$p \rightarrow \neg q$ q	$\neg p \rightarrow q$ $\neg q$	$\neg p \rightarrow \neg q$ q
AC	$p \rightarrow q$ q	$p \rightarrow \neg q$ $\neg q$	$\neg p \rightarrow q$ q	$\neg p \rightarrow \neg q$ $\neg q$
DA	$p \rightarrow q$ $\neg p$	$p \rightarrow \neg q$ $\neg p$	$\neg p \rightarrow q$ p	$\neg p \rightarrow \neg q$ p

\rightarrow means "if ... then"

$\neg p$, $\neg q$ denote "not- p ", and "not- q " respectively.

Note: Double negations: $\neg(\neg p)$ and $\neg(\neg q)$ are replaced by p , q respectively.

Even though this approach appears to be controlled and elegant, it involves tremendous complications when meaningful content is considered. These complications arise from the extreme difficulty of finding even three conditional sentences that are equally meaningful in all four negation modes. It often happens that only two of the four negation modes make sense, and usually these are either PP, NN or PN, NP. To give just one example consider O'Brien's sentence:

If the horse does jump, Jim will get hurt.

which O'Brien categorized as causal.

Transposing this PP sentence into the various modes, we get:

PN: If the horse does jump, Jim will not get hurt.

NP: If the horse does not jump, Jim will get hurt.

NN: If the horse does not jump, Jim will not get hurt.

The PN and NP modes of this sentence are harder to accept than the original PP sentence or its contrapositive NN sentence. For this reason it is more difficult, in the present example, to draw conclusions from the PN and NP modes, than from the PP and NN modes.

This makes the category of causal content questionable at least for these two modes*. Clearly, even though there is just a very small difference in the wording, content is not strictly preserved across negation modes. In the study discussed in this paper, it was considered far more important to ensure the reasonability of the conditional sentences, than to provide for similarity of verbal presentation. It was therefore decided to start each of the 32 items of the test with a new conditional sentence.

4. PSYCHOLOGICAL ASPECTS AFFECTING VALID REASONING

4.1. *The relations between symbolic logic and the psychological reasoning process.* Symbolic logic is not intended to be, and indeed is not, a model of the reasoning processes going on in the mind while deductive thought takes place. Logic provides objective criteria for judgment of the validity of the outcome of the reasoning process. To merit a favorable judgment, arguments should be arranged in a sequence, and inferences should be checked against the criteria provided by symbolic logic. However, only rarely does such a sequence bear similarity to the temporal sequence of thought by which one reaches these conclusions.

In the particular case of the test involved in the present study almost no thinking is needed to answer the test items once the algorithmic solution provided by symbolic logic has been mastered † (see Table 4).

The purely syntactical inferences established by symbolic logic should not be confused with the psychological thinking processes of human reasoning, which usually involve semantic and other considerations. The claim that symbolic logic does not serve as a model of thinking processes is true not only because the application of formal logic sometimes 'saves' thinking, but also in view of the many psychological factors that interfere with logical analysis, which are not reflected at all in the analysis of symbolic logic.

4.2. *Content effects.* In a study demonstrating the effect of verbal stereotypes on syllogistic reasoning, A. Lefford (1946) suggests that "Discerning

* See Section 4 for further analysis of content categories.

† To avoid any misunderstanding it is stressed again that the algorithmic solution was not introduced to the students through the experimental materials of the present study.

TABLE 4: ALGORITHMIC SOLUTION TO RELEVANT PROBLEMS OF THIS STUDY

p, q are distinct simple (atomic) sentence symbols, in which no negation occurs;
 + is the conditional symbol interpreted "if ... then ...";
 - is the negation symbol;
 p? q? are the interrogative sentences corresponding to p,q;
 "clues" is the term used in this study for the given information on which one should base the answer;
 "not enough clues" (NEC) is an abbreviated term for the idea that the information given in the clues does not suffice to yield a definite "yes" or "no" answer to the question;
 NOTE: The question never includes negation.

