DOCUMENT RESUME

ED 259 893 SE 045 857

AUTHOR	Mason, Emanuel J.
TITLE	The Development of Logical Thinking in Children.
PUB DATE	Feb 80
NOTE	93p.; Report to the Netherlands Ministry for Pure Science Research (ZWO). Appendices contain some pages of marginal legibility.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Abstract Reasoning; *Cognitive Development; Developmental Psychology; Elementary Secondary Education; Interviews; *Learning Processes; Logic; *Logical Thinking; Memory
IDENTIFIERS	*Netherlands; Piagetian Stages; *Syllogistic Reasoning

ABSTRACT

Reasoning and logical thinking can be defined and explained from different perspectives. Three approaches are reviewed in this report; they are: (1) the logical structure approach; (2) the Piagetian approach of developmental stages; and (3) the information processing or memory approach. Four hypotheses related to these approaches were investigated with children of ages ranging from 5 to 16 years. The subjects were interviewed about their interpretation of propositions and logical arguments. Interviews were designed to explore understanding of the premises and conclusions of syllogisms and the child's understanding of syllogisms in total. The report explains the design of the studies, scoring techniques, and the standardization of the materials that were used. Conclusions are presented for each of the hypotheses; the general pattern of correct responses with the Dutch students who participated in the study was found to be similar to previous results of English-speaking groups. Recommendations for continued and expanded studies are given. An extensive reference list and appendix containing interviews in Dutch and coding forms and scoring instructions in English are also included. (ML)

THE DEVELOPMENT OF LOGICAL THINKING IN CHILDREN

Emanuel J. Mason
University of Kentucky

Report to the Netherlands Ministry for Pure Science Research (ZWO) on research conducted by the author while at the Division of Developmental Psychology, Psychology Laboratory, Catholic University, Nijmegen, the Netherlands, supported by a visitor's grant from ZWO during the period from January 1979 to August 1979.

Table of Contents

I

Research on Reasoning and Logical Thinking

Introduction

Definitions of reasoning or logical thinking tend to be vague. Logicians generally provide a definition of logical thinking that refers to "correct or incorrect" thinking (Copi, 1961), or finding conclusions "on the basis of reasons" (Angell, 1964). Such definitions imply an interest in discovering truth and validity. Psychologists' definitions reflect an interest in the processes of thought. For example, Nolen (1976) defines logic as thought that conforms to some formalized rules. It is the rules that are of concern, not the truth, or validity of the thinking.

Further, psychologists and philosophers have long recognized that a person's logical thinking can be faulty. In keeping with their primary focuses, logicians have been interested in discovery and elimination of faulty reasoning (e.g., Ennis, 1975), while psychologists have been interested in studying the nature of correct and incorrect responses (Evans, 1972; Falmagne, 1975; Nolen, 1976; Piaget, 1953, 1963).

Most psychological approaches to reasoning and logical thinking use traditional Aristotelian logic as a model or shorthand for the thought processes (Johnson-Laird, 1975). Although some psychological research has been done on transitive assymetrical relationships (e.g., Brainerd, 1979), disjunctive relationships (e.g., Roberge, & Flexer, 1979 d) and other logical forms, the greater part of psychological research in the area has involved material implication (Hadar, 1975). In material implication, there is an antecedent statement which follows the word "if" and a consequent which follows the word "then." The antecedent in reality may or may not be related to the consequent, but the implication states that it cannot be the case that the antecedent is true when the consequence is false (Copi, 1961). For example, if p is the antecedent and q is the consequent, then an implication statement would be of the form "If this is a p, then it is a q. The implication, therefore, would not be true if there is a single case of p that is not a q. There are other criteria which must be met. These will be discussed later in this review. The importance of reasoning by implication has been discussed often in relation to mathematical thinking and problem-solving (e.g., Brainerd, 1979; Brown, 1979; Hadar, 1975; Roberge & Flexer, 1979(b)), scientific thinking (Ennis, 1975; Ennis & Paulus, 1965).

The differences between the focuses of the study of logic and the study of logical thinking has led various researchers to warn of over-dependence upon the models of logic to represent thought processes. For example, Evans (1972) questions the psychological relevance of logical validity to a person who is engaging in logical reasoning. Henle (1962) referred to deductive errors due to omitted, misinterpreted premises, an unwillingness to perform the systematic thinking required, or the introduction of knowledge outside of what is provided by the premises. Further, Braine (1978) made a distinction between the natural logic that people use in propositional reasoning, and the standard logic of logicians. In Braine's conception of human thinking, inferential rules are preferred to such tools of standard logic as truth tables, and axioms. Further, in Braine's scheme, connectives such as and, or, and other linguistic properties of propositions are closer to the usual or conversational meaning of the words than is the case in standard logic. On the other hand, Chapman and Chapman (1959) have proposed that people reason "illogically" by logicians' standards, but nonetheless systematically, while Woodworth and Sells (1935) proposed an "atmosphere effect" in which people were considered to evaluate conclusions of arguments according to the atmosphere of the premises. For example, if a major premise contained the word "all," then a conclusion containing the word "all" would tend to be considered valid regardless of the logic involved.

With the warnings regarding using logical structures as models of thought processes, in mind, it would be appropriate to turn to Piaget's theory of logical thinking (Inhelder & Piaget, 1958; Piaget, 1953, 1963). Piaget made a distinction between <u>logic</u> and <u>psychologic</u>. He suggested that the relationship between the two was similar in many respects to the relationship between pure logical notation as symbolic of the structures of thought. As Staudemeyer (1975) has pointed out, this puts a strain on the interpretation of symbolic operators, and the assumption that subjects perform systematic evaluation of propositions or elements in much the same way as logicians. Although many have recognized the difficulty in interpreting Piaget's writings (Brainerd, 1978; Flavell, 1977), the weaknesses in his logical models from a logician's viewpoint have been well documented (Ennis, 1975; Parsons, 1960). Further, Smedslund (1970) discussed the circular relationship between logic as a psychological variable and thought.

Staudenmayer (1975) took the position that an adequate model of human thought would necessarily include the variety of interpretations that a person bases on his or her own environmental and background experiences. Johnson-Laird (1975) suggested that no single logic would be adequate to explain human thinking, a position also supported by Evans (1972). Although alternative logical systems are available (e.g., Freedle, 1977; Rescher, 1976; von Wright, 1957), psychologists generally seem to prefer the traditional Aristotelian models despite the criticisms leveled at those models for the ability to depict human thought processes.

<u>Approaches to the Study of Logical Thinking</u>

Falmagne (1975a) has divided the study of logical reasoning into two camps. The first she identified with the Piagetian approach. The second, essentially a psycholinguistic enterprise, she called the propositional approach. Wason and Johnson-Laird (1968) had anot er formula for splitting the psychological study of reasoning. They traced one line of research to the Wurzburg School at the beginning of the Twentieth Century and gestalt psychology, and the other to behaviorism with its focus on learning.

The available literature in reasoning can be profitably organized a third way. This way recognizes three types of approaches. The first would include those who focus on the interpretations f propositions and structure of standard logic. This approach is best represented by Wason and Johnson-Laird (1972). Then there is the Piagetian approach which posits a series of developmental stages in the development of logic. Finally, the information processing or memory approach blends some of the same tasks and characteristics of the other two with some unique features of its own.

The logical structure approach. The thrust of research with this approach has been to determine the difficulty of reasoning using basic syllogistic structures (e.g., Roberge, 1971). Further, there is an interest in what influences make dealing with logical arguments more difficult. For example, it has been shown rather consistently that modus ponens (MP) is a relatively easy argument for adults and children to judge. Generally, modus tollens (MT) is the next easiest, followed by transivity (TR), denial of the antecedent (DA), and affirmation of the consequent (AC), respectively (Hadar, 1975; O'Brien & Shapiro, 1968; Roberge, 1970, 1971; Roberge & Mason, 1978). However, one study found MT the most difficult form when various permutations of negation in the premise were tried (Roberge, 1971).

Among the influences that affect the use of formal reasoning principles, negation of premises (Roberge, 1971; Roberge & Mason, 1978; Wason, & Johnson-Laird, 1972), and content of the premises (Mason, Bramble, & Mast, 1975; Roberge & Antonak, 1979; Wilkins, 1928), among othercharacteristics have been studied. Generally, it appears that subjects perform best on concrete content with which they are familiar as long as the premises make sense. On the basis of these findings, the present study used fairly concrete and familiar objects in the premises.

Piagetian Approach to the study of logic. Piaget's approach to logical thinking is embedded in his theory of the development of thinking. According to this theory, the child develops increasingly more mature thinking as it grows through a series of stages in which a priori organic structures interact with environmental experiences. Initially, the child develops sensory-motor schemes which become the basis for more mature thought. Through the processes of assimilation, adapting or interpreting external stimulation to correspond to existing mental structures, and accomodation, adapting mental sturctures to the external stimuli, the child's ability to reason develops. This development process continues through adolescence to the point at which formal operations have developed. At this final stage, the child can construct combinations of elements, isolate and manipulate variables, and form mental representations of abstract concepts and events (Pitt, 1976). Thus, all possibilities for a set of elements can be explored. The adolescent is not conscious of this system of possibility, but does them naturally (Piaget, 1953).

Contrary to the logical structure approach just discussed, Piaget was not interested in improving the structure of formal logic upon thinking and reasoning. Rather, his interest was in describing the thought processes as they existed in reasoning on various tasks. Piaget used the notation of formal logic to describe these thinking processes (Piaget, 1953). This may be one of the reasons that Piaget's models for logical thinking have been criticized as illogical (Brainerd, 1978; Ennis, 1975; Parsons, 1960).

The core of Piaget's formal operations stage at which complete logical thought becomes possible are 16 binary logical operations formed by the elements p and q (shown in Table 1), and the INRC group. The INRC group is a group of operations in the mathematical sense and meets the four requirements of such groups: 1) closure - when elements of the group are combined the result is always an element of the group; 2) associativity - when three elements are combined, the results are the same no matter what order of combination is used; 3) identity - there is an element of the group that does not change other elements when combined with them; and 4) inverse - for each element of the group there is another element of the group that when combined with it results in identity (Brainerd, 1978). The letters of the INRC group represent the following operations:

Identity: Application produces an equivalent representation of the same proposition
(i.e. $I[(p \cdot q) \vee (\bar{p} \cdot q) \vee (\bar{p} \cdot \bar{q})] = (p \supset q)$)
Implication — Implication

Negation: changes both signs and conjunctions
(i.e. $N\,(p \cdot q) = p \vee \bar{r}$)
Conjunction — Disjunction

Reciprocity: Changes the signs of a proposition
(i.e. $R(\bar{p} \cdot q) = p \cdot \bar{q}$)
Negation of Reciprocal Implication — Non-implication

Correlation: changes the connections or conjunction in a proposition
(i.e. $C(p \vee q) = p \cdot q$)
disjunction — conjunction

According to Piaget all children pass through the stages of development to reach formal operations. The speed or duration of this passage may be somewhat dependent on environmental factors, but the order will not change. Further, adults reaching the formal operational stage will be capable of performing these operations regardless of background or culture. However, they may be more able to demonstrate them in some areas than others (Piaget, 1972). On the other hand, the operationalization of formal operations is not as evident as Piaget suggested once one goes beyond the tasks that Piaget and his followers have used (Martorano, 1978).

Table 1

Piaget's 16 binary propositions formed by p and q[a]

	Notation	Name of proposition
1.	$p\cdot q$	conjunction
2.	$\bar{p}\cdot\bar{q}$	conjunctive negation
3.	$p\cdot\bar{q}$	Negation of Reciprocal Implication
4.	$\bar{p}\cdot q$	Non-implication
5.	$(p\cdot q)\ v\ (\bar{p}\cdot\bar{q})$	Equivalence ($p\equiv q$)
6.	$(p\cdot q)\ v\ (\bar{p}\cdot q)$	Affirmation of q
7.	$(\bar{p}\cdot q)\ v\ (p\cdot\bar{q})$	Affirmation of p
8.	$(\bar{p}\cdot q)\ v\ (\bar{p}\cdot\bar{q})$	Denial of p
9.	$(p\cdot\bar{q})\ v\ (\bar{p}\cdot\bar{q})$	Denial of q
10.	$(p\cdot\bar{q})\ v\ (\bar{p}\cdot q)$	Nonreverse Implication
11.	$(p\cdot q)\ v\ (\bar{p}\cdot q)\ v\ (\bar{p}\cdot\bar{q})$	Implication ($p\supset q$)
12.	$(p\cdot q)\ v\ (p\cdot\bar{q})\ v\ (\bar{p}\cdot\bar{q})$	Reverse Implication ($q\supset p$)
13.	$(p\cdot q)\ v\ (p\cdot\bar{q})\ v\ (\bar{p}\cdot\bar{q})$	Disjunction ($p\vee q$)
14.	$(p\cdot\bar{q})\ v\ (\bar{p}\cdot\bar{q})\ v\ (\bar{p}\cdot q)$	Nonconjunction (p/q)
15.	$(p\cdot q)\ v\ (p\cdot\bar{q})\ v\ (\bar{p}\cdot q)\ v\ (\bar{p}\cdot\bar{q})$	Complete Affirmation
16.	$\sim[(p\cdot q)\ v\ (p\cdot\bar{q})\ v\ (\bar{p}\cdot q)\ v\ (\bar{p}\cdot\bar{q})]$; or 0	Complete Negation

In this Table and the remainder of the paper, the following symbols will be used:

A line over a letter means negation (i.e., $\bar{p}$ should be read "not p"); the symbol $\supset$ means implication ($p\supset q$ should be read "p implies q"), and the $\therefore$ symbol means the conclusion follows, and should be read "therefore"; the symbol v means "or," and p v q should be read "p or q", finally, the symbol $\sim$ stands for negation.

Information processing and memory approaches to the study of logic. The information processing approaches, and there are several (e.g., Corso, 1967; Simon and Newell, 1974), are not as tied to the structures and notation of formal logic. However, certain representatives of this viewpoint do rely on traditional logic to some degree (e.g., Braine, 1978). Simon (1967) has argued that the logic essential for this approach only depends upon "the ordinary logic of declarative statements" (p. 20). Some researchers using this approach try to identify efficient algorithms for problem-solving while others using the same approach may be interested in simulating and describing the structures or mechanisms of reasoning. Therefore, information processing as it is discussed here represents a much less homogeneous approach than the first two that were discussed in this section. What makes representatives of this group similar is the assumption that human beings function on their environment in serial fashion using limited "built in" computational or operational abilities which allow them to solve problems of varying complexity.

Most information processing approaches require some notions of input of data, some central processing, storage and recall capabilities, and expressive or decoded "output." Braine (1978) discusses four essentials in the central processing and storage for logical reasoning, a comprehension mechanism, a mechanism for selecting steps, heuristics for planning and argument, and some short-term-memory or "computing" space. Thus, some processing mechanisms in addition to logical operators are required in going from reasoning to behavior. Case (1978) emphasized such things as practice, feedback, cue-highlighting, and the size of the space in working memory. Brown and De Loach (1978) referred to metamemory as contributing to the development of thinking. Other researchers have focused on task analysis (e.g., Mayer, 1978; Trabasso et al., 1978), the role of memory (e.g., Potts, 1978), representation of different kinds of tasks and objects (e.g., Mayer, 1978), and a variety of other mechanisms and processes

(Erickson, 1978). The present research is more related to the information processing research which focused on development of models to represent or explain reasoning and thinking, than those designed to produce effeicent problem-solving.

Comparison of the Three Approaches

By far, the Piagetian approach has been dominant in the psychological study of logic in the last forty years or so, partially because Piaget himself has been so productive a writer and researcher. Most developmental psychological research in this area begins with an affirmation, rejection, or query of something Piaget found or wrote. Similarly, most critical reviews of the literature focus on what Piaget did, said, or omitted, more than upon any other collection of work. Therefore, the present comparison will necessarily focus disproportionately upon Piaget.

The first point of comparison between the three approaches is a philosophical one. That is, each of the three approaches varies in its conception of logical thinking in people. The first group is routed in traditional logical structures, validity, truth functions, syllogisms such as modus ponens, modus tollens, and other concerns of traditional logic. Implicit in this approach is the assumption that good logical thinking (i.e. valid thinking) is what humans should strive toward. That is, the goal of logical thinking is to be able to chain propositions in ways in which their application will lead to logically correct reasoning. Not only is there a descriptive component to this research which assesses children's and adults' abilities to recognize, analyze or apply arguments under varying conditions, but also, there is a teaching interest represented by these researchers. That is, through the use of training which may focus upon language structures, Venn diagrams, logical truth tables, propositional tasks, and so on, a person's reasoning can be made more logical. Further, the assumption that humans are not naturally logical seems implicit in this approach.

This is different from the other two approaches, Piaget's and information processing, which seem to assume that there are existing logical structures which are revealed by research. Therefore, implicit in these approaches is the assumption that logical structures occur naturally in human thought. Further, there is a difference in the degree to which these resident structures resemble the notation system and structures of traditional logic with Piaget's being somewhat more similar to formal logic than the information processing view as it is broadly construed here. Although, as indicated earlier, many researchers who are here placed in the information processing category do research the processing of formal syllogisms (Braine, 1978; Mayer, 1978; Revlin & Leirer, 1978), their focus tends to be on the thought processes that can be identified rather than the performance of valid thinking.

Further distinctions between these approaches can be made on the basis of tasks (Danner & Day, 1977; Gelman, 1979; Martorano, 1978; Trabasso, Isen, Dolecki, McLanaham, Riley, & Tucker, 1978), degree to which validity of the logic is essential (Brainerd, 1978; Ennis, 1975; Piaget, 1953; Wason & Johnson-Laird, 1972), and the focus on linguistic matters (Evans, 1973; Roberge, 1978; Wason & Johnson-Laird, 1972). Piaget minimizes the importance of linguistic factors since the logical structures he proposes are organismic in nature and would therefore transcend language and culture.