LOGICAL FORM NEGATION MODE		DECIDABLE		UNDECIDABLE	
		AA	DC	AC	DA
PP	clues:	$\{ \begin{matrix} p \rightarrow q \\ p \end{matrix} \}$	$\{ \begin{matrix} p \rightarrow q \\ -q \end{matrix} \}$	$\{ \begin{matrix} p \rightarrow q \\ q \end{matrix} \}$	$\{ \begin{matrix} p \rightarrow q \\ -p \end{matrix} \}$
	question:	q?	p?	p?	q?
	answer:	yes	no	NEC	NEC
PN	clues:	$\{ \begin{matrix} p \rightarrow -q \\ p \end{matrix} \}$	$\{ \begin{matrix} p \rightarrow -q \\ q \end{matrix} \}$	$\{ \begin{matrix} p \rightarrow -q \\ -q \end{matrix} \}$	$\{ \begin{matrix} p \rightarrow -q \\ -p \end{matrix} \}$
	question:	q?	q?	p?	q?
	answer:	no	no	NEC	NEC
NP	clues:	$\{ \begin{matrix} -p \rightarrow q \\ -p \end{matrix} \}$	$\{ \begin{matrix} -p \rightarrow q \\ -q \end{matrix} \}$	$\{ \begin{matrix} -p \rightarrow q \\ q \end{matrix} \}$	$\{ \begin{matrix} -p \rightarrow q \\ p \end{matrix} \}$
	question:	q?	p?	p?	q?
	answer:	yes	yes	NEC	NEC
NN	clues:	$\{ \begin{matrix} -p \rightarrow -q \\ -p \end{matrix} \}$	$\{ \begin{matrix} -p \rightarrow -q \\ q \end{matrix} \}$	$\{ \begin{matrix} -p \rightarrow -q \\ -q \end{matrix} \}$	$\{ \begin{matrix} -p \rightarrow -q \\ p \end{matrix} \}$
	question:	q?	p?	p?	q?
	answer:	no	yes	NEC	NEC

observations and experimental evidence must lead one to conclude that rational thinking is not free from the influence of the affective processes". Lefford found that most subjects solve neutrally toned syllogisms more easily than syllogisms involving controversial matters which are likely to arouse some affective reaction to their subjects. (For a long period most psychologists tended to equate logical thinking with syllogistic reasoning).

As early as 1928 Minna Wilkins scrutinized the effect of syllogism content on the ability of students to accept valid conclusions. Wilkins found familiar material to be the least disturbing to logical thinking, suggestive and symbolic

material somewhat disturbing, and material of unfamiliar scientific or non-sense terms to be the most deleterious.

Feather (1964) found that the reasoning process was often distorted when subject matter was likely to play on emotions, prejudices, or attitudes. His results show that when subjects agreed with the conclusion, they made more errors in accepting invalid conclusions than in rejecting valid ones. When they disagreed with the conclusions more errors were made in rejecting valid conclusions than in accepting invalid ones.

Ennis and Paulus (1965) studied critical-thinking readiness in grades 1 – 12. In the first phase of this study they found that on their class-inclusion reasoning test, the concrete-familiar-content component was in general easier for adolescents than the symbolic presentation and the suggestive content. However, on their conditional reasoning test, which is most relevant to the present study, the three components were of about equal difficulty at each adolescent grade level.

4.3. Limitations of the test in the present study. The relevant literature points to a distinction between ‘causal relations’ and ‘class inclusion’ in the content of conditional reasoning problems (O’Brien 1971, 1972; Roberge 1969, 1970). Such a distinction is vague because any content in which causal relations are embedded can be expressed in terms of class inclusion. For example, even though the statement ‘If you enter the sea, you get wet’ sounds completely causal, it can be interpreted in the following manner: ‘the set of events consisting of your entering the sea is a subset of the set of events consisting of your getting wet’. In fact, one often says, ‘Whenever you enter the sea, you get wet’, which already has a class inclusion flavor. The difficulty in clearly defining ‘cause’ is familiar from studies of metaphysics. Since the distinction is not well defined, it is difficult to apply the distinction in constructing test items. However, about half of the items in the test discussed here have according to the subjective judgment of the researchers, a predominantly class-inclusion content.