The Present Studies

The present studies were designed to investigate hypotheses related to the three approaches identified in the review of the literature. Only the hypotheses which have been tested so far will be reported here. Analysis of the data is continuing and will be reported on subsequently.

Hypothesis 1: There will be differences across ages in the kinds of verbal explanations made by children and adolsecents in judging logical arguments.

Hypothesis 2: The Class inclusion will be related to provision of correct judgments of the truth of logical arguments containing implication.

Hypothesis 3: There will be an inverse relationship between response time and correctness of reasoning which reduces as children reach adolescence.

Hypothesis 4: The class meaning of a logical argument as reflected by subjects' explanations of their reasoning will correspond to four components of material implication $(p \cdot q, \bar{p} \cdot q, \bar{p} \cdot \bar{q}, \& \ p \cdot \bar{q})$.

To test these hypotheses, two data collection efforts were mounted. Children and adolsecents who participated ranged in age from five to 16 years of age. Subjects were interviewed about their interpretations of propositions and logical arguments, and these interviews were scored using several standards. The design of these two studies, including scoring, and standardization of the materials is described in the following section.

II

Methods

Of the various approaches that have been used to study the development of logic, each has posed unique difficulties. For example, some have used formal syllogisms to explore children's reasoning. However, this kind of logic may not correspond to the processes of thinking (Braine, 1978; Henle, 1962). A rather thorough discussion of this is provided by Osherson (1975, p. 16-25). Further, much of this research has involved the use of group paper-and-pencil examinations and required respondents to indicate whether the argument forms are valid or not (e.g., Hadar, 1975; Roberge, 1970; Roberge & Mason, 1978; Shapiro & O'Brien, 1970). The group administered paper-and-pencil approach is useful for testing large numbers of subjects, but does not offer an opportunity for the researcher to explore the meaning of the logic with the child. It would be difficult, for example, using this approach to determine whether subjects are giving correct answers to logical arguments because they have memorized an algorithm, have a partial understanding that is just enough to give the correct response, or completely understand the logic underlying the argument. This can be shown simply using the following syllogism:

If all men are mortals
and Socrates is a man.
Then Socrates is a mortal.

A child who is asked about the validity of this argument might respond affirmatively because he or she understands the premises in any of the following ways:

a) There are men and there are mortals.

b) There are mortals and there are men.

c) Men must be mortals but mortals do not have to be men.
Since Socrates is a man, he must also be a mortal.

d) To be mortal implies being a man also. For Socrates to be a man he must be a mortal.

e) Men are mortals, and my father is a mortal, so Socrates must be like my father.

Although all of the understandings listed above would probably lead a child to say that the argument is correct, (c) shows the most complete understanding, and (d) is clearly incorrect. The remaining understandings are not complete, although they would lead the child to respond correctly.

Individual interviews have been used to explore children's understanding of logic with some success (e.g., Falmagne, 1975; Inhelder & Piaget, 1958; Kodroff & Roberge, 1975; Kuhn, 1977). Interviews produce some problems, however, that are less troublesome with paper-and-pencil group administered instruments. These problems involve administration and scoring. Administration problems can be reduced by training interviewers not to influence the behavior of the respondent other than to encourage responses. Difficulties in scoring and interpreting interview data can also be reduced by training scorers and providing well-defined guidelines for interpreting and classifying responses. Therefore, such materials and techniques were developed for the present research, and will be described in the following sections along with the subjects studied and the procedures used.

DESIGN OF THE INTERVIEWS

The interviews were designed to explore two different kinds of information. The first interview explored the child's understanding of the premises and conclusions of syllogisms, the second explored the child's understanding of syllogisms in total. The two interviews were based on four syllogism forms: Modus Ponens (MP) ($P \supset Q$, $P \therefore Q$), Modus Tollens (MT) ($P \supset Q$, $\overline{Q} \therefore P$), Affirmation of the Consequent (AC) ($P \supset Q$, $Q \therefore P$), and Denial of the Antecedent (DA) ($P \supset Q$, $\overline{P} \therefore Q$).

Truth table analysis will reveal that MP and MT are valid arguments while AC and DA are invalid.

The two forms are described in the following sections. Since the first form was designed to compare understanding of premises individually and in the context of a whole argument, it was identified as the Part-Whole Form (Form PW). The second was only designed to study the meaning of the whole argument, and is therefore called the Complete Argument Form (Form CA).

FORM PW. This form was developed to explore relationships between understanding of the superordinate and subordinate clauses in the major premise and logical processing of the whole argument. As part of the understanding of the conditional implication, the child's understanding of the implicit class inclusion relationship in the premises was explored. Further, the form was designed to explore two different content dimensions at the same time, concrete and verbal. Since there was a possibility that experience with a partial presentation might influence a child's performance on the total argument, and vice versa, the interview was balanced for presentation of part first or whole first, and for the two content dimensions producing four parts as follows:

Part I - Whole first, concrete.

Part II - Part first, concrete.

Part III - Whole first, verbal.

Part IV - Part first, verbal.

Items were similar to the three-choice response paradigm used in other reasoning research (Ennis & Paulus, 1965; Hill, 1961; Roberge, 1970). That is, subjects were required to state whether they thought the conclusion was true or invalid with the following three choices: Yes (it is true), No (it is not true), Maybe (it may be or may not be true depending upon circumstances). However, in the present interviews subjects were asked to give reasons for their answers.

In each concrete presentation the examiner used two trays (clear plastic 18.0 cm x 11.6 cm x 3.2 cm). One had ten small plastic animals in it (varying sizes ranging in height from about 2 cm to about 5.5 cm and including one each of a lion, deer, antelope, rhinocerous, elephant, kangeroo, black cheeta, white polar bear, moose, and a white dog; the dog statue came from the standard 1959 Stanford-Binet Intelligence Test Kit.). The other tray (identical to the first) contained 12 green blocks (2.5 cm^3) from the Stanford-Binet Kit. In the animal tray the dog was placed separately from the cluster of other animals at one end of the tray. Further, outside of the trays, the examiners had an unpainted wood block of 2.5 cm x 2.5 cm x 8.0 cm, and a second dog identical to the one in the animal tray. For the verbal presentation, cards (16 cm x 10 cm) containing drawings of a tree and a football (soccer ball) were used. Illustrations of the materials described here are in the appendix.

The format of the whole-first presentation (Parts I and II) for each item is illustrated below using an example from the animals-dog concrete content.[a] (Of course, the items were all presented in Dutch).

1. Examiner puts tray A before the subject and says[b]:

IF ALL DOGS ARE ANIMALS	(as words are spoken, E points to dogs and then rest of animals).
AND THIS IS A DOG	(E shows S the statue of other dog that had been kept out of sight).
THEN, I SHOULD PUT THIS	
WITH THE ANIMALS	(E places dog in box near dogs).

---

[a] The reader can easily construct the remainder of the items by substituting the "cubes and blocks," "trees and plants" and "balls and round things" in the premise of the four syllogism forms. Table 2 should be of some help here.

[b] The words the examiner says are in upper case letters; instructions to the examiner are in lower case.

A. DOES THAT CONCLUSION FOLLOW?

If the child does not give a standard answer, (yes, no, maybe) E says:

REMEMBER, YOU CAN ONLY SAY YES, WHICH MEANS IT MUST FOLLOW:

NO, WHICH MEANS IT CANNOT FOLLOW, OR

MAYBE, WHICH MEANS IT MAY BE TRUE SOMETIMES, BUT YOU HAVE NOT BEEN TOLD ENOUGH TO BE CERTAIN.

When S gives a standard answer, E asks:

B. WHY DO YOU SAY THAT?

After S's answer, E continues:

C. TELL ME WHAT IT MEANS TO YOU WHEN I SAY,

ALL DOGS ARE ANIMALS.

When S responds, E asks:

D. ARE THERE MORE DOGS OR ANIMALS?

When S responds, E continues, NOW TELL ME WHAT THIS MEANS:

E. THIS IS A DOG (E points to larger dog in tray)

THEN, I SHOULD PUT THIS WITH THE ANIMALS.

The following is an illustration of an item from the part-first presentation (Parts II and IV). Again the animals-dog content is used for the illustration.

4. Examiner puts tray A before the subject and says:

A. TELL ME WHAT IT MEANS TO YOU WHEN I SAY,

ALL DOGS ARE ANIMALS.

When S responds, E asks:

B. ARE THERE MORE DOGS OR ANIMALS?

When S responds, E continues, NOW TELL ME WHAT THIS MEANS:

C. THIS IS NOT AN ANIMAL (E holds up dog)

THEN, I SHOULD PUT THIS WITH THE DOGS.

After S responds, E says,

NOW LISTEN TO THIS ONE.

IF ALL DOGS ARE ANIMALS (E points to dogs then animals as words are spoken)

THIS IS NOT AN ANIMAL

THEN, I PUT THIS WITH THE DOGS.

D. DOES THAT CONCLUSION FOLLOW:

If the child does not give a standard answer, (yes, no, maybe), E says:

REMEMBER, YOU CAN ONLY SAY YES, WHICH MEANS IT MUST BE TRUE.

NO, WHICH MEANS IT CANNOT FOLLOW. OR-

MAYBE, WHICH MEANS IT MAY BE TRUE SOMETIMES. BUT YOU HAVE NOT BEEN TOLD ENOUGH TO BE CERTAIN.

When S gives a standard answer, E says:

E. WHY DO YOU SAY THAT?

The content by syllogism matrix in Table 2 shows the schematic used to construct the items. There were two items in every part for each kind of content and syllogism. Since there were four syllogisms, and two kinds of content crossed in each part. There was a total of eight items in a part, and 32 items in the whole interview.

Table 2

Content matrix for contructing test items for Form PW[a)]

	Concrete Content				Verbal Content			
	Part - First		Whole - First		Part - First		Whole -First	
	Dogs/ Animals	Cubes/ Blocks	Dogs/ Animals	Cubes/ Blocks	Trees/ Plants	Balls/ Round things	Trees/ Plants	Balls/ Round things
MP	1	5	1	5	1	5	1	5
MT	2	6	2	6	2	6	2	6
DA	3	7	3	7	3	7	3	7
AC	4	8	4	8	4	8	4	8

a) Numerals in cells of table used to randomly order items.

THEN IT IS ....

A. E then says: DOES THAT FOLLOW?

If the child does not give a standard answer, E says:

REMEMBER YOU CAN ONLY SAY YES, WHICH MEANS IT MUST BE TRUE,

NO, WHICH MEANS IT CANNOT BE TRUE, OR

MAYBE, WHICH MEANS YOU HAVE NOT BEEN TOLD ENOUGH TO BE CERTAIN.

B. When the child has given an answer, E asks:

WHY DO YOU SAY THAT?

From the arguments (MP, MT, DA, AC in two forms each), shape of the minor premise block (circle or cube), and color (yellow or green), a total of 32 items were constructed. This is illustrated in Table 3. As with the last from, the items were ordered randomly in the interview. In addition, the 32 items were divided in half, so that 16 items could be administered in an administration. This required the child to attend for periods of about 10 to 20 minutes, usually about 12 minutes.

Table 3

Content Matrix for forming Test Items for Form CA[a)]

Syllogism form	circles		cubes	
	yellow	green	yellow	green
MP 1.	1	9	17	25
2.	2	10	18	26
MT 1.	3	11	19	27
2.	4	12	20	28
DA 1.	5	13	21	29
2.	6	14	22	30
AC 1.	7.	15	23	31
2.	8	16	24	32

a) Numerals in cells used to randomly order items

a) Numerals in cells used to randomly order items

For research questions not included in the present discussion, half of the subjects were administered the items with the content of the minor premise placed separately from the remainder of the blocks on the carpet using the red tape strip as a barrier. The rest of the subjects were shown the carpet with content of the minor premise included with the other blocks.

ADMINISTRATION

The interviews were administered during the regular school day individually by graduate students in developmental psychology from the University of Nijmegen. Interviews were recorded on cassette tape for later analysis. Examiners had specific instructions describing what to say. Copies of these instructions are included with the interviews in the Appendices of this report.

Children and adolescents were interviewed individually in a quiet room outside of their regular classroom. The setting varied slightly in each school, however, generally the examiner sat at a table directly across from the child with the test materials arranged between them and the cassette tape recorder (with built-in microphone) off to one side.

Each session was designed to take no more than 20 minutes. Generally one or two parts of Form PW were administered during a single session. With Form CA, half the items (16) administered in a typical session, however, with older or more able subjects, occasionally all items could be administered within the 20 minute session. The order of administration of the parts for Form PW, and the halves for Form CA was randomly determined.

Each child's interview was preceded by a series of instructions and some practice examples. The practice examples were devised to give the child some experience with the interview format and the use of the terms YES, NO and MAYBE format for responding. The practice problems did not contain any of the formal arguments used in the interview. The practice examples and accompanying instructions are included in the Appendices.

SUBJECTS

Subjects for the studies described here were students in nursery, primary, and secondary schools in the region of Nijmegen, the Netherlands. Kindergarten, 2nd, 4th, and 6th grade children were randomly sampled as were students from the second year of secondary school and administered Form PW. Secondary school students were selected in the ratio for HAVO, MAVO, and VWO at that level. Although it should be stated that at the particular secondary school which supplied the sample, a larger portion of the students go on to a VWO program than is the case nationally.

Table 4 summarizes the characteristics of the students.

Table 4

Characteristics of Students who participated[a]

Grade	Number of boys	Number of girls	Average age (yrs-months)
Kindergarten	5	5	6-0
1	7	5	7-0
2	6	5	8-2
3	6	6	9-1
4	6	6	10-2
5	6	6	1 [illegible]
6	6	6	12-3
Secondary 1	9	9	13-0
Secondary 2	6	5	15-8
	57	53	

a) Kindergarten, second, fourth, sixth graders and Secondary 2 students responded to Form PW, and first, third, fifth graders, and Secondary 1 students responded to Form CA.

TABULATING AND SCORING

In this section, three stages in the processing of the data collected in the interviews are discussed. The first is the transcription of the data from the cassette tapes to handwritten protocols. This was done to facilitate scoring, and to make a record of each interview. The second stage involved the actual scoring of the child's assessment of the syllogism. The third stage included the classifying of the child's verbal explanations into categories of verbal and logical explanations.

Transcription. All interviews were recorded using a cassette tape recorder (Philips, N2215 automatic portable cassette). The recordings were made so that there would be an accurate record of each interview. However, it was felt that scoring from the recorded interview would be difficult because responses to specific questions cannot be addressed easily on a cassette tape. Also, when scoring, the time required to play a tape in order to listen to one testing session would be greater than the time needed for reading a transcript of the session if it were in written form. For that reason, the children's responses were transcribed. Forms were used which were designed to facilitate scoring and at the same time reduce the amount of writing necessary to record the essence of the child's responses.

For Parts I and III of Form PW in which the whole syllogism was presented first, responses were transcribed using the following format for each item:

A.	YES	NO	MAYBE
B.			
C.			
D.	P		Q
E.			

The transcription format used with each item of Parts II and IV of Form PW was:

A.			
B.		P	Q
C.			
D.	YES	NO	MAYBE
E.			

Note that in both formats, the letters correspond to the order of the questions asked.

Transcribers used these forms in the following way. They circled the child's response to the question about whether the argument was true (YES, NO, or MAYBE), and to the class inclusion question (subordinate (P) or superordinate (Q)). For the remaining sections, the transcribers were told to record in the child's words what they thought were the essentials of the child's response. They were told to omit any irrelevant verbalization, and to keep in mind that they would be required to score the child's understanding of his or her response on the basis of what was written in the transcripts. Since there was thus an element of judgment required on the part of the persons doing the transcription, the question of inter-transcriber reliability was raised. This question was investigated and will be discussed further in the section on reliability.

The format for transcribing responses to each item in the Form CA interviews is shown below.

A.	YES	NO	MAYBE
B.			
C.			

In the section labelled "A" the transcriber circled YES, NO, or MAYBE to indicate the child's response. In section B, the essentials of the child's response were noted. Finally, in section C, the time elapsed in seconds from when the examiner began the question to the moment when the child responded with "YES," "NO," "MAYBE." The time was recorded using a hand-held stopwatch with a sweep second hand, and was recorded to 1/10 second accuracy. The reliability of the timing recorded by transcribers was studies as well as the reliability of the transcription and will be discussed later.

Scoring. In Form PW two questions in each item, and in Form CA one question in each item could be scored directly. That is, the items for which the child was responding with a one word answer that was either correct or incorrect could be scored directly without further interpretation. These were scored using a key. An incorrect answer was credited as "0" and a correct one as "1." The total score was the sum of the 1's and 0's.

Classification of Verbal Explanations. Two instruments were used for classifying verbal responses. One, the Kodroff-Roberge system, focused on categories of verbal explanation based on techniques used with some success to classify children's verbal reasoning (Antonak & Roberge, 1978; Kodroff & Roberge, 1975). The second classification system (C-L) was directed more at the class and logical relationships expressed by the children. Copies of the forms designed for these two classification systems may be found in the Appendices. The specific instructions given to the scorers are also included in the Appendices. Scorers practiced the two systems until they felt that they were familiar enough with them to use them. Then inter-rater reliability was determined.