All items in the test express, in English, content assumed to be familiar to fifth graders in California. This notion of ‘familiarity’ gives rise to another ambiguity discussed in the literature, namely the distinction between factual content and non-contradictory hypothetical content. Actually the two concepts differ only in the relative size of a relevant model that a reasonable student may readily have in mind for interpretation of the sentence. The word ‘model’ is used here in the sense used in logic, i.e., a universe in which the sentence becomes true. To illustrate this point consider the sentence ‘If the car is fast, then it is shiny’ the converse of which appears in O’Brien’s studies.

The collection of cars which are both shiny and fast is one of many models for this sentence, since for this collection the sentence is certainly true. If our subjects were living in a world where only shiny fast cars existed, the sentence would be factual for them. However, for a Californian child this is not only a hypothetical sentence but also contradictory, in some sense, to his everyday experience with fast cars which are not at all shiny. It was not the purpose of this study to teach children to deal with hypothetical situations for which it is difficult to imagine a model, nor to test the extent to which they are able to do so. Therefore, items based on sentences like that in the example above were omitted from the test.

The following is an example of a sentence which may be factual for some people but hypothetical for others. 'If he has a driver's license, then his age is at least sixteen'. This sounds factual to a Californian, but is not necessarily so for residents of another state. Nevertheless, such sentences were considered to be familiar in content and therefore legitimate in the test. Since the distinction between factual and hypothetical content is also not absolute, in the present test an effort was made to use only conditional sentences whose content seemed to the experimenters to be reasonable for a nine-to-eleven-year-old California child. An effort was made to select items of content that were easy to picture, familiar, or described in previously learned terms, without distinction between factual and hypothetical content. None of the items had content contradictory to experience. The words 'easy to picture' refer to sentences which a child of nine to eleven may have used before, or have heard in a natural conversation, and for which the child can easily devise a concrete model in which that conditional sentence holds. The next example is intended to illuminate this point.

For children whose mothers were working it was difficult to accept the following sentence: 'If George is sick today, then his mother will stay home with him'. Their first reaction was, 'What if she *has* to go to work?'. Nevertheless this sentence was not considered contradictory to experience even for those students, for it was considered that they could, without too much difficulty, imagine a suitable model. The same is true for: 'If someone plays too much football, then he does not do enough homework'. This sentence may raise emotional objection in some children; however, it was not regarded as contradictory to experience, and was not seen as causing difficulties in the creation of an appropriate model.

It may happen that immediate experience distracts the student from a given item. For example, the item: 'If it is raining, then it is cloudy. It is not cloudy. Is it raining?', has a negative answer, which may contradict reality on a rainy day. Efforts were made to avoid items leading to such situations, but

since the test dealt with familiar content, these efforts may not have been entirely successful.

To build a list of items, conditional sentences were taken from previous studies (Hill 1961, Miller 1968, Carroll 1970) and modified or replaced according to the limitations stated above. The sentences were also carefully checked to ensure that the content of each conditional sentence would not suggest its inverse or converse. For example, Hill's (1961) sentence, 'If Ann is at school, then she is the leader today,' was omitted because the truth of its inverse – 'If Ann is not at school, then she is not the leader today' – may be inferred from the original sentence, as Ann could clearly not be the leader if she did not come to school. Since the original sentence functions as a biconditional sentence, it leads to a justifiable but technically incorrect yes or no answer to AC or DA items built from this sentence. The same is true for the following example of Miller's (1968): 'If Harry finds his meal ticket, then Harry can eat his lunch'. This sentence, even though it does not logically imply its inverse, is very likely to suggest its inverse to many students who have experienced a lack of lunch consequent upon a lost lunch card.

It is particularly difficult to design sentences that do not suggest their inverse in the NP negation mode (If not- p , then q), for in ordinary language this mode is usually used for dichotomous situations. For example, 'If you do not feel all right, you should see your doctor'. Obviously, if you are all right you do not need to see your doctor. Suppose, now, we include an item like this

Clues: a. If you don't feel all right, then you should go see your doctor.
 b. You feel all right.