Kodroff-Roberge System. The K-R system used in the present studies was adapted somewhat from the instrumentation of Kodroff and Roberge. To begin with, the original system had six categories of verbal behavior. The first four categories were used as proposed by Kodroff and Roberge. They included verbalization deficits (an inability to discuss the logic behind a deficit, e.g., "I don't know"), memory deficits (repeating a premise incorrectly, losing train of thought or becoming confused, e.g., "Because this is a plant and this ... I forgot."), arbitrary explanations (bringing in facts not given in the premises, e.g., "I had a dog like that once and he was an animal"), and structural deficits (response indicating an awareness of conditional logic but not adequate conditional reasoning, e.g., "There are dogs and there are animals."), patterned explanations, the sixth K-R category (response indicating that the premises are organized in a pattern, at least three terms in the item are included ("e.g., Dogs and animals. This is a dog and an animal"). The fifth category, conditional reasoning was changed slightly from the Kodroff-Roberge meaning. In the current system a response was only scored as conditional reasoning if the response showed some understanding of conditional reasoning that was incorrect. The child must have included the words "If" and "then" in the response (e.g., "If this is a plant, then it must be a tree.")

Two other changes were made in the original version of the K-R scoring system for the present studies. One was the addition of a category called other for verbal explanations which did not fit the other categories. The second was the addition of a category entitled good logic. To be classified as good logic the response must not have fit in any of the other categories, must have shown mature interpretation of the premises and the implications involved, and the proper relationship between the superordinate and subordinate terms in the major premise (e.g., "Since all dogs are animals, there can be no dogs that are not animals. Therefore, this is a dog, so it is also an animal").

C-L Coding Form. The C-L coding form was devised to identify the kinds of logic represented in subjects' responses. There were two kinds of ratings made of each response using this system. The first kind was classification, the second was logical relationships.

Classification relationships referred to the subordinate and superordinate characteristics of the terms in the premise. For example, in the premise "If all trees are plants," trees are part of the group of things called plants. However, the child's verbalization may reflect an understanding of the premise with trees and plants as separate and equal concepts, or some trees as plants but others not, etc. These kinds of relationships were coded on the class relationships side of Form C-L. To illustrate, "If all trees are plants," may be understood by the child to mean any of the following:

a) There are equal numbers of trees and plants ($P=Q$).

b) There are more trees than plants ($P>Q$).

c) There are less trees than plants ($P<Q$).

d) Trees and plants both exist as separate unrelated concepts ((P) (Q)).

e) Trees and plants are parts of the same concept ((PQ)).

f) Some trees are plants and some plants are trees ((P(Q)).

g) Trees are part of the larger group of plants (((P) Q)).

h) Plants are part of the larger group of trees (((Q) P)).

These categories of meaning formed the class relationship categories of the C-L coding form.

The logical ratings on the other hand were based on Piaget's suggestions about the components of logical implication (see Ennis, 1975; Inhelder & Piaget, 1958). To these were added a few categories that were to be important from the 16 binary propositions formed by P and Q (Inhelder & Piaget, 1958). The following categories were checked when evidence for them could be found in the child's response:

a) P : The subordinate (P) clause or term is affirmed.

b) Q : The superordinate (Q) clause or term is affirmed.

c) $\bar{P}$: The negation of P is affirmed.

d) $\overline{Q}$: The negation of Q is affirmed.

e) $P.Q$: Both P and Q are recognized.

f) $\overline{P}.Q$: The negation of P and Q are affirmed.

g) $\overline{P}.\overline{Q}$: The negation of both P and Q are affirmed.

h) $P.\overline{Q}$: The existence of P without Q is affirmed.

i) $N(P.\overline{Q})$: The existence of P without Q is denied.

j) $P \rightarrow R$, $R \rightarrow Q$: Evidence of transitive reasoning (from P to R and R to Q) is given.

Logical relationships were checked if they were implied by the child's response. For example, if the response was "When these are dogs, these are animals because dogs are animals," the item would be scored as P.Q (where P represents the subordinate and Q the superordinate categories in the premises). But also, since it is implied that not all animals are dogs, or that dogs constitute a subordinate group of the category of things known as animals, the logical relationships $P.\overline{Q}$ and $\overline{P}.\overline{Q}$ could be inferred.

These kinds of inferences on the part of the scorers required that the scorers be trained. Further, with such a procedure, inter-rater reliability should be established to demonstrate that different raters can rate a protocol with some consistency, or to put it another way, that different people can see the same things in a given protocol.

The logical and the class sides of the C-L coding form contained a column for "other," which was used when a response could not be placed in one of the categories. Scorers placed a check mark in the appropriate column for each of the class and logical relationships that could be inferred from a response.

Scorers were trained to use coding Form C-L by coding two or three protocols together after they had read and discussed the procedures. When they felt that their ratings were agreeing substantially, they rated the reliability protocols. After reliability was established, the remainder of the protocols were scored.

RELIABILITY

Reliability refers to the degree to which a test or procedure yields consistent information. The present section discusses reliability estimations of various types for each of the procedures used. The reliability of the transcription procedures are discussed first. Then the reliabilities of the total reasoning scores obtainable from Forms PW and CA are described. Following this, the Kodroff-Roberge verbal classification system is discussed from the point of view of reliability of the raters. Finally, reliability estimations for the C-L coding Form are presented. These reliability studies were made using the same three third year graduate students in developmental psychology at the Catholic University of Nijmegen, who collected the data, then transcribed and recorded it. Details of the reliability studies are reported elsewhere (Mason, Haars, Dingemans, Bierman van Duuren, 1979).

A. <u>RELIABILITY OF TRANSCRIPTIONS</u>

<u>Form PW</u> - A random sample of five cassette tapes was chosen from a pool of 112 taped interviews, one tape from each of the five grades tested. Each cassette contained two parts of the four part form. Two of the tapes chosen contained parts I and II, the remainder contained Parts II and IV. The raters averaged about 12 to 15 words per response for the 120 responses they each transcribed on the five tapes. There were no significant differences between the raters in the number of words they transcribed. Further, the average correlations between the three raters for the number of words they transcribed ranged from .77 to .99 and suggested a high degree of consistency between raters in terms of the number of words transcribed.

However, agreement in the number of words does not suggest agreement in the meaning of the words or the words used. The degree to which the raters agreed upon the meaning of their transcriptions was investigated by comparing their ratings using the K-R and C-L scoring systems, and is discussed later in this report. In regard to the degree to which there was agreement in the actual words used, and the order in which they appeared in the transcripts, a count was made with each rater paired with every other one. An average of about 10 to 13 of the same words were written in the same order by each pair of raters. Further analysis revealed correlations between pairs of raters ranging from .86 to .99 suggesting a high degree of consistency. It was concluded on the basis of these data that there was a high degree of consistency in transcribing from the recorded interviews using the instructions for the raters that were given.

Form CA - A random sample of eight taped interviews was chosen from a pool of 54 interviews. Two interviews were chosen from each of the first grade, third grade, fifth grade, and first year secondary school (seventh year of formal education). Again, the same three scorers were used. Two transcribed an average of about 19 words per item, the third averaged about 12.5 words. Although the difference between the third and the first two was statistically significant ($t_{\alpha/2}$ = 5.13 and 4.26, respectively, df = 510, $p < .001$), this lack of agreement between one of the raters and the other two does not suggest that the shorter transcriptions were necessarily less adequate than the longer ones for extracting meaning. This would be revealed in the application of the C-L and K-R scoring systems to the data if it is the case.

The degree to which the number of words transcribed correlated among the three transcribers was investigated. Correlations ranged from .70 to .96 suggesting that even though there was a difference in the number of words recorded among observers, there was moderate to high agreement on the items for which more transcribed words were required to determine the meaning of the child's response. Similar correlation patterns were found for the correlations among pairs of raters.

Using the hand held stopwatch, each rater averaged about 12.5 seconds in response time measured for each item. There were no significant differences amond raters, and the response times correlated .96 between pairs of raters. That is, the data suggested that each rater timed the various responses almost identically indicating a high reliability for this measure.

B. RELIABILITY OF REASONING SCORES

Each form PW and AC was capable of yielding a total reasoning score. This total score was computed by summing the correct responses (YES, NO, or MAYBE) for the question about whether each argument was true. When the child was correct, a score of 1 was assigned to the item, and when incorrect, a score of zero was assigned. The total reasoning was a summation of all the 1's and 0's. The internal consistency approach was used to investigate the reliability of this total score.

Two internal consistency formulas were used, the Spearman-Brown, and one Cronbach (1960) suggested as a more adequate method. For Form PW, both methods yielded reliability estimates of .88. For Form CA, both methods provided estimates of .93. It was concluded that internal-consistency reliability was adequate for these instruments.

C. RELIABILITY OF THE REVISED KODROFF-ROBERGE SYSTEM

Since the Kodroff-Roberge system is a method for classification of verbal responses to questions requiring reasoning, the reliability question with this system concerns the consistency with which different raters use the response categories. Raters made their ratings from the hand-written transcriptions of the taped interviews. Each rater used his or her own transcription since it was felt that in a practical scoring stiuation, the rater would use his or her own transcription. Furthermore, the transcribers were given the instructions to only transcribe the parts of the child's verbal response that they felt were necessary to show the child's understanding of the reasoning principles it used. Therefore, the rater's classifications on this and the C-L coding forms could be considered evidence of content validity of the transcripts if the inter-rater reliability were high.

The determination of inter-rater agreement with categorical data is somewhat more complex than simply counting the number of times two or more raters agree and the number of times that they do not. Many of the problems involved have been discussed by Cohen (1960), Light (1971), and Robinson (1957). Basically, the problems involve a distinction between demonstrating patterns of agreement versus patterns of association. The methods suggested by Light (1971) and Cohen (1960) were used with the present data.

For each of the Forms PW and CA, a three dimensional matrix of agreement was constructed. There was one dimension for each rater. The eight categories of ratings were placed along each dimension to form an 8 x 8 x 8 cube containing 512 "cells." Using these cubes, coefficients of agreement were calculated for all three raters using Light's procedure (1971, p. 365-370), and for pairs of raters using Cohen's (1960) method.

Form PW - The overall coefficient of reliability (Km) of the ratings given by the three raters was computed to be .74 with the Light formula. Inter-rater reliabilities (K) were computed to be .75, .77, and .70, respectively, for pairs of raters. These reliabilities were considered moderate and were statistically significant ($p < .05$). It should be noted that there was somewhat greater agreement between raters for the responses of children who were younger (Kindergarten and second grade) and those who were older (second year of secondary school). At those ages raters said the children seemed better able to express their understanding, or lack of it. The fourth and sixth graders may have been in a transitional stage between more mature logic (or less), and between more (or less) mature expressive ability. This, of course, is a broader research question than can be discussed here, and will be the subject of later reports. However, it should be recognized that if at some ages children are not sure what they mean to say about their logic, or are not sure how to explain it, this could affect the consistency of those trying to rate what they say they mean.

In addition, the percent of agreements between pairs of raters were computed. These ranged between 78.13% and 83.75%. This was considered substantial, and to reflect further on the consistency between raters. A total of 160 responses was rated in the five tapes by the three raters.

Form CA - The overall coefficient of reliability (Km) for the three raters was computed to be .70. For pairs of raters inter-rater reliability estimates ranged from .66 to .74 using Cohen's method. Further, the pairs of raters agreed on from 77.08% to 82.92% of the 240 responses they considered. Again, these reliability estimates were considered moderate, and seemed to be influenced by unclearness of responses by the middle age range.

The moderate reliabilities obtained for this form are slightly lower than those reported by previous users of the original Kodroff-Roberge system. For example, using Cohen's method, Kodroff and Roberge (1975) reported a reliability of .86 for a pair of raters. However, their data were from children who were younger than those in the present studies (first to third grades). The moderate reliabilities obtained for the three observers in the present study suggest that researchers using the K-R system with a broad age range should not try to interpret small differences, especially among fourth to sixth graders, because such differences may not be reliable even if they do approach statistical significance.

D. RELIABILITY OF THE CLASS - LOGIC SCORING SYSTEM

The same three raters and tapes were used to study the reliability of the C-L scoring system. Since the numbers involved were small (10 parts for Form PW, and 8 tapes for Form CA), the data from the two were combined. Even so, there were not enough ratings in some of the categories to produce dependable reliability estimates. However, it should be noted that these categories were not used with consistency by the three raters and although this contributed to the lower correlations in some cases, it does show a consistency among the raters. Tables 5 and 6 show the correlations between raters for the frequency with which they rated categories (with tape as the unit of analysis (n = 18). Also shown in the tables are the number of times both raters in a pair used a category for the same tape. Generally, it can be seen that the categories that were used more frequently show higher inter-rater correlations, the exceptions being (P = Q) in Table 10 and (P) in Table 11. One can conclude from these data that certain of the categories, notably $(\overline{P})$, $(\overline{P}.\overline{Q})$ and $(P.\overline{Q})$ from the logic ratings, and (P < Q) (P) (Q), and (P (Q)) in the class ratings can be used with some confidence of reliability.

Table 5

Correlations between raters for the frequencies that they used class categories, and the number of tapes (n) in which each category was used by both raters for the 18 reliability tapes.[a]

Pairs of Raters	P=Q	P>Q	P<Q	(P) (Q)	(PQ)	(P(Q)	((P) Q)	((Q) P)	other
1 with 2									
r_{xy}	n.s.	n.s.	.87	.69	.50	.77	.76	n.s.	n.s.
n	13	1	11	9	13	8	13	3	2
1 with 3									
r_{xy}	n.s.	n.s.	.87	.87	.68	.97	.89	n.s.	n.s.
n	11	1	11	11	14	8	12	2	4
2 with 3									
r_{xy}	n.s.	n.s.	.77	.90	n.s.	.74	.83	n.s.	.80
n	9	0	9	9	13	7	11	6	14

[a] Note: $p < .05$ for significant correlations.

Table 6

Correlations between raters for the frequencies that they used logic categories, and the number of tapes (n) in which each category was used by both raters for the 18 reliability tapes.[a]

Pairs of Raters	P	Q	$\overline{P}$	$\overline{Q}$	$\overline{P}.Q.$	$\overline{P}.\overline{Q}$	$P.\overline{Q}$	$P.\overline{Q}$	$N(P.\overline{Q})$	$P \nrightarrow R$ $R \rightarrow Q$	other
1 with 2											
r_{xy}	n.s	n.s.	.92	n.s.	.53	.59	.86	.67	.52	n.s.	n.s.
n	4	3	2	0	18	9	11	12	2	0	2
1 with 3											
r_{xy}	n.s.	n.s.	.83	n.s.	.51	.97	.92	.93	.56	n.s.	n.s.
n	10	5	2	1	18	12	12	10	4	0	3
2 with 3											
r_{xy}	n.s.	.5[illegible]	.79	n.s.	.86	.66	.83	.68	n.s.	n.s.	.71
n	4	4	1	0	18	12	8	11	2	0	9

Summary

The procedures outlined in this section represent a fairly complex set of perspectives for the researcher to use in assessing children's reasoning. The procedures are not meant to be comprehensive. However, they do comprise a system which assesses verbal expression (K-R), understanding of implication (C-L), and ability to reason logically (total score). It also assesses understanding of parts of conditional arguments, the arguments presented as a whole, and understanding of the class inclusion aspects of implication. Further, it has been used with children ranging in age from five to 15 years old. The following section of this report addresses four hypotheses tested using the methods and materials described in this section.

III

Analyses of Data

Study 1

Hypothesis: There will be differences across ages in the kinds of verbal explanations made by children and adolescents in judging logical arguments.

Introduction

The importance of verbal explanations for the understanding of children's reasoning behavior has been recognized in mathematical reasoning (Austin & Howsam, 1979; Vergnaud, 1979), and more traditional logical reasoning tasks such as syllogisms (Evans, 1972; Roberge, 1978; Wason, & Johnson-Laird, 1972). Piaget (1953, 1972) has suggested that at the stage of formal operations (early adolescence), reasoning with verbal symbols as abstract representations of concrete elements becomes more prominent. One would suspect, therefore, that adolescents' verbal explanations for their logical thinking would be a better reflection of their actual judgments than would be the case with younger children.

The importance of the linguistic structure on the reasoning task has been recognized (e.g., Evans, 1972; Johnson-Laird, & Taggart, 1969; Roberge, 1978). For example, Evans (1972) reported that wording a conditional as "If ..., then..." led to more correct response for MP than MT, while MT responses improved when the conditional was worded in the "Only if ..., then..." form. One of the ways of exploring the bases for this kind of finding is by asking subjects questions about why they gave the reasons that they did. This approach has been criticized by Ennis (1975), Brainerd (1978), and others. Basically the criticism deals with a lack of assurance that verbal explanations coincide with the reasoning that takes place. Verbal ability depends upon age, vocabulary, practice in explaining, background experience, formal learning, and other factors which may or may not be independent of the simple judgment of an argument as valid. Further, there is a question about whether thinking is verbal, and can be explained easily using language. Another criticism can be raised about the experimenter's ability to interpret the verbalization of the subject in the correct way. On the other hand, Evans (1972) has called for more "thinking aloud" procedures to analyze human thought and reason.

The present study attempted to reduce some of the previous criticism of verbal explanation procedures, by using the Kodroff-Roberge (K-R) system for classification of verbal explanation. Using this system, Kodroff and Roberge (1975) found arbitrary explanations and patterned explanations prominent in a sample of children in first, second and third grades of elementary school. Further, they found more correct reasoning with concrete than with verbal presentation, and MP appeared less difficult than MT. Verbal deficits was the predominant category of erroneous explanation in first grade, while arbitrary explanation predominated in second grade. Using the same procedure with mentally retarded children and adolescents, Antonak and Roberge (1978) found patterned explanations, structural defects, and verbal deficits the predominant errors made with MP, MT, and AC argument forms. Patterned explanations were most prominent with MP, verbal deficits with AC, and structural defects with MT. Finally, both studies found a lack of correspondence between correctness of reasoning responses, and correctness of the verbal explanation for the responses.

The present study explored the patterns of verbal errors in explanation among Dutch children and adolescents. All previous work cited was compiled on English-speaking children. If similar patterns of verbal explanations are found across ages, it would be convincing evidence of underlying processes not subject to language or cultural effects as Piaget (1953) has claimed.

Methods

Subjects were first, third, fifth, and seventh graders (first year secondary school) described in the subjects section earlier.

For this analysis, subjects' responses to the items on Form CA for which the four argument types appeared in the form shown in Table 7 were used. Since each item was repeated in four ways, every child responded to a total of 16 items. Further, to investigate the first hypothesis, the Kodroff-Roberge scoring system was used to classify responses.

Results

The means and standard deviations of correctness of judgments for each of the argument forms are shown in Table 7. It can be seen that most first graders could respond correctly to MP items, but that this high level of performance apparently reduced in later years. It should also be noted that on DA and AC, the youngest children performed substantially below the chance level which would suggest a performance average of 1.33 correct strictly by chance (there were three choices for each of the four presentations of the arguments). The third and fifth graders performance approximated this chance level for DA and AC, while the oldest children reached a level substantially above chance. Further, the performance of the oldest subjects was more similar for the four arguments than was true for younger children.

These patterns were further investigated with the Kodroff-Roberge scoring system as modified for the present studies. The frequencies of categories of explanations tabulated by grade and argument form are shown in Table 8. As noted in earlier studies (Antonak & Roberge, 1978; Kodroff & Roberge, 1975), there was a widespread tendency to give faulty verbal explanations despite judging the argument correctly. There was greater tendency toward verbal deficits (VD) in first graders than any other group. Arbitrary explanation (AE) were the most prominent kind of erroneous verbal explanation used by first graders, and to lesser degrees, third and fifth graders. The seventh graders tended to give proportionately greater structural deficits (SD). However, good

Table 7

Means and standard deviations of subjects' responses to four argument forms (40 is perfect score)

Argument form

Grade		MP $P \supset Q$ P $\therefore Q$	MT $P \supset Q$ $\overline{Q}$ $\therefore P$	AC $P \supset Q$ Q $\therefore P$	DA $P \supset Q$ $\overline{P}$ $\therefore Q$
1	$\overline{X}$	3.67	3.00	.33	.58
	SD	.65	1.41	1.16	1.24
	n = 12				
3	$\overline{X}$	2.25	2.33	1.33	2.00
	SD	1.42	1.45	1.56	1.71
	n = 12				
5	$\overline{X}$	3.08	2.75	1.17	1.33
	SD	1.31	1.42	1.47	1.75
	n = 12				
7	$\overline{X}$	3.22	2.72	2.33	2.78
	SD	1.11	1.45	1.53	1.52
	n = 18				
Correct response		Yes	No	Maybe	Maybe

Table 8

Frequencies of Categories of Verbal Explanations for reasoning responses (Frequency of correct responses in patentheses)[a]

Grade	Argument Form	Verbal Explanation Category[b] VD	MD	AE	SD	SR	PE	GL
1 (n = 12)	MP	4(3)	1(1)	25(23)	1(1)	0(0)	0(0)	17(16)
	MT	2(1)	0(0)	33(22)	6(6)	0(0)	0(0)	7(7)
	AC	1(0)	0(0)	27(1)	17(1)	0(0)	0(0)	3(2)
	DA	4(2)	0(0)	30(2)	9(1)	0(0)	0(0)	5(2)
	Total	11(6)	1(1)	115(48)	33(9)	0(0)	0(0)	32(27)
3 (n = 12)	MP	2(2)	3(0)	29(12)	3(2)	0(0)	0(0)	10(10)
	MT	1(1)	0(0)	29(13)	5(2)	0(0)	0(0)	13(12)
	AC	0(0)	0(0)	17(3)	13(1)	0(0)	0(0)	18(12)
	DA	0(0)	0(0)	13(5)	12(2)	0(0)	0(0)	21(17)
	Total	3(3)	3(0)	88(33)	33(7)	0(0)	0(0)	62(51)
5 (n = 12)	MP	0(0)	0(0)	20(12)	4(4)	0(0)	1(1)	23(19)
	MT	0(0)	0(0)	18(7)	10(6)	2(2)	0(0)	18(18)
	AC	0(0)	0(0)	19(4)	15(0)	1(0)	0(0)	13(10)
	DA	0(0)	0(0)	17(5)	17(5)	2(0)	0(0)	12(6)
	Total	0(0)	0(0)	74(28)	4[illegible](15)	5(2)	1(1)	66(53)
7 (n = 18)	MP	0(0)	0(0)	12(3)	8(5)	0(0)	0(0)	51(50)
	MT	0(0)	0(0)	15(1)	12(3)	0(0)	1(1)	44(44)
	AC	0(0)	1(1)	8(2)	22(2)	0(0)	1(0)	40(37)
	DA	0(0)	2(1)	3(3)	[illegible](14)	0(0)	1(0)	36(32)
	Total	0(0)	3(2)	38(9)	[illegible](24)	0(0)	3(1)	171(163)

[a] The "other" category was omitted from the table because less than 1% of the total responses fall into this category

[b] VD(Verbal deficits), MD(Memory Deficits), AE(arbitrary explanations), SD(Structural deficits), CR(Conditional reasoning), PE(Patterned Explanations), GL (Good logic).

logic (GL) was the major category of response among the oldest group.

Discussion

Generally, the kind of erroneous verbal explanation children gave when describing their reasons for logical judgments appeared more related to the age than to the form of the logical argument or syllogism. Older children in the sample (about 13 years old) gave more good logic explanations, but when they gave erroneous explanations, their preference was for structural deficits (SD). When corrected for number of students in the sample, the structural deficit kind of error did not seem to be reduced very much by age, however. That is, according to the definition of structural deficits, the child seemed aware of the structure of conditional reasoning but did not describe conditional reasoning adequately in the explanation. This suggests that there may be a tendency to explain reasoning with some deficits regardless of age, even when reasoning correctly. Further, it would seem that the improved good logic explanations among the older children may be due to a reduced tendency to use explanations that can be characterized as verbal deficits and arbitrary explanations.

Clearly MP was the easiest argument to deal with. However, it is interesting to note that the MP and MT items seemed more difficult for older children. There may be two concerns operating here. The first is a cieling effect. Since there were only four items presented in each form, the maximum score was reached by most children, and older groups could only do equally well or worse. But this does not explain the large variation between 3.67 and 2.25 found for the first and third graders, respectively. The second concern is directed more toward explanation of this difference. Specifically, data in Table 8 reflect a reduction in Verbal deficits and arbitrary explanations between first and third grade. If verbal explanations are taken as reflecting the underlying thought processes, then it would seem that more mature thought would require more consideration in responding. The matured thought process might make the easier questions more difficult, and the more difficult ones easier. Indeed, there seems to be a constant increase in good logic (GL) responses over the years relative to erroneous or faulty explanations. Further, there is a reduction in the ratio of correct judgments of the truth of the arguments to erroneous explanations, supporting the contention that more complex thinking was going on in the older subjects.

Past research of this type was done only with English-speaking children in the United States (Antonak & Roberge, 1978; Kodroff & Roberge, 1975). This study was different from the earlier ones in three ways. First, it was done with Dutch-speaking children in Holland. Second, more argument forms were used; and third, the age range was greater in the present study. With these differences in mind, it can be noted that the present study did not find the prominence of verbal deficits and patterned explanations reported in the earlier studies. Further, there was not a great degree of difference across arguments in the kind of verbal explanation provided. Thus, neither age nor argument form provides a satisfactory explanation for the differences in the present study and the others. Therefore, the native language of the children may be suspected of causing the differences in verbal explanation. It would not be at all surprising if the kind of verbal errors displayed in a task are dependent upon the language one speaks. This finding should be explored further with children who speak other languages.

A final comment before leaving discussion of the data in this first study concerns the interpretation of children's verbal explanations. It is difficult to say how much of this depends upon the verbal fluency and skill of the child, and how much it depends upon his or her reasoning skill. To some extent, this question was addressed in the previous paragraph. However, the nature of the reasoning processes involved in answering "Klopt dat?" may be different from those involved in answering "Waarom zeg je dat?" Some suggestion that these processes are different can be found in the high number of correct judgments that were explained erroneously, especially with younger children. Again, this seems to be fertile area for research.

Study 2

Hypothesis: Class inclusion will be related to correct judgment of the truth of logical arguments containing material implication.

Introduction

The relationship between class inclusion and the logic of implication has been inferred in the literature for some time (Brainerd, 1978; Ennis, 1975; Knifong, 1974). Basically, this relationship suggests that children must be able to recognize superordinate and subordinate relationships before mature implication can be demonstrated. In the typical class inclusion (CI) problem described by Piaget (1953), a class of objects or elements A, and another A^1, are combined to form a larger class B. Thus,

$$A \,\&\, A^1 = B$$

The child is presented the elements of A, A^1, and B, and asked if there are more A or B. For example, B could symbolize a set of objects called all wooden beads, A would be "blue, wooden beads," and A^1 would be "wooden beads that are not blue." It can easily be seen that A can be added to A^1 to get B. Although a child of five to seven years of age can typically say this relationship, it is not until later that the child becomes able to recognize that:

$$A < B,$$
$$\text{and} \quad B - A^1 = A .$$

Thus, the child is described by Piaget (1952, 1953; Inhelder & Piaget, 1958) as finding it difficult to think about part-whole relationships. Therefore, prior to being able to deal with these relationships, it would be probable that after presenting a child with an array of objects, say beads as described above, the child would not be able to consistently say that the wooden beads outnumber the blue wooden beads. This has been established in a number of empirical studies that have been reviewed elsewhere (e.g., Brainerd, 1978, 1979; Dodwell, 1962; Judd & Mervis, 1979; Kuhn, 1977; Wilkinson, 1976). Some researchers have specifically found results suggesting a compatibility of class inclusion skill and implication reasoning (Elkind, Anagnostopoulou, Malone, 1970; Falmagne, 1975; Kuhn, 1977). Others have shown that the class inclusion problem is actually a complicated task involving a number of different skills and that it can be solved with ordering or transitive algorithms without actually recognizing all the features of mature implication (Trabasso, <u>et</u>. <u>al</u>., 1978; Trabasso, Riley, & Wilson, 1975).

As has been discussed by Ennis (1975) the logic of implication, and that of class inclusion imply the identical propositions: $p\ q$, $\bar{p}\ q$, $\bar{p}\ \bar{q}$, not $(p\ \bar{q})$. On the basis of this and the research, therefore, one would expect the psychological processes involved in responding to "are there more (A) or (B)? to be similar to responding to a question about a conditional argument such as: "If all (A)s are (B)'s, and this is an (A), then does it follow that it is a (B)?" which contains the "A implies B" implication. The present study explored the class inclusion aspects of such implication reasoning.

<u>Method</u>

<u>Subjects</u>. The 45 subjects who participated in the study were Dutch children in kindergarten, second, fourth and sixth grades who had responded to Form PW. There was one more girl in the sample than there were boys. Average ages were: 5.9 years (Kleuter), 8 5 (second), 9.7 (fourth), and 12.3 (sixth). The upper two age levels from the PW sample was not used because those children almost always responded correctly to the class inclusion question producing very little variance to study at the older ages.

<u>Materials</u>. As described earlier, using Form PW, each of the four argument types, MP, MT, DA, and AC, were presented four times. The conditional arguments were not only read to each child, but were demonstrated using pictures, or toy animals or blocks. All of the items were about the blocks (i.e. "If all blocks are wood, and this is a block (one is shown separate from the group), then it is wood.") The child responded YES, NO, or MAYBE. In addition, each major premise

was administered separately from the whole argument, (half of the time this was done before the full argument was given, and half the time afterwards), and the class inclusion question was asked. An example of the procedure is as follows: The child is first told "All dogs are aminals." He or she would then be asked the class-inclusion question "Are there more dogs than animals?"

Procedures. The subjects were given three practice items as discussed earlier in the section describing Form PW in detail. Following this the four part interview was administered, recorded, transcribed and scored using methods already discussed in this report. Then the number of correct responses to the four argument forms was correlated to the number of correct responses on the class inclusion questions.

Results

Tables 9 and 10 show the means, standard deviations and correlations between reasoning scores and class inclusion scores for each grade level. For the kindergarten and sixth graders, there were no statistically significant correlations found. However, the moderate relationship between class inclusion score and MP score was significant ($p < .05$) at the fourth grade level, and the rather substantial correlations between class inclusion and DA, AC, and total scores at the second grade level were significant ($p < .001$). Further, it can be seen in Table 9 that total reasoning score improved at each successive age level. This systematic improvement with age seemed to be due to consistent increases in scores on DA and AC, rather than the less consistent pattern found for MP and MT across age levels.

Table 9

Means and standard deviations of scores on each argument,

total reasoning scores, and class inclusion

Grade		Class Inclusion	Argument				Reasoning Total
			MP	MT	DA	AC	
K (n = 10)	$\overline{X}$	18.80	3.90	2.70	.20	.10	6.90
	SD	18.97	1.60	2.06	.63	.32	2.08
2 (n = 11)	$\overline{X}$	31.64	5.00	3.36	.73	1.09	10.18
	SD	28.91	1.00	1.29	1.27	1.70	4.14
4 (n = 12)	$\overline{X}$	26.92	4.83	4.92	1.08	1.75	12.58
	SD	8.23	.94	1.16	1.31	1.35	3.26
6 (n = 12)	$\overline{X}$	27.75	5.25	3.50	1.75	2.50	13.00
	SD	6.66	.96	1.68	1.96	2.15	5.19

Table 10

Correlations of reasoning scores on each argument and total reasoning score with class inclusion (CI) scores for each grade level

Group	Correlations of reasoning scores with CI				
	MP	MT	DA	AC	Total
Kindergarten	.16	.18	.11	.11	-.01
Second	.39	.44	.89**	.83**	.85**
Fourth	.60*	.05	-.29	-.30	-.09
Sixth	.55	.36	.25	.12	.36

* $p < .05$

** $p < .001$

Discussion

In regard to the original hypothesis that responses to the class inclusion question should correlate with responses to arguments containing conditional implication, the results seem mixed at first glance. On the one hand, class inclusion score seemed significantly related to reasoning in only one instance (MP) in the fourth grade group, and three in the second grade group (DA, AC, and total). On the other hand, a closer look at the descriptive statistics in Table 9 suggests that the results may reflect a greater relationship between the CI response and reasoning scores than the correlations along suggest. For example, the CI responses in Kindergarten appeared to be a little better than chance (on the basis of two choices, and 32 items). The large standard deviation of CI in the kindergarten sample further supports the supposition that many of those children responded randomly.

Looking at the oldest group, another kind of problem appeared with regard to interpreting the correlations. Specifically, the relatively high scores and small standard dev-ations for CI in fourth and sixth grades suggested attenuation of the CI scores at those levels. This would have the effect of depressing the correlation coefficients. Thus, since students tend to do well on CI above 4th grade, the relationship between CI and reasoning cannot be explored at those levels using the instrumentation of the present study.

Considering the statistical characteristics of the data, it does seem that correctness of response to the class inclusion question is related to correctness of reasoning. This conclusion is made based on the omission of data from grade six and kindergarten due to attenuation and random responding, respectively. Therefore, the data which seem most useful to this hypothesis are those of the middle two grades at which four out of ten correlations were positive and statistically significant as proposed in the hypothesis.

The present research suggests some recommendations for studying the relationship between implication reasoning and class inclusion. First, the ability to handle class inclusion questions seems to be emerging at about eight years of age. This is not very different from what was suggested by other researchers (e.g., Dodwell, 1961, 1962; Inhelder & Piaget, 1958; Piaget, 1953). Although it has been shown that younger children can be taught to answer class inclusion questions correctly (e.g., Trabasso, et. al., 1978), their level of understanding of class inclusion has not been demonstrated as equal to older children who answer CI questions correctly. Therefore, children of five years old or kindergarten age are probably too young for a meaningful study of class inclusion. Second, the present study did not investigate the range of tasks usually considered to represent class inclusion (Brainerd, 1979; Inhelder & Piaget, 1958; Klahr & Wallace, 1972; Trabasso, et. al, 1978). A future study of class inclusion and implication reasoning should use a greater range of class inclusion tasks. Not only would such a range of tasks provide a more complete assessment of class inclusion, but also might provide enough range of response to reduce the problem of attenuation of the CI variable that was found in the present study.

In conclusion, the present study provided evidence that class inclusion is related to conditional implication reasoning. From the data it appeared that the youngest children (about five years of age) were responding to the CI question randomly, while the oldest ones (about 12 years old) tended to respond to it correctly. Further, the CI performance seemed to be somewhat in transition and related to reasoning at the second and fourth grade level. Finally, improvement in total reasoning seemed to be most related to imporvement in reasoning with invalid arguments requiring the MAYBE response (A and AC), a finding which corresponds to earlier research with English-speaking subjects (e.g., Hadar, 1975, Roberge, 1970).

Study 3

Hypothesis: There will be an inverse relationship between response time and correctness of reasoning which reduces as children become adolescents.

Introduction

A considerable body of knowledge has been developed regarding the difficulty of the four logical arguments MP, MT, DA, and AC for children and adults. Roberge (1970) reported that the invalid principles of affirmation of the consequent (AC) and denial of the antecedent (DA) were more difficult than the valid modus ponens (MP) and modus tollens (MT). Research by Hadar (1975) with elementary school children and O'Brien (1973) with high school subjects, have reported the following order of difficulty

$$MP < MT < DA < AC.$$

In other research, Roberge (1970), Roberge and Mason (1978) and Mason, Bramble and Mast (1975) found the following slightly different order with small differences between difficulty of DA and AC:

$$MP < MT < AC < DA$$

using fourth through tenth graders as subjects.