Question: Should you go see your doctor?

A child using his common sense (along with logic) may answer – no. For the child will see no need to see his doctor if he is all right (even though, in fact, he may need to see him for some other purpose). The correct answer – not enough clues – is therefore unlikely to be used even by the good logical thinkers, since in addition to the given information they use some implicit knowledge that the content suggests.

The next and final example is also in the NP negation mode:

Clues: a. If he is not here, then he is there.
 b. He is here.

Question: Is he there?

The common sense answer to this question is 'no'. This conclusion does *not* follow logically from the clues, but from the clues together with the additional implicit premise that no one can be in two different places at the same time.

In other words, the first conditional sentence is known to be a biconditional, because its inverse – if he is here, then he is not there – is always true. We cannot expect children to ignore this knowledge and rely just on the logical validity of an inference, when their experience adds more information than is explicitly given in an item. It is in fact a general goal in education that, in the process of problem solving, a student *will* associate relevant knowledge and relate it to data explicitly given. It was certainly not the purpose of this study to destroy this intuition. For this reason items such as the above were regarded as misleading and were not included in the test.

It should be noted, however, that as long as we do not deal with content-free items, problems of content are confounded with those of logic, and we can never totally avoid confusion caused by content, as in the above examples.

Finally, it should be admitted that criteria for including a certain item in the test were based more on intuition and careful examination of content than on definitions relating to the above ambiguities. Revisions of the initial test version were designed to increase the internal consistency of the test, using item analysis.

5. VALIDITY PROBLEMS INHERENT IN ITEM FORMAT

5.1 Positive and negative questions. The question following the two premises in each test item may be stated in two ways: positively and negatively. A negatively stated question seems in some cases to be the ‘correct’ one to ask in view of the logical form. For example:

- Clues: a. If the Parkers’ car is not in their garage, then they are not home.
 b. The Parkers’ car is not in their garage.

By applying modus ponens to the above premises we infer: ‘The Parkers are not home’. The direct interrogative form for this inference is ‘Are the Parkers not home?’. Most people would answer this question by ‘no’, meaning ‘no’, they are not home.’ This ‘no’ answer reinforces the negation of the question rather than negating it. In other words, the ‘no’ is an affirmation and takes the place of the usual affirmative answer ‘yes’ – which is the correct answer to the negatively stated question.

Since scoring is based upon logic as the sole objective criteria for correctness judgment, a ‘no’ answer to the negatively posed question would be marked as wrong, despite the fact that the replier may have used the ‘no’ answer in the right affirmative sense. To avoid such misjudgments questions were always stated positively.

5.2 *The cost of purifying the logical form.* The logic required in order to arrive at the right answer is slightly more complicated than a simple application of the relevant rule of inference. For instance, in the previous example, inference by modus ponens that the Parkers are not home, must be followed by reference to the question and giving of the answer. An alternative item format was considered with the aim of restricting the required inference pattern to a pure application of the rule.

In this format the same AA item states the two premises (clues) as before (see above), but puts the question differently as follows:-
 ‘Which of the following sentences is therefore correct?’:

1. The Parkers are home.
2. The Parkers’ car is in their garage.
3. The Parkers are not home.
4. None of the above is correct.

Despite the purer logic involved in answering the question in this format, it was found unsatisfactory because:

- (i) It requires much more reading and is therefore much more tiring.
- (ii) One of the alternative answers always contradicts the second clue (in this case the second alternative does).
- (iii) The answer is remote from the clues, making it harder to relate the clues to the conclusion.
- (iv) The phrasing on the whole is much more awkward than in the previous format.

A third format suggested by Paulus (1967) was also considered. In his test Paulus presents the items in the following way:

‘Suppose you know:
 If the shirt is green, then I will not wear it.
 I will not wear the shirt.
 Then would this be true?
 The shirt is not green.’