Roberge (1971) added a fifth principle, transivity (TR) to his study of difficulty of conditional reasoning in college students, (the transivity argument form is: $P \supset Q$, $Q \supset R$, $P \supset R$) and found the following unique pattern of difficulties among the adult subjects:

$$MP < TR < DA < AC < MT$$

Although there was no explanation for the unusual difficulty of the MT form in this study, Roberge noted that many subjects tended to respond that the argument was neither valid nor invalid. Excepting this one study, according to most research, the easiest argument to recognize as true seems to be MP, followed closely by MT. Affirmation of the consequent and denial of the antecedent seem about equally difficult, but more difficult than the other two.

The present research was undertaken to explore the use of response time as a measure of response difficulty on judging logical arguments. Response time would seem to be an appropriate measure of task difficulty from psychological information processing. The rationale for this point of view would be that more difficult problems require more steps for locating solutions. If a problem is easy to a subject, the solution should require fewer steps (Homa, 1973; Newell & Simon, 1972; Wohlwill, 1962). Response time, and a closely related variable, reaction time, are among the earliest variables studied by psychologists and those interested in mental processes (e.g., Donders, 1869 cited in W. G. Koster, 1969). Although there has been considerable research done on the time one takes to respond to different kinds of tasks, relatively little has been directed toward judging of logical arguments. Although Revlin and Leirer (1978) have argued that increased time will allow more task scanning, and Huttenlocher (1968) has given evidence of difficulty of judging a premise in a transitive ordering task and time required in responding, little has been done regarding the relationship between argument difficulty, and response time. The present study explored this with respect to reasoning with conditional syllogisms whose relative difficulty had been established in research discussed previously. Specifically, the study was conducted to determine whether the response latency for MP, MT, DA, and AC argument forms would match the order of difficulty reported in earlier studies.

Methods

Subjects. Fifty-four Dutch students' responses to the Form CA were used in the present study. The students were from the first, third, and fifth grades and first year of secondary school. Their characteristics were described in an earlier section of this paper.

Materials. The children were administered the Form CA interview in the manner discussed earlier. Since the present study was only concerned with the four arguments in their most common form, only the 16 items that were in the form shown in Table 7 were used in the present study.

Procedures. Data were collected as described earlier. Students were given one point for each correct response, and scores for each argument were summed.

Results

Table 11 shows the means of time and reasoning score (average total number of correct items). As can be seen from the table, the average scores suggest patterns of difficulty described earlier in other studies. However, the time scores did not follow such a pattern. Correlations between time and score at each grade level for the for argument forms produced only three statistically significant correlations among the total of sixteen possible. The MT form produced a significant correlation ($P < .05$) for third grade (-.50), and fifth grade (-.63). MT had a significant correlation at the secondary school level (-.43) ($p < .05$). In short, for these significant correlations the shorter the average response time, the higher the reasoning score as expected. However, the relationship does not appear to be substantial. Further, the total score obtained by summing performance across arguments was not significantly correlated to total time at any grade level.

Conclusions

There seemed to be no systematic significant correlations between reasoning scores and response time. However, this lack of product-moment correlation should not be taken to mean lack of some kind of consistent relationship. Figure 1 shows the graphs of responses to the four argument forms with the corresponding times across the four age groups of the present study. There appears to be a

Table 1

Scores and response times on four conditional argument forms

Grade		Argument Forms: MP Score	MP Time[a]	MT Score	MT Time	DA Score	DA Time	AC Score	AC Time
1 (n = 12)	$\bar{X}$	3.66	9.60	2.66	12.49	.58	12.50	.33	9.38
	SD	.65	4.18	1.50	4.06	1.24	3.75	1.15	3.19
3 (n = 12)	$\bar{X}$	2.25	12.08	2.33	13.76	2.00	13.89	1.33	9.68
	SD	1.42	8.33	1.44	7.91	1.70	8.10	1.56	3.12
5 (n = 12)	$\bar{X}$	3.08	12.02	2.75	12.09	1.33	12.93	1.67	9.16
	SD	1.31	8.68	1.42	4.92	1.78	8.66	1.47	3.11
first year secondary (n = 18)	$\bar{X}$	3.22	9.69	2.72	12.76	2.33	13.02	2.77	9.16
	SD	1.11	4.71	1.45	4.92	1.53	6.79	1.52	3.57

[a] Time is indicated in seconds.

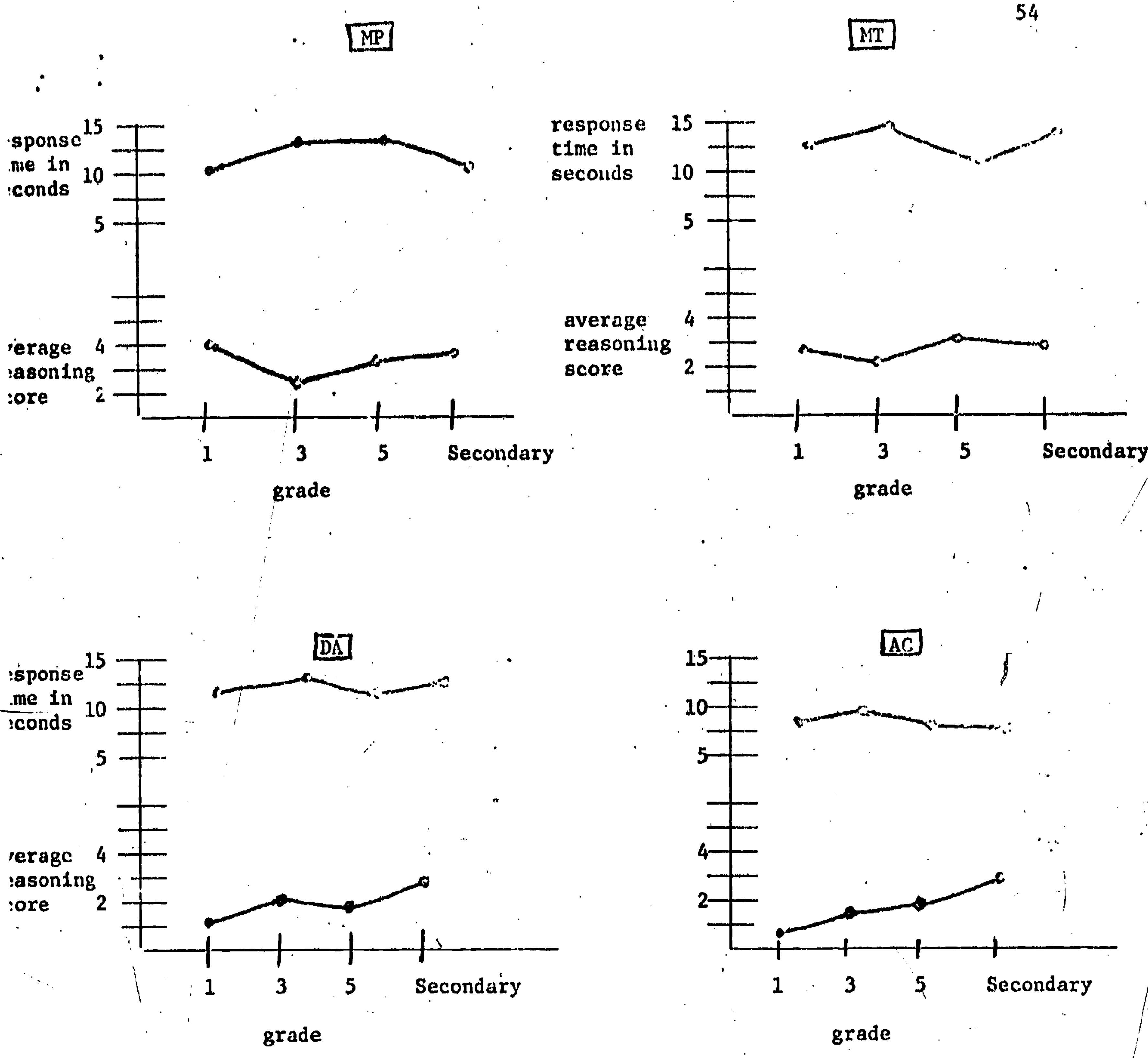


Figure 1. Response times and reasoning scores for the four arguments plotted against age.

fairly consistent tendency for the younger and older groups to respond relatively quickly. On two arguments, MP and MT, the more rapid response times are associated with higher scores, and the slower response times with lower scores. However, for DA, the elementary grade pattern seems to suggest that when children took more time, they tended to give more correct answers. But the secondary students performed best on these items although they had short response times.

All this seems to suggest that there are relationships between arguments and response times but that they are complex. Therefore, referring to the original hypothesis, response time seems to be inadequate as an indicator of difficulty. Further, results of the present analysis suggest that investigation of the strategies or processes underlying the correct responses might explain the complex relationships observed between response time and reasoning scores.

Study 4

Hypothesis: The class meaning of a logical argument as reflected by subjects' explanations of their reasoning will correspond to the four components of material implication ($p \cdot q$, $\bar{p} \cdot q$, $\bar{p} \cdot \bar{q}$, $\sim(p \cdot \bar{q})$).

Introduction

It has been recognized that behavior that is observed may be the result of different internal processes, and that describing observable behavior does not necessarily describe the thoughts that occured with the behavior (DeCorte, 1977). Numerous attempts have been made to explore the thoughts behind the logical behavior of children and adults. Most attention has been focused on processes underlying erroneous reasoning. For example, Helsabeck (1975) suggested that a cause of faulty reasoning was a tendency to overlook alternatives in choosing or judging a conclusion. Yet, Ceraso and Provitera (1971) argued that faulty understanding of premises can produce erroneous reasoning.

Researchers have questioned the existence of Piaget's 16 binary operations in mature reasoning. For example, the 16 binary operations have not been easy to demonstrate (Bynam, Thomas, & Weitz, 1972; Roberge & Flexer, 1979(a); Weitz, Bynam, & Thomas, 1973). Kuhn (1977) has reported evidence supporting the four components of implication that are available to the formal operations thinker from the 16 binary propositions ($p \cdot q$, $\bar{p} \cdot q$, $\bar{p} \cdot \bar{q}$, $N(p \cdot \bar{q})$). However, the last component, the denial of non-implication, could not be supported by van Duyne (1974). Further, it has been suggested that transductive reasoning, in which the implication connection is understood as "and" (i.e. $p \cdot q$ rather than $p \supset q$), is used by young children to provide correct judgments in some forms of propositional logic (Knifong, 1974). As Kuhn (1977) suggested, a child using transductive reasoning would not understand conditional or causal relationships between elements of propositions.

The objections to Piaget's logic of material implication raised by Ennis (1975) and others was discussed earlier in this report. Ennis' criticisms deal more with the logic of Piaget than with the thinking processes underlying reasoning behavior. For example, Ennis objected to Piaget's unclear use of the "v" symbol to variously mean "and," "or," "and/or," the possibility under Piaget's system of the existence of something both (p) and (not p). However, such inconsistencies may be possible in thought even if they are not possible in reality.

Smedslund (1977) has taken the position that logical operations do not exist as Piaget has suggested. He based this assertion on the observations that children behave differently when faced with a change in the content of a task having the same logical structure, that people do not seem to ordinarily behave abstractly, and that the direct evidence for logical is scant. Smedslund argued that the researcher in psychology usually pursues a person's understanding, but in the Piaget system, one assumes understanding and pursues the logic fostered by the understanding. According to Smedslund, there is a circularity between understanding and logical performance that is difficult to untangle. Staudenmayer (1975) suggested that interpretation of premises in logical arguments is an inductive process, and that when individuals do not seem to reason according to formal logic, they may be behaving logically according to their understanding of the premises. Further, Halpern (1965) discussed a perceptual dependence in logical thinking that is more prominent in children 5-7 years old. During this period, according to Halpern, a perceptual dependence could hinder the child from correctly judging a logical argument by focusing understanding of the premises in terms of irrelevant (or less than salient) observable traits of the elements in the premises.

Thus, while researchers have generally recognized the role of understanding of the premises in logical reasoning, there has not been much closure on how understanding is determined, or what influences it. The present study was designed to explore a child's understanding of the premises in logical arguments through their explanations of their judgments of the validity of their arguments.

Method

Subjects: The students in the first, third, and fifth grades, and the first year of secondary school that were described in an earlier section of this report served as subjects in the present study.

Materials: Form CA responds to the 16 items for which the conclusion was not negated were scored by the three Dutch graduate students in developmental psychology who had collected the data.

Procedure: Scoring procedures and criteria for the C-L scoring system that were described earlier were used in the present study. After the scoring was completed, scores were transferred to computer cards for analysis.

Results

Tables 12 and 13 show the frequency and proportions of the categories of class and logical scoring systems for each grade. It is immediately apparent that the children's explanations covered a variety of both class and propositional categories. However, the (P)(Q) interpretation seemed to occur somewhat more often than most other class categories. Further, the most frequently used class relationships category had $P < Q$ paired with (P) Q. This pair was represented between 16 and 23 percent of the time until 5th grade. In the first year secondary school students, it appeared much more frequently than that (42%). This category would seem to represent the most complete understanding of the class relationships underlying material implication, and it is noteworthy that the secondary school children used this category about twice as frequently as the younger children.

A similarly dispersed pattern of categories of propositions was found by classifying explanations using the logical system. As can be seen in Table 13, $P \cdot Q$ was generally used in about 10 to 15 percent at all age levels, and $P \cdot Q$,

Table 12

Response categories for all arguments combined for class scoring system (proportions in parenthesis)[a]

Response Category[b]	(n=12) One	Grade (n=12) Three	(n=12) Five	(n=18) First year secondary
No Response	30 (.16)	18 (.09)	5 (.03)	4 (.01)
1. P = Q	1 (.01)	3 (.02)	3 (.02)	1 (.00)
2. P > Q	2 (.01)	2 (.01)	0	0
3. P < Q	0	1 (.01)	1 (.01)	1 (.00)
4. Ⓟ Ⓠ	7 (.04)	3 (.02)	3 (.02)	0
5. (Ⓠ P)	20 (.10)	19 (.10)	15 (.08)	27 (.09)
6. (P)(Q)	31 (.16)	45 (.23)	39 (.20)	36 (.13)
7. (Ⓟ Q)	16 (.08)	1 (.01)	20 (.10)	21 (.07)
8. P Q	0	2 (.01)	0	2 (.00)
9. P = Q, (P Q)	32 (.17)	18 (.09)	24 (.13)	24 (.08)
10. P < Q, (Ⓟ Q)	31 (.16)	36 (.19)	44 (.23)	120 (.42)
11. P > Q, (Ⓠ P)	7 (.04)	23 (.12)	22 (.11)	15 (.05)
12. Ⓟ Ⓠ, (P)(Q)	2 (.01)	5 (.03)	4 (.02)	13 (.05)
13. P > Q, Ⓟ Ⓠ, (Ⓟ Q)	1 (.01)	6 (.03)	3 (.02)	12 (.04)
14. P < Q, Ⓟ Ⓠ, (Ⓠ P)	0	1 (.01)	0	1 (.00)
15. P < Q, (P)(Q)	0	0	3 (.02)	0
16. (P)(Q), (Ⓟ Q)	0	1 (.01)	1 (.01)	0
17. P = Q, P < Q, (P Q), (Ⓟ Q)	2 (.01)	0	2	0
18. P > Q, (P)(Q)	3 (.02)	1 (.01)	0	2 (.00)
19. (Ⓟ Q) other	0	1 (.01)	0	1 (.00)
20. P = Q, Ⓟ Ⓠ, (P Q)	3 (.02)	0	0	0
21. P = Q, (P Q) (P)(Q)	3 (.02)	2 (.01)	0	5 (.02)
22. P > Q, (Ⓟ Q)	1 (.01)	3 (.02)	2 (.01)	3 (.01)
23. (P)(Q) (Ⓠ P)	0	1 (.01)	1 (.01)	0
Total	192	192	192	288

[a] Proportions rounded to nearest one-hundredth

[b] Categories 9 - 23 are combinations of the first eight.

Table 13

Propositional logic categories of explanations for reasoning on MP, MT, AC, DA[a]

Class logic[b] category	Grade One (n=12)	Grade Three (n=12)	Grade Five (n=12)	Grade Eerste (n=18)
No Response	24 (.13)	16 (.08)	2 (.01)	7 (.02)
1. P	5 (.03)	0	0	1 (.00)
2. Q	2 (.01)	1 (.01)	1 (.01)	0
3. $\overline{P}$	1 (.01)	1 (.01)	1 (.01)	1 (.00)
4. $\overline{Q}$	5 (.03)	9 (.05)	2 (.01)	0
5. $P \cdot Q$	20 (.10)	28 (.15)	20 (.10)	34 (.12)
6. $\overline{P} \cdot Q$	24 (.13)	31 (.16)	40 (.21)	44 (.15)
7. $\overline{P} \cdot \overline{Q}$	22 (.11)	2 (.01)	25 (.13)	22 (.08)
8. $(P \cdot \overline{Q})$	2 (.01)	1 (.01)	0	1 (.00)
9. $N(P \cdot \overline{Q})$	0	1 (.01)	0	0
10. $P \cdot Q$, $\overline{P} \cdot Q$	35 (.18)	16 (.08)	26 (.14)	31 (.11)
11. $P \cdot Q$, $\overline{P} \cdot \overline{Q}$	26 (.14)	42 (.22)	36 (.19)	106 (.37)
12. $P \cdot Q$, $(P \cdot \overline{Q})$	7 (.04)	21 (.11)	12 (.06)	12 (.04)
13. $P \cdot Q$, $\sim(P \cdot \overline{Q})$	9 (.05)	1 (.01)	7 (.04)	6 (.02)
14. $P \cdot Q$, $\overline{P} \cdot Q$, $\overline{P} \cdot \overline{Q}$	0	3 (.02)	3 (.02)	11 (.04)
15. $P \cdot Q$, $\overline{P} \cdot Q$, $\overline{P} \cdot Q$, $\sim(\overline{P} \cdot \overline{Q})$	0	4 (.02)	0	3 (.01)
16. $P \cdot Q$, $\overline{P} \cdot Q$, $\overline{P} \cdot \overline{Q}$, $(P \cdot Q)$	0	5 (.03)	8 (.04)	1 (.00)
17. $P \cdot Q$, $\overline{P} \cdot Q$, $(P \cdot \overline{Q})$	0	0	6 (.03)	1 (.00)
18. $P \cdot Q$, $\overline{P} \cdot \overline{Q}$, $(P \cdot \overline{Q})$	2 (.01)	2 (.01)	0	0
19. P, $\overline{p}$	3 (.02)	0	0	1 (.00)
20. $P \cdot Q$, $\overline{P} \cdot \overline{Q}$, $\sim(P \cdot \overline{Q})$	3 (.02)	0	0	4 (.01)
21. $P \cdot Q$, $\overline{P} \cdot Q$, $\sim(P \cdot \overline{Q})$	1 (.01)	0	0	0
22. $\overline{P} \cdot Q$, $\overline{P} \cdot Q$	0	6 (.03)	2 (.01)	1 (.00)
23. $\overline{P} \cdot \overline{Q}$, $(P \cdot \overline{Q})$	0	1 (.01)	0	0
24. $P \cdot Q$, $\overline{P} \cdot Q$, $\overline{P} \cdot \overline{Q}$, $(P \cdot \overline{Q})$	0	1 (.01)	0	0
25. Q, $\overline{Q}$	1 (.01)	0	0	0
26. $P \cdot Q$, $\overline{P} \cdot Q$	0	0	1 (.01)	0
27. $\overline{P} \cdot Q$, $\overline{P} \cdot \overline{Q}$, $\sim(\overline{P} \cdot Q)$	0	0	0	1 (.00)
Total	192	192	192	288

[a] Proportions rounded to nearest one-hundredth.

[b] Categories 9 - 27 are combinations of the first eight.

$\overline{P} \cdot Q$ was generally used in 8 to 18 percent at all age levels. The same was true of $P \cdot Q$, $\overline{P} \cdot \overline{Q}$ at the first, third, and fifth grade levels, but for this category at the secondary level, the proportion of responses nearly doubled.

The variation in C-L response categories across the four argument forms is shown in Tables 14 and 15. With the class scoring system, between a third and a quarter used the most complete kind of response ($P < Q$, (P) Q). There did not seem to be very much difference in the kinds of categories of explanation given for responses to the four argument forms with either the class or the logical scoring. About a third to a quarter provided the most frequently given category of response on the logical proposition scoring form, $P \cdot Q$, $\overline{P} \cdot \overline{Q}$.

Table 14

Response categories on the Class scoring system for each argument (proportions in parentheses)

Category[b]	MP	MT	AC	DA	TOTAL
	Argument Form				
No Response	14 (.06)	18 (.08)	16 (.07)	9 (.04)	57 (.07)
P = Q	3 (.01)	1 (.00)	0	4 (.02)	8 (.01)
P > Q	2 (.01)	0	2 (.01)	0	4 (.00)
P < Q	0	0	1 (.00)	2 (.01)	3 (.00)
(P) (Q)	5 (.02)	2 (.01)	4 (.02)	2 (.01)	13 (.02)
(P Q)	13 (.06)	25 (.12)	17 (.08)	26 (.12)	81 (.09)
(P)(Q)	38 (.18)	62 (.29)	42 (.19)	9 (.04)	151 (.17)
((P) Q)	16 (.07)	16 (.07)	17 (.08)	9 (.04)	58 (.07)
((Q) P)	0	0	2 (.01)	2 (.01)	4 (.00)
P = Q, (P Q)	28 (.13)	18 (.08)	24 (.11)	28 (.13)	98 (.11)
P < Q, ((P) Q)	70 (.32)	52 (.24)	56 (.26)	53 (.25)	231 (.27)
P > Q, ((Q) P)	11 (.05)	8 (.04)	15 (.07)	33 (.15)	67 (.08)
(P) (Q) (P)(Q)	6 (.03)	1 (.00)	8 (.04)	9 (.04)	24 (.03)
P > Q, (P) (Q) ((P) Q)	4 (.02)	8 (.04)	2 (.01)	8 (.04)	22 (.03)
P < Q, (P) (Q), ((Q) P)	0	0	0	2 (.01)	2 (.00)
P < Q, (P)(Q)	0	0	1 (.00)	2 (.01)	3 (.00)
(P)(Q) ((P) Q)	1 (.00)	1 (.00)	0	0	2 (.00)
P = Q, P < Q, (P Q) ((P) Q)	0	1 (.00)	1 (.00)	2 (.01)	4 (.00)
P > Q, (P)(Q)	1 (.00)	0	2 (.01)	3 (.01)	6 (.01)
P = Q, (P)(Q) (P Q)	0	0	1 (.00)	1 (.00)	2 (.00)
P = Q, (P Q) (P)(Q)	0	1 (.00)	1 (.00)	1 (.00)	3 (.00)
P > Q, ((P) Q)	2 (.01)	0	3 (.01)	5 (.02)	10 (.01)
(P)(Q) ((P) Q)	2 (.01)	2 (.01)	0	5 (.02)	9 (.01)
P < Q ((Q) P)	0	0	1 (.00)	1 (.00)	1 (.00)
Total	216	216	216	216	864

[a] Proportions rounded to nearest hundredth.

[b] Categories 9-23 are combinations of the first eight.

Table 15

Response Categories on the Propositional
Logic scoring system for each argument (propositions in parentheses)[a]

	Category[b]	MP	MT	AC	DA	TOTAL
	No Response	14 (.06)	12 (.06)	14 (.06)	9 (.04)	49 (.06)
1.	P	1 (.00)	3 (.01)	1 (.00)	1 (.00)	6 (.01)
2.	Q	1 (.00)	0	1 (.00)	2 (.01)	4 (.00)
3.	$\overline{P}$	0	0	1 (.00)	3 (.01)	4 (.00)
4.	$\overline{Q}$	6 (.03)	4 (.02)	5 (.02)	1 (.00)	16 (.02)
5.	$P \cdot Q$	16 (.07)	24 (.11)	27 (.13)	35 (.16)	102 (.12)
6.	$\overline{P} \cdot Q$	55 (.25)	44 (.20)	28 (.13)	12 (.06)	139 (.16)
7.	$\overline{P} \cdot \overline{Q}$	24 (.11)	24 (.11)	18 (.08)	5 (.02)	71 (.08)
8.	$(P \cdot \overline{Q})$	1 (.00)	1 (.00)	0	2 (.01)	4 (.00)
9.	$N(P \cdot \overline{Q})$	0	1 (.00)	0	0	1 (.00)
10.	$P \cdot Q, \overline{P} \cdot Q$	29 (.13)	29 (.13)	26 (.12)	24 (.11)	108 (.13)
11.	$P \cdot Q, \overline{P} \cdot \overline{Q}$	43 (.20)	50 (.23)	68 (.31)	49 (.23)	210 (.24)
12.	$P \cdot Q, (P \cdot \overline{Q})$	8 (.04)	7 (.03)	13 (.06)	24 (.11)	52 (.06)
13.	$P \cdot Q, \sim(P \cdot \overline{Q})$	4 (.02)	4 (.02)	2 (.01)	13 (.06)	23 (.03)
14.	$P \cdot Q, \overline{P} \cdot Q, \overline{P} \cdot \overline{Q}$	4 (.02)	2 (.01)	0	11 (.05)	17 (.02)
15.	$P \cdot Q, \overline{P} \cdot Q, \overline{P} \cdot Q, \sim(P \cdot \overline{Q})$	2 (.01)	3 (.01)	1 (.00)	1 (.00)	7 (.01)
16.	$P \cdot Q, \overline{P} \cdot Q, \overline{P} \cdot \overline{Q}, (P \cdot Q)$	2 (.01)	5 (.02)	3 (.01)	4 (.02)	14 (.02)
17.	$P \cdot Q, \overline{P} \cdot Q, (P \cdot \overline{Q})$	1 (.00)	1 (.00)	0	5 (.02)	7 (.01)
18.	$P \cdot Q, \overline{P} \cdot \overline{Q}, (P \cdot \overline{Q})$	1 (.00)	1 (.00)	1 (.00)	1 (.00)	4 (.00)
19.	$P, \overline{P}$	0	0	2 (.01)	2 (.01)	4 (.00)
20.	$P \cdot Q, \overline{P} \cdot \overline{Q}, \sim(P \cdot \overline{Q})$	1 (.00)	0	4 (.02)	2 (.01)	7 (.01)
21.	$P \cdot Q, \overline{P}\ Q, \sim(P\ \overline{Q})$	0	0	0	1 (.00)	1 (.00)
22.	$\overline{P} \cdot Q, \overline{P} \cdot \overline{Q}$	3 (.01)	0	0	6 (.03)	9 (.01)
23.	$\overline{P} \cdot \overline{Q}, (P \cdot \overline{Q})$	0	0	1 (.00)	0	1 (.00)
24.	$P \cdot Q, \overline{P} \cdot Q, \overline{P} \cdot \overline{Q}, (P \cdot \overline{Q})$	0	0	0	1 (.00)	1 (.00)
25.	$Q, \overline{Q}$	0	0	0	1 (.00)	1 (.00)
26.	$P \cdot Q, \overline{P} \cdot Q$	0	1 (.00)	0	0	1 (.00)
27.	$\overline{P} \cdot \overline{Q}, \overline{P} \cdot \overline{Q}, \sim(\overline{P} \cdot Q)$	0	0	0	1 (.00)	1 (.00)
	TOTAL	216	216	216	216	864

[a]Propositions rounded to nearest hundredth.

[b]Categories 13 - 27 are combinations of the first eight.

Discussion

The results were interpreted as suggesting that the hypothesis received only limited support from the data. Specifically, it appeared that there was some growth, as children become older, toward more complete recognition of the combinatorial subordinate and superordinate class relationships underlying the elements of propositions in material implication. However, there appeared to be great variation in the kind of relationship that could be inferred from subjects' explanations, even with the oldest children in the sample.

Further, with regard to the original hypothesis, the four components of material implication $\left[(p \cdot q, \bar{p} \cdot q, \bar{p} \cdot \bar{q}, \sim(\bar{p} \cdot \bar{q})\right]$ were not generally observable in the subjects' explanations. This suggested that the logical propositions underlying children's responses to logical tasks may not be complete in the formal logical sense, even with children who are supposedly operating at a formal level (13 years old).

One should approach these data with a number of cautions in mind. First, as Brianerd (1973) pointed out, explanations may not be good reflections of reasoning behind judgments. There are a variety of tasks to perform in formulating and providing an explanation that are greatly simplified when only a one word judgment response is required. Thus, the intervening processes could inhibit accurate representation of understanding in a verbal explanation.

A second area of caution concerns the C-L scoring form. While the interrater reliability was adequate, only large differences should be interpreted. For that reason, only the categories getting proportions of .10 or more were discussed in thie report. When response rate was at a lower level, the stability of the category might be questionable.

With these cautions in mind, the findings were rather consistent with the literature. For example, logical judgment tasks have been suggested as separate from understanding of the logic involved (Smedslund, 1970). The relatively higher rate of correct judgments of the truth status of the four arguments reported earlier (Table 7) suggests that such a task is different from giving explanations for those judgments on the basis of the inconsistency of categories of response seen in the C-L scoring. Further, previous research questioning the existence of the 16 binary operations in logical thinking (i.e. Bynam, et al., 1973) was supported by the present data.

Researchers have suggested a variety of explanations for the difference between performance on logical tasks and understanding of that performance. For example, Begg and Denny (1969), Chapman and Chapman (1959), and others have proposed an "atmosphere" effect in which quantifiers, negation, and other characteristics of the premises of an argument affect the conclusion. Johnson-Laird (1975) stressed the role of prior beliefs. Northrup (1977) emphasized the nature of the task, and so on.

This study raises many questions that will be addressed in subsequent analyses of these data. For example, the relationship of class to logical category, and the relationship of category of explanation to correctness of response will be explored. Further, the effects of part-whole presentation (Form PW) and verbal versus visual presentation of the premise material seem to offer promising avenues for further analysis.

IV

Conclusions

The four studies that have been completed so far represent only a small part of the possible analyses created by the data. However, they do suggest some insight into the complex reasoning processes of children and adolescents. For example, the data suggested that verbal deficits in explanations were only a major problem for first graders. Children in the older groups gave progressively less arbitrary explanatic and more good logic explanations. These findings suggested that the means of explaining seemed to mature with age. There were not significant differences in this pattern at different ages.

Further, with regard to explanations, the logical premises expected to become more prominent in older students (according to Piaget's (1953) logical components of implication) did not surface. Rather, explanations seemed to reflect a wide variety of understandings of the premises at all ages. A similar pattern was found for the class scoring suggesting that class relationships in logical premises containing implication may not be routinely inferred. Again, the argument form had little effect on the patterns of explanations. This is particulary noteworthy when compared to the increased rate of correct responding with age, and the differential rate of correct responses to each argument form. One obvious conclusion might be that the judgment of the truth of a syllogism represents a different, and perhaps easier task than the explanation of that judgment.

The data suggested that class inclusion was related to the ability to judge the truth of logical arguments containing implication. This is compatible with what would be expected, and has often been assumed.

Response time seemed to be related to response correctness in an unpredicted way. That is, when correct responding was highest, response time was shortest.

This occured even when younger children were more correct than older children on the modus ponens. This suggests that the younger children were responding with a rather simple algorithm which led them to respond correctly to MP. Further, the simplicity of the algorithm would lead to shorter response times. The algorithm did not work as well with the other argument forms, however. Meanwhile, as the children were older, they responded on the basis of a more complex algorithm, and were only familiar enough with it at the oldest age level to have increased number of correct responses and shorter response times. This line of reasoning will be explored in subsequent analyses of the data.

The pattern of correct responses was similar with the Dutch students who participated in the present study to previous results mostly with English-speaking groups. This suggests that the underlying structure for logical reasoning develops independent of language and culture, as suggested by Piaget. However, there might be some subtle differences to be found in the logic employed by English and Dutch speaking children. This will be pursued in subsequent studies.

The present studies provided for the development and standardization of two interviews designed to pursue childrens' understanding of logical arguments and premises. Hopefully, these interviews will help fill the void recognized by other researchers in the field (e.g., Evans, 1972).

Finally, analysis of the data collected in Nijmegen in 1979 is continuing. Some future analyses will explore the effects of negation on explanation, the effects of separation variables on logical explanation and judgment, the relationship of materials in the premises to logical explanation, and part vs. whole presentation on reasoning.

References

Angell, R. B. Reasoning and logic. New York: Appleton-Century-Crofts, 1964.

Antonak, R. F., & Roberge, J. J. Characteristics of the conditional reasoning of educable retardates, American Educational Research Journal, 1978, 15, 519-528.

Austin, J. L. & Howsam, A. G. Language and mathematical education. Educational Studies in Mathematics, 1979, 10, 161-197.

Begg, I., & Denny, J. P. Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning errors. Journal of Experimental Psychology, 1969, 81, 351-354.

Braine, M. D. On the relation between the natural logic of reasoning and standard logic. Psychological Review, 1978, 85, 1-21.

Brainerd, C. J. The origins of the number concept. New York: Praeger, 1979.

Brainerd, C. J. Piaget's theory of intelligence. Englewood Cliffs: Prentice-Hall, 1978.

Brainerd, C. J. Judgments and explanations as criteria for the presence of cognitive structures. Psychological Bulletin, 1973, 79, 172-179.

Brown, M. Cognitive development and the learning of mathematics, In A. Floyd (Ed.) Cognitive development in the school years. New York: Halstead, 1979.

Brown, A. C. & DeLoach, J. S. Skills, plans, and other self-regulation. In R. Siegler (Ed.) Children's thinking: What develops? Hillsdale, New Jersey: Erlbaum, 1978.

Bynum, T. W., Thomas, J. A. & Weitz, L. J. Truth-functional logic in formal operational thinking. Inhelder and Piaget's evidence. Developmental Psychology, 1972, 7, 129-132.

Case, R. Intellectual development from birth to adulthood: A Neo-Piagetian interpretation. In R. S. Siegler (Ed.) Children's thinking: What develops? Hillsdale, New Jersey: Erlbaum, 1978.

Cersaso, J. & Provitera, A. Sources of error in syllogistic reasoning. Cognitive Psychology, 1971, 2, 400-410.

Chapman, L. J. & Chapman, J. P. Atmosphere effect re-examined. Journal of Experimental Psychology, 1959, 58, 220-226.

Cohen, J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 1960, 20, 37-46.

Connolly, F. M. & Binns, R. W. Logical reasoning in science education and an annotated bibliography. Unpublished report. 1974 (ERIC Document Reproduction Service No. ED099727)

Copeland, R. W. How children learn mathematics. New York: McMillan, 1979.

Copi, I. M. Introduction to logic. New York: Macmillan, 1961.

Corso, J. F. The experimental psychology of sensory behavior. New York: Holt, Rinehart & Winston, 1967.

Cronbach, L. J. Essentials of Psychological testing. (2nd Ed.), New York: Harper & Row, 1960.

Danner, F. & Day, M. C. Eliciting formal operations. Child Development, 1977, 48, 1600-1606.