Paulus’ format was rejected for reasons similar both to those mentioned about phrasing a negative question and those mentioned for the second format. In addition it was felt that the words ‘be true’ in the question are a poor substitute for the words ‘necessarily follow’ which are more accurate, but less familiar to young children.

5.3 *Answer choice*. Each test item is a three-choice one. The correct answer for AA and DC items is either yes or no, depending upon their negation mode. For AC and DA items the right answer is indefinite. Several expressions of the undecidable nature of these items were used by previous researchers: 'Can't tell', 'Maybe', 'Not-enough-clues'. There is a psychological difference between declaring: 'I can't tell' and 'there are not enough clues to decide'. In the first one the blame for not reaching a definite answer is on the student. In the second alternative the blame is on the item. It was found that the answer 'maybe' was considered by some students to be an easier escape in case of doubts. Therefore 'not-enough-clues' was chosen to express the undecidable nature of AC and DA items. The full meaning of this alternative was explained to test subjects before the test began, including consideration of examples.

6. VALIDITY AND RELIABILITY

Apart from its particular content, the test is a direct measure of the ability validly to infer from conditional premises. As such, its validity is by definition unquestionable. The fact that content is sometimes confounded with logic does not reduce the test's validity at all. In fact, the opposite is true. Since it is difficult to imagine any content-free application of logic in a 'real life' setting, the variety of content which appears in the test strengthens its content-wise generalizability within the limitations discussed earlier.*

A split half Pearson-product-moment-correlation-coefficient with Spearman-Brown formula for double length was applied to pretest total scores for the whole sample (210 subjects). This yielded a reliability of .79.

7. ADMINISTRATION

The introduction, i.e., reading the instructions, working out examples and filling in of personal data such as age, name etc. took 8 – 10 minutes in each class. Within the next 20 – 25 minutes students worked on their own, while one experimenter circulated to answer the very few requests for help in reading a word. Two mixed order versions of the same test were used. The two versions of the test were duplicated on differently colored paper. The test was presented to the students as a green-versus-yellow team contest in puzzle solving. The teams were the two halves of the class created by handing out different

* Since reliability is regarded as a necessary condition for validity (Anastasi, 1968), the reliability of the test follows. However, reliability has many facets. It is not uniquely defined, and hence its necessity relative to validity is not in the strict 'by definition' mathematical sense. Because of such doubts, an independent reliability study was carried out.

versions (colors) to neighboring students. This was intended to serve two purposes:

- (i) Increase the student motivation to take the test, and in this way reduce random or careless answers by creating a competitive team atmosphere, and,
- (ii) Insure independent work by each student. Due to the manner of distribution each student was surrounded by students of the opposite team.

The instructions given prior to taking the test included two examples, one of which had 'Not enough clues' as the right answer. In addition, a verbal explanation of the 'Not-enough-clues' alternative was given. A no-time-limit announcement was made, and help in reading was offered when necessary.

To prevent peer pressure of faster on slower students towards the end of the testing period, tests were not collected until all the students in the class had finished their work. Early finishers were asked to draw a picture or a design on the back of any page of the test booklet. (Many were reluctant to stop drawing when the test was over.)

Team scores were given to the classes a few days after each test. Individual scores were given to teachers. No correction period followed the test and puzzle books were not returned to students. Students seemed to enjoy the pretest a great deal, but were less enthusiastic, even though cooperative, during the posttest. This change of attitude was particularly noticeable in the experimental group.

8. PRETEST RESULTS*

For each test item the number of yes's, no's and NEC's (not enough clues) were counted. Table 5 shows item profiles for the whole sample of 210 students. Data were collected before instruction of the experimental unit started.

Table 5 shows that for each of the 16 decidable items most of the students gave the right answer before any learning took place. With the undecidable items the opposite was the case.

No matter what negation mode the premises took in each of the sixteen undecidable items, at least half of the students gave an incorrect definite (yes or no) answer instead of NEC. In 10 of the items more than 75% of the students did so.

* Whole sample results are presented here. Further analysis of pre- and posttest results is presented in the next paper, the third in this sequence. The third paper displays the research design and discusses in detail the measured impact of the experimental materials described in the first paper.

TABLE 5: TEST-ITEM PROFILES BASED ON PRETEST RESULTS OF 210 PARTICIPATING STUDENTS (Right answers are circled)

Item No.	Logical Form	Negation Mode	No. of Yes's	No. of No's	No. of NEC's	No. Skipped	Total	
D E C I D A B L E	1	AA	180	12	17	1	210	
	23	AA	172	25	13	0	210	
	25	AA	61	119	30	0	210	
	30	AA	6	190	13	1	210	
	11	AA	189	7	14	0	210	
	13	AA	163	23	24	0	210	
	19	AA	4	199	6	1	210	
	3	AA	25	169	16	0	210	
	22	DC	++	5	196	9	0	210
	31	DC	++	4	179	24	3	210
	8	DC	+-	37	134	37	2	210
	16	DC	+-	8	193	9	0	210
	5	DC	-+	111	66	31	2	210
	9	DC	-+	182	6	22	0	210
20	DC	--	172	29	9	0	210	
28	DC	--	152	48	10	0	210	
U N D E C I D A B L E	2	AC	159	3	47	1	210	
	14	AC	155	14	38	3	210	
	4	AC	140	2	64	4	210	
	7	AC	159	8	43	0	210	
	21	AC	87	60	63	0	210	
	27	AC	4	159	46	1	210	
	12	AC	64	101	45	0	210	
	18	AC	5	160	44	1	210	
D A B L E	17	DA	12	150	47	1	210	
	10	DA	14	146	49	1	210	
	6	DA	92	28	90	0	210	
	32	DA	60	83	64	3	210	
	29	DA	10	140	56	4	210	
	15	DA	17	81	112	0	210	
	24	DA	109	18	83	0	210	
	26	DA	--	165	6	39	0	210

The initial situation leaves no doubt as to the need for improving subject's ability to separate fallacious from valid inferences.

Despite the differences in the instruments, the present results qualitatively replicate, and in fact even quantitatively sharpen previous researches reporting the dominance of 'child's logic' (O'Brien et al., 1971). This may have been due to the efforts invested in validating the instrument. Figure 1 shows score distribution on each logical form subtest for the control and experimental groups.

The equivalence of the experimental and control groups prior to implementation of experimental teaching is demonstrated by Figure 1. In general, in the pretest, both groups exhibited a mastery of the AA and DC cases, and a lack of ability to distinguish these cases from the AC and DA ones.

Only 15% of the whole sample gave correct answers to at least 75% of the

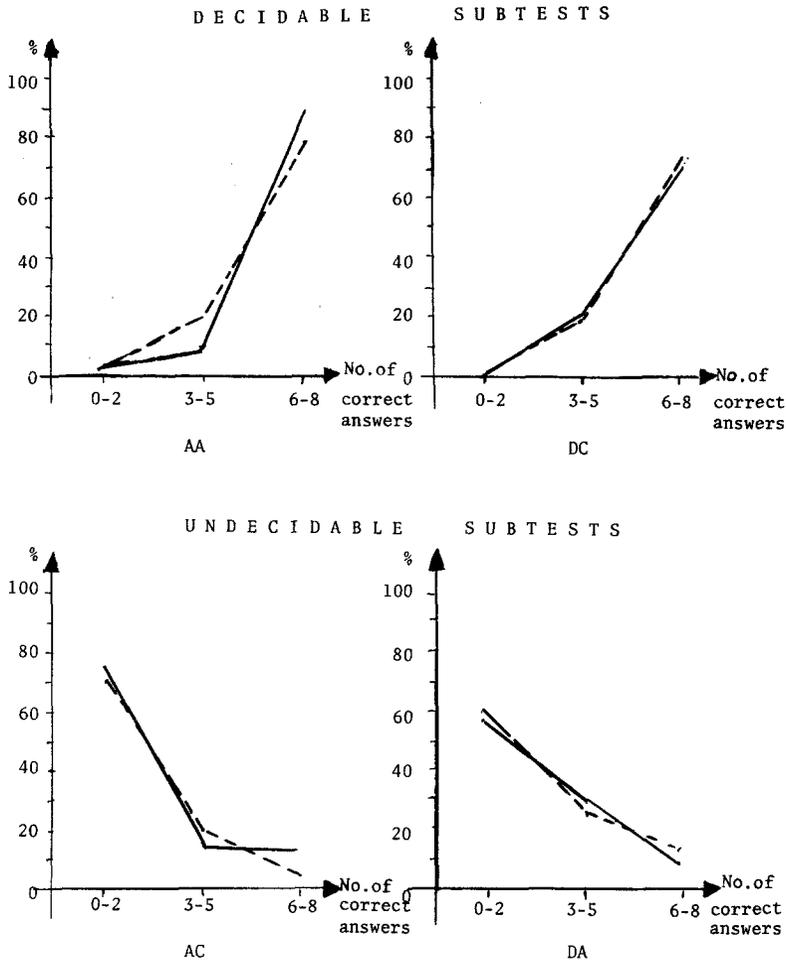


Fig. 1 Percentages of students in the experimental and in the control groups having various pretest scores on logical form subtests.

— experimental group
 --- control group

undecidable items. This general lack of ability to point out that a certain conclusion does not necessarily follow from the given premises, points to a difficulty. There is, namely, a doubt concerning the extent to which the successful performance on decidable items is attributable to logical analysis, rather than to a fortunate mechanism that a child applies and which has nothing to do with logical analysis. This mechanism may be responsible both for wrong answers on

undecidable items, and for right answers on decidable items. A conjecture concerning the possible nature of such a mechanism will be developed in a future paper.

*Technion,
Haifa, Israel*

*University of California
Berkeley, USA*

REFERENCES

- Anastasi, A. (1968), *Psychological Testing*, Third Edition, New York: Macmillan Company.
- Carroll, C.A. (1970), 'Low Achievers' Understanding of Four Logical Inference Forms: An Analysis of Difficulties and of the Effect of Instruction,' Unpublished Ph.D. dissertation, New York: Columbia University.
- Ennis, R.H. and Paulus, D.H. (1965), *Critical Thinking Readiness in Grades 1 - 12, Phase I: Deductive Logic in Adolescence*, Ithaca: Cornell University.
- Feather, N.T. (1964), 'Acceptance and Rejection of Arguments in Relation to Attitude Strength, Critical Ability and Intolerance of Inconsistency,' *Journal of Abnormal Social Psychology* LXII, 127 - 137.
- Hill, S.A. (1961), 'A Study of the Logical Abilities of Children.' Unpublished Ph.D. dissertation. California', Stanford University.
- Lefford, A. (1946). 'The influence of emotional subject matter on logical reasoning,' *The Journal of General Psychology* 28, 367 - 373.
- Miller, W.A. (1968). 'The Acceptance and Recognition of Six Logical Inference Patterns by Secondary Students,' Unpublished Ph.D. dissertation, University of Wisconsin.
- O'Brien, T.C. (1970), 'Logical thinking in children ages 6 - 13.' *Child Development* 41, 823 - 829.
- O'Brien, T.C. (1973), 'Logical thinking in college students,' *Educational Studies in Mathematics* 5, 71 - 79.
- O'Brien, T.C., and Reali, N.C. (1971), 'Logical Thinking - language and context,' *Educational Studies in Mathematics* 4, 201 - 219.
- Paulus, D.H. (1967), 'A Study of Children's Ability to Deduce and to Judge Deductions,' Unpublished Ph.D. dissertation, Cornell University.
- Roberge, J.J. (1969). 'Negation in the major premise as a factor in children's deductive reasoning', *School Science and Mathematics* 69(8), 715 - 722.
- Roberge, J.J. (1970), 'A study of children's abilities to reason with basic principles of deductive reasoning,' *American Educational Research Journal*, 583 - 596.
- Wilkins, M.C. (1928), 'The effect of changed material on the ability to do formal syllogistic reasoning,' *Archive of Psychology* 16, 83.