DeCorte, E. Some aspects of research on learning and cognitive development in Europe. Educational Psychologist, 1977, 12, 197-206.

DeLeeuw, L. Onderzoek naar het effect van het aanlernen van algorithmische en het heuristische oplossingsmethoden mede in verband met persoonskenmerken. Lisse, Holland: Swets & Zeitlinger, 1979.

Dewey, J. Logic: The theory of inquiry. New York: Holt, 1938.

Dodwell, P. C. Relations between the understanding of the logic of classes and of cardinal number in children. Canadian Journal of Psychology, 1962, 16, 152-160.

Dodwell, P. C. Children's understanding of number and related concepts. Canadian Journal of Psychology, 1960, 14, 191-205.

Elkind, D., Anagnostopoulou, R., & Malone, S. Determinants of part-whole perception in children. Child Development, 1970, 41, 391-398.

Ennis, R. H. Children's ability to handle Piaget's propositional logic: A conceptual critique. Review of Educational Research, 1975, 45, 1-41.

Ennis, R. H. & Paulus, D. Critical thinking in grades 1-12, Phase 1: Deductive reasoning in adolescence. Ithaca, N.Y.: Cornell University, 1965 (ERIC Document Reproduction Service No., ED 003 818).

Erikson, J. R. Research on Syllogistic reasoning. In R. Revlin & R. E. Mayer (Eds.) Human Reasoning, Washington, D.C.: Winston, 1978.

Evans, J. On the problems of interpreting reasoning data: Logical and psychological approaches. Cognition, 1973, 1, 373-384.

Farnham-Diggory, S. Cognitive processes in education: A psychological preparation for teaching and curriculum development. New York: Harper & Row, 1972.

Flavell, J. H. Cognitive development. Englewood Cliffs, N.J.: Prentice-Hall, 1977.

Freedle, R. Psychology, thomian topologies, deviant logics, and human development. In N. Datan, & H. W. Reese (Eds.) Life-span developmental psychology: Dialectical perspectives on experimental research. New York: Academic Press, 1977.

Gelman, R. Preschool thought. American Psychologist, 1979, 34, 900-905.

Groza, V. S. A survey of mathematics: Elementary concepts and their historical development. New York: Holt, Rinehart and Winston, 1968.

Hadar, N. B. Children's conditional reasoning: An investigation of fifth graders' ability to learn to distinguish between valid and fallacious inferences. Unpublished doctoral dissertation, University of California, Berkley, 1975.

Halpern, E. The effect of incompatibility between perception and logic in Piaget's stage of concrete operations. Child Development, 1965, 36, 491-497.

Helsabeck, F. Syllogistic reasoning: Generation of counter examples. Journal of Educational Psychology, 1975, 67, 102-108.

Henle, M. On the relation between logic and thinking. Psychological Review, 1962, 69, 366-378.

Hill, S. A. A study of the logical abilities of children. (Doctoral dissertation, Stanford University) Dissertation Abstracts, 1961, 21, 3359. (University Microfilms No. 61-1229).

Homa, D. Organization and long-term memory search. Memory and Cognition, 1973, 1, 369-379.

Huttenlocher, J. Constructing spatial images: A strategy in reasoning. Psychological Review, 1968, 75, 550-560.

Inhelder, B. & Piaget, J. The growth of logical thinking from childhood to adolescence. New York: Basic Books, 1958.

Johnson-Laird, P. N. Models of deduction. In R. J. Falmagne (Ed.) Reasoning: Representation and process in children and adults. Hillsdale, N.J.: Erlbaum, 1975.

Johnson-Laird, P. N., & Taggart, J. How implication is understood. The American Journal of Psychology. 1969, 82, 367-373.

Judd, S. A., & Mervis, C. B. Learning to solve class-inclusion problems: The roles of quantification and recognition of contradiction. Child Development, 1976, 50, 163-169.

Klahr, D., & Wallace, J. G. Class-Inclusion processes. In S. Farnham-Diggory (Ed.). Information processing in children. New York: Academic Press, 1972.

Knifong, J. Logical abilities of young children - two styles of approach. Child Development, 1974, 11, 21-28.

Kodroff, J. K., & Roberge, J. J. Developmental analysis of the conditional reasoning abilities of primary-grade children. Developmental Psychology, 1975, 11, 21-28.

Kuhn, D. Conditional reasoning in children. Developmental Psychology, 1977, 13, 342-353.

Koster, W. G. Attention and performance II: Acta Psychologica, 1969, 30, 412-431.

Light, R. J. Measures of response agreement for qualitative data: some generalizations and alternatives." Psychological Bulletin, 1971, 76, 365-377.

Martorano, S. R. What is a formal operations task? Paper presented at the Annual meeting of the American Educational Research Association, Toronto, Canada, 1978.

Mason, E. J., & Bramble, W. J., T. A. Familiarity with content and conditional reasoning. Journal of Educational Psychology, 1975, 67, 238-242.

Mason, E. J. Haars, V. J. E., Dingemans, P. J. M., Bierman van Duuren, P. Two structured interviews for studying development of logical reasoning in children and adolsecents. Nijmegen, Netherlands: Vakgroep Ontwikkelings-psychologie, University of Nijmegen, 1979.

Mayer, R. E. Effects of meaningfulness on the representation of knowledge and process of inference for mathematical problem-solving. In R. Revlis & R. E. Mayer (Eds.) Human Reasoning. New York: Wiley, 1978.

Newell, A., & Simon, H. A. Human problem solving. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Nolen, P.A. Implications of formal operational thinking at the college level. Contemporary Educational Psychology, 1976, 1, 269-273.

Northrup, L. D. The definition and measurement of reasoning. Washington, D.C.: U.S. Civil Service Commission, Personnel Measurement Research and Development Center, 1977, (ERIC document No. ED 150 178).

O'Brien, T. C. Logical thinking in College Students. Educational Studies in Mathematics, 1973, 5. 71-79.

O'Brien, T. C., & Shapiro, B. J. The development of logical thinking in children. American Educational Research Journal, 1968, 5, 531-542.

Osherson, D. N. Logical abilities in children. Vol. 3. Reasoning in adolescents: Deductive Inference. Hillsdale, N. J.: Erlbaum, 1975.

Parsons, C. Inhelder and Piaget's The growth of logical thinking, II.: A logician's viewpoint. British Journal of Psychology, 1960, 51, 75-84.

Piaget, J. Logic and psychology. Manchester, England: Manchester University Press, 1953.

Piaget, J. Psychology of intelligence. Paterson, N.J.: Littlefield, Adams, 1963.

Piaget, J. Intellectual evolution from adolescence to adulthood. Human Development, 1972, 15, 1-12.

Pitt, R. B. Toward a comprehensive model of problem-solving: Application to solution of chemistry problems by high school and college students. Unpublished doctoral dissertation, University of California, San Diego, 1976.

Potts, C. R. The role of inference in memory for real and artificial information. In R. Revlin, & R. E. Mayer (Eds.) Human reasoning, New York: Wiley, 1978.

Rescher, N. Plausible reasoning: An introduction to the theory and practice of plausible inference. Atlantic Highlands, N.J.: Humanities, 1976.

Revlin, R., & Lierer, V. O. The effects of personal biases on syllogistic reasoning: Rational decisions from personalized representations. In R. Revlin, & R. E. Mayer (Eds.) Human reasoning, New York: Wiley, 1978.

Roberge, J. J. A study of children's abilities to reason with basic principles of deductive reasoning. American Educational Research Journal, 1970, 7, 583-596.

Roberge, J. J. An analysis of response patterns for conditional reasoning. Psychonomic Science, 1971, 22, 338-339.

Roberge, J. J. Linguistic and psychometric factors in propositional reasoning. Quarterly Journal of Experimental Psychology, 1978, 30, 705-716.

Roberge, J. J., & Antonak, R. F. Effects of familiarity with content on propositional reasoning. Journal of General Psychology, 1979, 100, 35-41.

Roberge, J. J. & Flexer, B. K. Further examination of formal operations reasoning abilities. Child development, 1979, 50, 478-484. (a)

Roberge, J. J., & Flexer, B. K. Propositional reasoning in adolescence. Journal of General Psychology, 1979, 100, 85-91. (b)

Roberge, J. J. & Mason, E. J. Effects of negation on adolescents' class and conditional reasoning abilities. Journal of General Psychology, 1978, 98,

Robinson, W. S. The statistical measurement of agreement. American Sociological Review, 1957, 22, 17-27.

Shapiro, B. J., & O'Brien, T. C. Logical thinking in children ages six through thirteen. Child Development, 1970, 41, 823-829.

Siegler, R. S. The origins of scientific reasoning. In R. S. Siegler (Ed.), Children's thinking: What develops? Hillsdale, N.J.: Erlbaum, 1978.

Simon, H. A. The logic of decision-making. In N. Rescher (Ed.) The logic of decision and action. Pittsburg: University of Pittsburg Press, 1967.

Simon, H. A., & Newell, A. Thinking processes. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), Contemporary developments in mathematical psychology. Vol.I. Learning, memory, and thinking. San Francisco: Freeman, 1974.

Smedslund, J. Circular relation between understanding and logic. Scandanavian Journal of Psychology, 1970, 11, 217-219,

Smedslund, J. Piaget's psychology in practice, British Journal of Educational Psychology, 1977, 47, 1-6.

Staudenmayer, H. Understanding conditional reasoning with meaningful propositions. In R. J. Falmagne (Ed.) Reasoning: Representation and process in children and adults. Hillsdale, N.J.: Erlbaum, 1975.

Trabasso, T., Isen, A. M., Dolecki, P., McLanahan, A. G., Riley, C. A., & Tucker, T. How do children solve class-inclusion problems? In R. S. Siegler (Ed.) Children's thinking: What develops?, Hillsdale, N.J.: Erlbaum, 1978.

Trabasso, T., Riley, C. A., & Wilson, E. G. The representation of linear order and spatial strategies in reasoning: A developmental study. In R. F. Falmagne (Ed.) Reasoning: Representation and process in children and adults. Hillsdale, N.J.: Erlbaum, 1975.

Van Duyne, P. C. Realism and complexity in reasoning. British Journal of Psychology, 1974, 65, 59-67.

Vergnaud, G. The acquisition of arithmetical concepts. Educational Studies in Mathematics, 1979, 10, 263-274.

Von Wright, G. H. Logical studies. London: Routledge & Kegan Paul, 1957.

Weitz, L. J., Bynum, T. W., Th[illegible], J. A., & Steger, J. A. Piaget's system of binary operationals. Am[illegible]irical investigation. Journal of Genetic Psychology, 1973, 123, 279-[illegible]

Wilkins, M. C. The effect of [illegible]d material on the ability to do formal syllogistic reasoning. Arc[illegible] of Psychology, 1928, 16, No. 102.

Wilkinson, A. Counting strateg[illegible] and semantic analysis as applied to class inlcusion. Cognitive Psych[illegible]y, 1976, 8, 64-85.

Wohlwill, J. F. The learning of absolute and relational number discriminations by children. Journal of Genetic Psychology, 1962, 101, 217-228.

Woodworth, R. S. & Sells, S. B. An atmosphere effect in formal syllogistic reasoning. Journal of Experimental Psychology, 1935, 18, 451-460.

APPENDIX

INSTRUCTIES AAN DE PROEFPERSONEN

Ik ga je een paar dingen zeggen, zoals bijvoorbeeld:

DIT IS MARIA

en dan zal ik iets over haar vertellen zoals: ze heeft een blauwe jurk aan.

Ook zal ik je er direct een paar vragen over stellen. Jij moet er achter zien te komen of het klopt wat ik zeg.

Je kunt op drie manieren antwoord geven:

1) JA. DAT BETEKENT DAT HET MOET KLOPPEN.
2) NEE. DAT BETEKENT DAT HET NIET KAN KLOPPEN.
3) MISSCHIEN. DAT BETEKENT DAT HET SOMS MISSCHIEN WEL KLOPT, MAAR DAT IK JE NIET GENOEG VERTELD HEB OM HET ZEKER TE WETEN.

Bijvoorbeeld, stel je voor dat ik zeg:

MARIA IS (BEVINDT ZICH) NAAST DE KNOOP.

DAN IS (BEVINDT ZICH) DE KNOOP NAAST MARIA.

Is het juist als ik dan zeg dat de knoop naast Maria is?

(Wacht op het antwoord van het kind, zeg daarna:)

Ja is het goede antwoord.

Laten we een ander voorbeeld proberen:

EEN KOE IS GROTER DAN EEN KAT.

DAN IS EEN KAT GROTER DAN EEN KOE.

Is het juist als ik dat zeg?

(Wacht op het antwoord van het kind, zeg daarna:)

Nee is het goede antwoord omdat een koe niet groter en kleiner kan zijn dan een kat.

Laten we nog een ander voorbeeld proberen:

KAAS IS VOEDSEL.

DAN IS VOEDSEL KAAS.

Klopt het als ik zeg dat voedsel kaas is?

(Wacht op het antwoord van het kind, zeg daarna:)

Denk er aan dat je alleen JA, NEE of MISSCHIEN kunt zeggen.

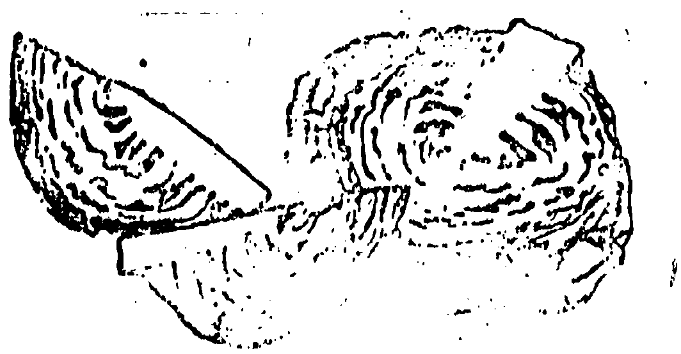
(Wacht op het antwoord van het kind, zeg daarna:)

Het antwoord hier is MISSCHIEN omdat voedsel kaas kan zijn of een ander ding dat je kunt eten zoals brood of vlees. Begrijp je dat?

Als het kind ja zegt, vraag hem dan om uitleg.

Zegt het kind nee, leg het dan daarna nog eens uit en gebruik daarbij plaatjes.

1a



1. De proefleider zet bak A voor het kind neer en zegt:

ALS ALLE HONDEN DIEREN ZIJN (terwijl de woorden uitgesproken worden, wijst de proefleider op de hond en dan op de rest van de dieren).

EN DIT IS EEN HOND. (de proefleider laat het beeldje van de hond zien die buiten het gezichtsveld was gehouden).

DAN MOET IK DEZE BIJ DE DIEREN ZETTEN (de proefleider zet de hond in de bak).

A. KLOPT DAT?

Als het kind geen standaard antwoord geeft, (ja, nee, misschien) zegt proefleider:

DENK ERAAN DAT JE ALLEEN JA KUNT ZEGGEN. DAT BETEKENT DAT HET MOET KLOPPEN. NEE, DAT BETEKENT DAT HET NIET KAN KLOPPEN. OF-
MISSCHIEN, DAT BETEKENT DAT HET SOMS KAN KLOPPEN, MAAR DAT IK JE NIET GENOEG VERTELD HEB OM ER ZEKER VAN TE ZIJN.

Wanneer het kind een standaard antwoord geeft, vraagt de proefleider:

B. WAAROM ZEG JE DAT?

Na het antwoord van het kind, gaat de proefleider door:

C. VERTEL ME WAT HET BETEKENT ALS IK TEGEN JE ZEG,
ALLE HONDEN ZIJN DIEREN.

Wanneer het kind antwoord geeft, vraagt de proefleider:

D. ZIJN ER MEER HONDEN OF DIEREN? Dan vraagt de proefleider: WAAROM?

Wanneer het kind antwoord geeft, gaat de proefleider door: VERTEL ME NU WAT DIT BETEKENT:

E. DIT IS EEN HOND (De proefleider wijst op de hond in de bak)
DAN MOET IK DEZE BIJ DE DIEREN ZETTEN.

(1.) De proefleider plaatst een blad voor het kind en legt erop:

9 groene kubussen (bij elkaar)

Hij zegt dan: ALS ALLE BLOKKEN VAN HOUT ZIJN (Pl. laat op blad geplaatste blokken zien)

EN DIT IS NIET VAN HOUT (Pl. laat groene kubus zien)

DAN IS HET EEN BLOK (Pl. plaatst groene kubus buiten blad)

A.............KLOPT DAT?

Opmerking 1: Als het kind geen standaard-antwoord geeft, zegt de proefleider: DENK ERAAN, JE KUNT ALLEEN MAAR JA ZEGGEN, DAT BETEKENT DAT HET MOET KLOPPEN, OF NEE, DAT BETEKENT DAT HET NIET KAN KLOPPEN, OF MISSCHIEN, DAT BETEKENT DAT IK JE NIET GENOEG VERTELD HEB OM ER ZEKER VAN TE ZIJN.

Wanneer het kind een standaard-antwoord heeft gegeven, vraagt de proefleider:

B.............WAAROM ZEG JE DAT?

Opmerking 2: Het kind moet in staat gesteld kunnen worden de proefleider duidelijk aan te tonen dat hij/zij de volledige betekenis van de implicatie in de premissen en de conclusie beschreven heeft zoals hij/zij die begrijpt. De proefleider moet het kind blijven ondervragen m.b.t. het antwoord totdat dat duidelijk is.

(2.) De proefleider plaatst een blad voor het kind en legt erop:

8 groene kubussen en 1 gele cirkel (bij elkaar)

Hij zegt dan: ALS ALLE BLOKKEN VAN HOUT ZIJN (Pl. laat op blad geplaatste blokken zien)

EN DIT IS EEN BLOK (Pl. laat gele cirkel zien)

DAN IS HET NIET VAN HOUT (Pl. plaatst gele cirkel buiten blad)

A.............KLOPT DAT?

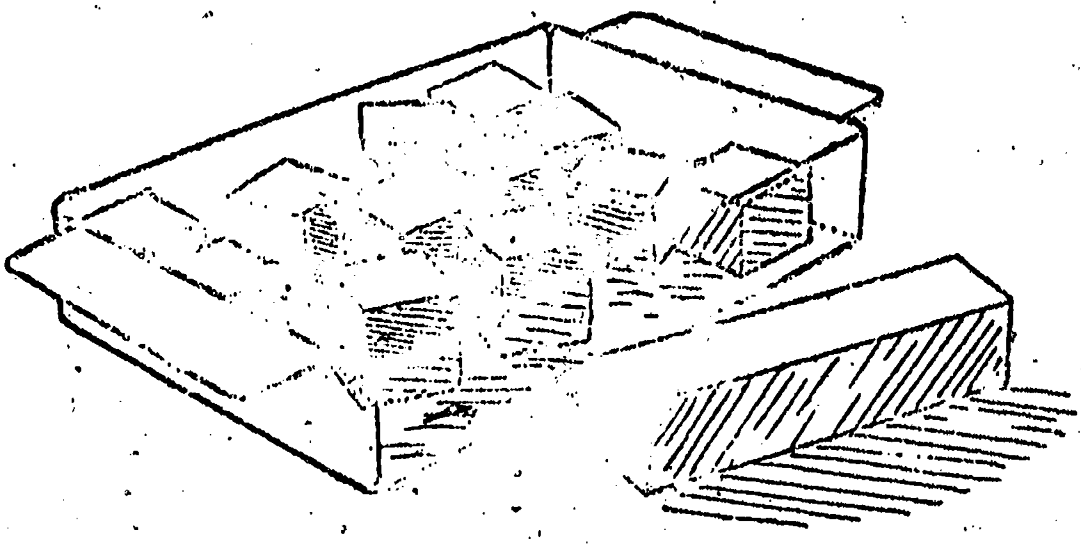
Als het kind geen standaard-antwoord geeft, zie Opmerking 1.

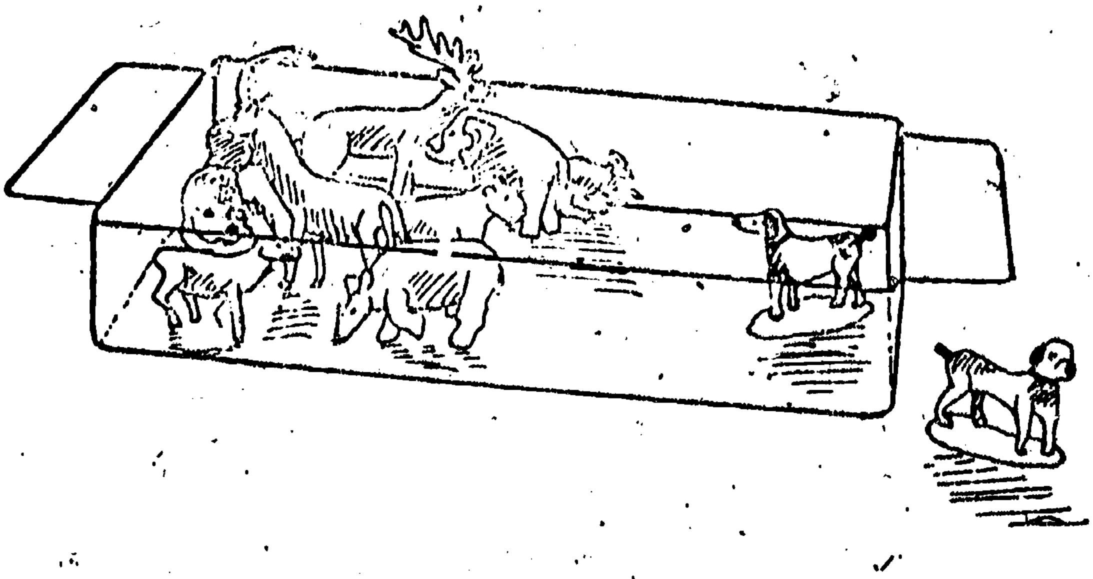
Wanneer het kind een standaard-antwoord heeft gegeven, vraagt de proefleider:

B.............WAAROM ZEG JE DAT?

Zie Opmerking 2 m.b.t. het antwoord.

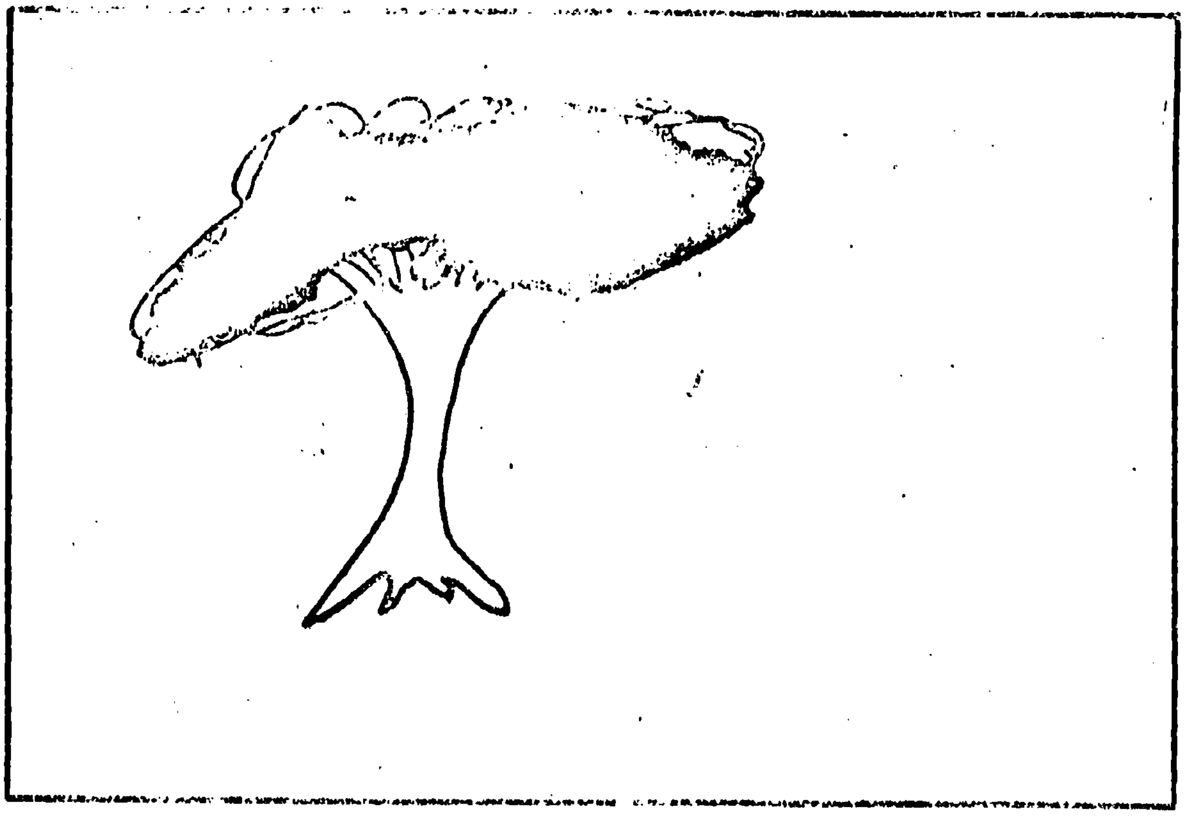
1d





1b

1c



DEEL I

Presentatie: Geheel-Concreet

De proefleider heeft twee bakken. De éne bak (A) bevat 10 verschillende kleine beeldjes van dieren (waaronder één hond) die at random neergezet zijn. De andere bak (B) bevat verscheidene groene kubussen van 2.5 cm.
Er is één hond en een niet-gekleurd blokje van 2.5x2.5x8 cm die buiten het gezichtsveld van het kind worden gehouden.

De éne hond in de bak komt apart te staan van de andere 9 dieren.

De mogelijkheid bestaat dat het kind, vooral in de jongere leeftijdsgroepen, niet weet wat het woord 'kubus' betekent. De proefleider moet er daarom naar vragen en eventueel uitleggen.

Wanneer je het kind een object (hond, kubus) laat zien en je ontkent het (Dit is geen hond, geen kubus), bedoel je: laten we het doen alsof.

De proefleider moet het kind ertoe aanzetten redenen te geven. Acceptabele vragen die gebruikt kunnen worden om het kind aan te sporen zijn bijv.:
"Waarom zeg je dat?"
"Kun je me er meer over vertellen?"
"Wil je dat verder uitleggen?"
"Wat bedoel je daarmee?" etc.

Het principe dat de proefleider steeds in de gaten moet houden bij ondervraging van het kind is dat het kind geen aanwijzing gegeven mag worden hoe de vraag beantwoord moet worden.

Indien het kind de instructies in A,B,C,D of E niet begrijpt, herinner het er dan aan wat er gevraagd wordt.
Bijvoorbeeld: Indien de proefleider een reden verwacht en het kind schijnt te antwoorden m.b.t. de validiteit (of het klopt of niet), dan moet het ertoe aangespoord worden om het gewenste type antwoord te geven.
Let er ook goed op dat het kind geen aanwijzing gegeven wordt m.b.t. het correcte antwoord.

Vermijd, nadat het kind antwoord gegeven heeft, uitspraken als: "Dat is een goed antwoord". Beter is niet zozeer een uitspraak te doen m.b.t. de juistheid van het antwoord, maar meer gericht op het feit dat het kind goed zijn best doet, zoals bijv. "Je doet het goed", "Mooi zo!", od.

1.Enkele vragen om meer te weten te komen over het antwoord van het kind:

Wil je dat verder uitleggen?

Vertel er eens meer over.

Kun je me er meer over vertellen?

Waarom zeg je dat?'

Wat bedoel je?

2.Te antwoorden op"ik weet het niet"

a)Herhaal de vraag

b)Weet je wat P is? Qis?

indien "nee"-stop!

indien "ja" - herhaal de vraag

c)"Ik weet het niet"- ga verder met het volgende item.

1) Vermijd, nadat het kind antwoord gegeven heeft, uitspraken als: "Dat is een goed antwoord". Geen uitspraak dus m.b.t. de (on)juistheid van het antwoord, maar een uitspraak gericht op het feit dat het kind goed zijn best doet, zoals bijv. "Je doet het goed", "Mooi zo!", ed.

---

2) Het principe dat de proefleider steeds in de gaten moet houden bij ondervraging van het kind is dat het geen aanwijzing gegeven mag worden hoe de vraag beantwoord moet worden. Indien het kind de instructies in A of B niet begrijpt, herinner het er dan aan wat er gevraagd wordt.

---

3) De proefleider moet het kind ertoe aanzetten redenen te geven. Enkele vragen om meer te weten te komen over het antwoord van het kind:

"Wil je dat verder uitleggen?", "Vertel er eens meer over.", "Waarom zeg je dat precies?" of "Wat bedoel je daarmee?".

4) Wat er gedaan moet worden na het antwoord "Ik weet het niet.":

a) herhaal de vraag; indien hetzelfde antwoord:

b) "Weet je wat P is?", "Weet je wat Q is?"; indien "Nee.": stop met de vraag te stellen; indien "Ja.": herhaal de vraag nog een keer; indien "Ik weet het niet.":

c) stop met de vraag te stellen en ga verder met het volgende item.

---

tape______ Name________ Grade__________ School__________ Examiner__________ Scorer__________ FORM CA

K-R Coding Form.

ITEMS

Categories of Explanation	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1 Verbal Deficits																																
2 Memory Deficits																																
3 Arbitrary Explanations																																
4 Structural Deficits																																
5 Conditional Reasoning																																
6 Patterned Explanations																																
7 Others																																
8 Good Logic																																

Place response to syllogism (yes, no, maybe) in box corresponding to explanation, and circle correct responses.

SCORING INSTRUCTIONS FOR
CODING FORM C-L

1. In general, the scorer will place a check in the boxes where the kind of statement the child made is represented at the top of the column. For example, if the child says, "Dogs and animals are the same thing.", checks should be placed in the P=Q column and the PQ column. Also, a check would be place in the P.Q column because the child recogizes that both P and Q can exist at the same time. The specific criteria to use with scoring each column are listed below. The only violation to the check placement in a column is that if the child reverses P.Q in his verbal response, place an R in the appropriate columns instead of a check. For example, if the major premise says, "All P are Q" and the child says "There are Qs and there are Ps". Since the child said that Q and P exist in his answer, we score it as P.Q, and place an R in the column because of the reversal. More than one column can be scored (✓ or R) for each item.

2. The three sections of each item that have several-word responses are scored. Each section has a line on the form. There are two forms, one for Parts I and III, and the other for II and IV items.

3. The meanings of the symbols in each column is as follows:

A. Class Relationships

a) P=Q : Response indicates P and Q are the same in number or size.
ex: "They are the same". "There are dogs and there are animals".

b) P<Q : There are more Q than P.

c) P>Q : There are more P than Q.

d) (P)(Q) : P and Q exist as separate variables.
ex: "There are dogs and there are animals". "A thing is either a tree or a plant".

e) (PQ) : P and Q are the same.
ex: "All dogs are animals and all animals are dogs".

f) (P(Q) : There is some overlap between P and Q.
ex: "Some dogs are animals". "Some animals are dogs".

g) ((P)Q) : P exists as a subclass of Q.
ex: "A dog is an animal". "All dogs are animals".

h) ((Q)P) : Q is a subordinate group to P.

ex: "All animals are dogs","An animal is a dog".

i) Other: Put a symbolfor an other kind of response that needs to be récorded. In the notes area below, the symbol should be explained. Ex:"There are dogs". This would be scored P only.

B. Logical Relationships

a) P : only the existence of P is recognized.

ex:"There is a dog".

b) Q : only the existence of Q is recognized.

ex: "There are plants".

c) $\bar{P}$: only the existence of the negative of P is recognized.

ex: "There is no dog". "This not a ball".

d) $\bar{Q}$: only the existence of the negation of Q is recognized.

ex : "There is no animal."

e) P.Q : The existence of both P <u>and</u> Q is recognized.

ex : "There are dogs and there are animals.", "You said all dogs are animals", "Animals can be dogs and dogs can be animals."

f) $\bar{P}$.Q : recognizes that the negation of P can exist while Q exist.

ex : "There is no dog, but there are animals", "All dogs are animals, but not all animals are dogs."

It can be inferred from some answers that P Q is intended.

ex : "Every dog must be an animal too.", implies that animal is superordinate, and animals can exist that are not dogs.

g) $\bar{P}.\bar{Q}$: recognizes that the negation of P and Q can exist at the same time.

ex : "There are no dogs or animals."

This answer can also be inferred from correctly applied implication or other answers in which P.Q is not stated directly.

ex :"All dogs are animals, but not all animals are dogs.", "If this is a dog, than it must be an animal, because all dogs are animals.", "If this were not a dog, it might or might not be an animal."

h) (P.$\bar{Q}$) : says that P can exist even when Q does not.

ex : "There are no round things, but this is a ball anyway.", "This is a dog but not an animal."

i) $N(P.\bar{Q})$: denies that P and $\bar{Q}$ can exist at the same time.

ex: "There are no dogs that are not animals".

This can sometimes be inferred from a good implication response, ex:

"This dog <u>must</u> be an animal because all dogs are animals."

j) $P \rightarrow R$ / $R \rightarrow S$ means that a term was added. For example,"all trees are green and all plants are green so this must be a plant."

k) Other: Mark any other kind of response that should be recorded (dealing with logical understanding) with a symbol and explain the symbol at the bottom of the page under "notes".

—FORM CA—

C-L Coding Form

Name________________ Examiner________________

Tape________________ Scorer________________

Date scored__________

ITEMS	A. Class relationships									B. Logical Relationships										
	P=Q	P>Q	P<Q	(P) (Q)	(PQ)	(P(Q))	((P) Q)	((Q) P)	other	P	Q	$\bar{P}$	$\bar{Q}$	P·Q	$\bar{P}$·Q	$\bar{P}\cdot\bar{Q}$	$(P\cdot\bar{Q})$	$N(P\cdot\bar{Q})$	P→R R→Q	OTHER
1																				
2																				
3																				
4																				
5																				
6																				
7																				
8																				
9																				
10																				
11																				
12																				
13																				
14																				
15																				
16																				
17																				
18																				
19																				
20																				
21																				
22																				
23																				
24																				
25																				
26																				
27																				
28																				
29																				
30																				
31																				
32																				

-FORM PW-

C-L Coding Form

Name______________ Examiner______________

Tape______________ Scorer______________

Date scored__________ Part I III (circle)

ITEMS	A. Class relationships									B. Logical Relationships										
	P=Q	P>Q	P<Q	(P) (Q)	(P Q)	(P)(Q)	((P) Q)	((Q) P)	other	P	Q	$\bar{P}$	$\bar{Q}$	P•Q	$\bar{P}$•Q	$\bar{P}$•$\bar{Q}$	(P•$\bar{Q}$)	N(P•$\bar{Q}$)	P→R R→Q	OTHER
1B																				
1C																				
1E																				
2 B																				
2 C																				
2 E																				
3 B																				
3 C																				
3 E																				
4 B																				
4 C																				
4 E																				
5 B																				
5 C																				
5 E																				
6 B																				
6 C																				
6 E																				
7 B																				
7 C																				
7 E																				
8 B																				
8 C																				
8 E																				

NOTES: