

1980

Adolescent Reasoning: Influence of Question Format and Type of Variables on Ability to Control Variables

David I. Levine

UC Hastings College of the Law, levined@uchastings.edu

Marcia C. Linn

Follow this and additional works at: http://repository.uchastings.edu/faculty_scholarship

Recommended Citation

David I. Levine and Marcia C. Linn, *Adolescent Reasoning: Influence of Question Format and Type of Variables on Ability to Control Variables*, 62 *Sci. Educ.* 337 (1980).

Available at: http://repository.uchastings.edu/faculty_scholarship/1335

This Article is brought to you for free and open access by UC Hastings Scholarship Repository. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of UC Hastings Scholarship Repository. For more information, please contact marcus@uchastings.edu.

Adolescent Reasoning: Influence of Question Format and Type of Variables on Ability to Control Variables

MARCIA C. LINN

University of California, Berkeley, California 94720

DAVID I. LEVINE

University College, London, England

Introduction

The ability to control variables, to keep all but one variable the same so the effect of the variable that is changed can be investigated, is relevant to many everyday decisions. For instance, one must control for container size when comparing prices.

Inhelder and Piaget[1] conducted the first major investigation of adolescent thinking. Part of their research involved controlling variables, or the schema Piaget calls "all other things equal." By conducting experiments using 15 varied tasks, they found that learners progress from what they call concrete operations to what they call formal operations during adolescence. A clearer understanding of how children process questions about controlling variables and validation of possible theoretical explanations for changes in ability to control variables are needed. For a survey of recent work see Levine and Linn[2].

Pascual-Leone[3] has recently suggested a neo-Piagetian theory which places importance on the way information is presented in a question. This paper focuses on the kind of information presented in the question (question format) and the subjects' familiarity with the variable in the questions. To investigate question format, we asked questions which included three different kinds of information. The three questions and the information in each were:

(1) *Free response question (Free)*. Subjects were simply asked to propose an investigation of the effect of a particular variable in the problem. No further information was given.

(2) *Multiple choice question (MC)*. Subjects were asked to help design an experiment to investigate a particular variable in the problem. Several alternative approaches were proposed. The subject chose the one he preferred and explained his choice.

(3) *Screen question (Screen)*: Subjects were shown the results of an experiment about one variable while a screen concealed the exact procedure. Subjects were then asked what could be found out from the information that they had.

To investigate the role of the subjects' familiarity with the variables and apparatus: (a) Three problems were compared which were of varying degrees of familiarity to the subjects; (b) for one of the problems (Ramp) we assessed subjects' knowledge of the variables involved.

Methods

Subjects

The 120 subjects came from a large comprehensive school (for 11- to 17-year olds) in a middle-class suburban area of London. Forty subjects, half boys and half girls, came from each of three age groups: 11.9 to 12.8 (12-year olds); 13.10 to 14.7 (14-year olds); 15.9 to 16.8 (16-year olds). Subjects were selected randomly from the approximately 90% who volunteered to participate.

Procedure

Subjects were randomly assigned to one of two interviewers who randomly selected one of two 20 minute interviews: (1) the Ramp Problem (Ramp) or (2) the Circuit Problem (Circuit) and the Seed Problem (Seeds).

Ramp Problem

Wollman[4,5] developed a group version of the ramp problem. The current version employs some similar questions but differs in important respects (see Linn[6]).

The apparatus consisted of a curved ramp made of 1 cm wide aluminum channeling, and glass or steel marbles of different sizes and colors (Fig. 1). For Ramp, the variables were height of the release point, weight of the marble, and weight of the target.

Ramp: Free response question (Free). The interviewer explained the apparatus and then introduced the Free question by saying: "Suppose you have been challenged to a marble rolling contest with a boy/girl in your class. He/she will choose the position that both of the marbles in this box will be released from. [Take out closed box with steel and blue marbles inside.]

"You can decide which of the two marbles he/she will use and which one you will use. [Sex was chosen to agree with that of the student.] Before the contest, you can try each marble once to find out which marble is best. For these trials, you can use whatever positions you want, but remember, for the contest, your friend is going to choose one position for both marbles."

After the subject suggested a way to test the marbles, the interviewer showed the marbles to the subjects and used one of two probes: (a) If the subject said to test the marbles from the same position, the experimenter asked if it would be just as good to test from unequal positions, or (b) if the subject wanted to test the marbles from unequal positions, the experimenter asked about using the same position.

Ramp: Variables. Subjects were asked eight questions to measure their knowledge about the three basic variables involved: weight of the marble, weight of the target, and height of the release point.

Ramp: Multiple choice question (MC). A picture of a large marble ready to be released from position 2 to hit a blue target and four choice pictures of possible experiments with the ramp, target, and marble, drawn on index cards, were placed directly in front of the

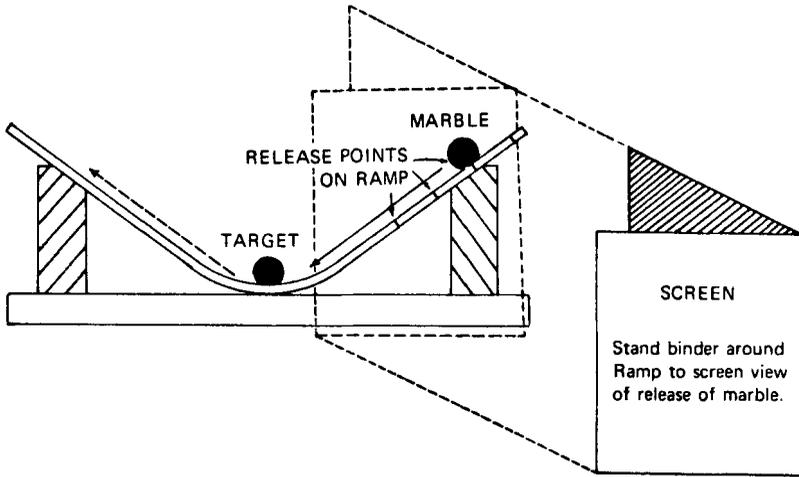


Figure 1. Apparatus for Ramp problem.

subject. The choice pictures showed large or small marbles ready to be released from position 2 or 4 to hit either the blue target or another target. Each subject was then told: "Another child in your class has also been challenged to a marble rolling contest, but the marbles will be different. He tried the large marble from position 2 with the blue target. He doesn't know which of these other experiments to do to find the best marble for the contest. Can you tell which would be the test to do to find out which marble to use in the contest?" After the subject chose a response, he was asked why he chose that one and why he did not choose each of the others.

Ramp: Screen. A blinder was placed over half of the ramp as in Figure 1. The subject could see how far up the ramp the target went when it was hit by each marble in turn but not where the marbles were released from. (In fact, two different release points were used such that a smaller marble hit the target farther.) The subject was then asked: "Which marble hit the target the farthest? From what you just saw, which marble would you choose for the contest? Why?" The screen was removed and the interviewer said: "Now I'll show you what I did. I released the clear marble from here [high position] and the other from here [low position]. Will the clear one still win when they are released from the same position?"

Circuit Problem

The apparatus for the Circuit problem consisted of a metal box (Fig. 2) with a labeled on-off switch, and three pairs (or "trees") of wires on the top. Each pair came out of its own hole. For Circuit the three variables were the trees: the blue-white tree, the green-red tree, and the black-yellow tree. The circuit was wired so that it could not buzz unless the one critical wire was selected as part of a three wire group. The interviewer introduced the apparatus and explained that the subject was to find out if either the blue or white wire was connected to the buzzer by trying the wires in groups of threes. As in the Ramp problem, subjects were asked a Free, MC, and Screen question about the Circuit problem.

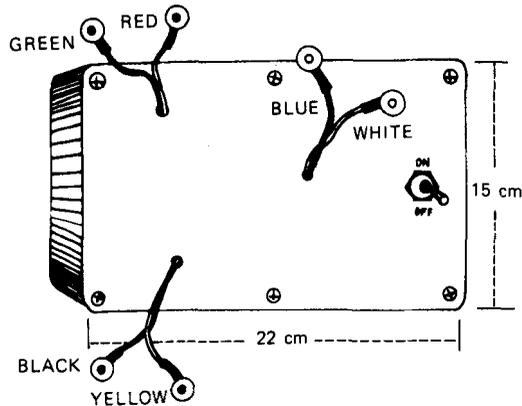


Figure 2. Apparatus for Circuit problem.

Seed Problem

This problem was included in the interview with the circuit problem. Only the Free question was used for Seeds. The three variables were: the amount of soil, amount of water, and amount of fertilizer. At the end of the Circuit problem, subjects were asked the Seed problem as follows: "Suppose you had two kinds of seeds, white radish seeds and red radish seeds. You want to see which kind will grow the tallest in a week. You have two flower pots that are just the same, some soil, water, and fertilizer pellets. What would you do? How much soil would you put in the first pot? How much in the second?" [Similar questions were asked about the water and fertilizer.] [Finally they were asked:]

"Would it be just as good to use two spoons of water in the first and six spoons of water in the second pot?" Unlike Ramp or Circuit, Seeds was a purely verbal problem; no materials were presented to the subjects. Also unlike Circuit and Ramp, the Seeds question involved a concurrent comparison rather than a sequential comparison.

Sequence of Questions

We could not readily present the questions in all the possible sequences with the available sample size. Pilot tests indicated that no significant order effect occurs for these questions. Thus we chose a question sequence on the basis of the amount of information presented in each question. The question with the least information came first (Free) followed by MC and Screen. It should be noted that the easiest question was MC which occurred in the middle of the sequence.

Scoring

The two authors scored all interviews independently. Interrater reliability for each question ranged from 90–95%. Scores for subjects assigned to each interviewer did not differ significantly. The scoring system will not be explained in detail here. (Detailed descriptions of the categories are available from the authors.) Table I briefly provides the criterion for failure and success on each question.

Relationship to Piagetian theory. Relevant to our Free question, Piaget has described

the development of the “all other things equal schema” as proceeding from a preference for deliberately varying two factors at a time to a stage where the child performs “one by one multiplications,” that is, has a limited ability to control variables, to a final “structured” whole where the child realizes all variables not under investigation must be controlled. We found that for Free many subjects preferred to vary two factors at a time (scores of 1, 2, or 3 for the Ramp and 1, 2, 3, or 4 for the Circuit). Some had a limited understanding of controlling variables (scores of 5), and others understood the general principle of controlling variables (score of 6). For Free, only scores of 5 and 6 counted as successes.

Our category 4 for the Ramp Free question includes subjects who preferred to use the highest position on the ramp. These subjects appeared to be controlling variables, but when probed, had chosen high because it would “hit hardest” and were just as happy to have one trial at the high position, and the next trial at the low position. This preference

TABLE I
Criterion for Success and Failure on Each Question

FREE			
	Ramp	Circuit	Seeds
Failure	Proposes to test the two marbles from different positions Proposes to test both from high but accepts unequal positions	Proposes a six wire solution, such as White-Yellow-Green and then Blue-Black-Red	Proposes that the amount of either water, fertilizer, or soil should be different for the two seeds.
Success	Realizes that the marbles must be place at the same position.	Proposes a four* wire solution, such as White-Yellow-Green and then Blue-Yellow-Green	Proposes that the amounts should be the same for the two seeds
Screen			
	Ramp	Circuit	
Failure	With screen in place, subject chooses the marble that makes the target go farther	With screen in place, subject chooses the wire mentioned by the experimenter	
Success	With screen in place, subject will not make a decision based on the partial information available	With screen in place, subject will not make a decision based on the partial information available.	
MC - Ramp or Circuit		Variables - Ramp	
Failure	Incorrect choice	0-6 questions answered correctly	
Success	Correct choice	7 or 8 questions answered correctly	

* A five wire solution such as blue-black-red and blue-yellow-green was also correct, but did not occur in this sample.

for high as “best” is consistent with concrete thought, but is not a characteristic reported by Inhelder and Piaget.

As noted by Inhelder and Piaget in their research, some subjects prefer to test the same object twice, rather than to vary the factor under consideration. We prevented this from occurring in the format of our Free question, but not in the MC question. Many subjects who scored 1, 2, or 3 on MC wanted to test the same marble or wire twice. The remainder of those scoring 1, 2, or 3 wished to vary two factors at once. Thus, using Piaget’s system we could classify subjects who score 4 and below on Free or 3 and below on MC as concrete. In our system all of these scores were counted as “failures.”

Scoring of the Screen is more difficult to reconcile with Piaget’s system, since he has not used this approach, but it appears to be related to his concept of the “real” and the “possible.” Inhelder and Piaget [1] said that the concrete child is limited to noting the relationships that seem to “thrust themselves on him [p. 256]” while the formal child, “in order to avoid inconsistencies as new facts emerge [p. 256],” considers relationships he observes as possible and seeks to determine the actual relationships that prevail. Thus we could say that subjects who accept the evidence from the screen condition and score 1 or 2 (“failures”) are considering the real information only. Those who succeed recognize that the evidence from the screen is incomplete because they consider the possible relationships (scores of 3 or 4). By this formulation, subjects who are successful on each of our tasks are capable of thinking beyond the level of concrete operations.

Results

Each subject received either the Ramp problem or the Circuit problem and the Seed problem. The Ramp and Circuit problems each had three questions (Free, MC, Screen). The Seeds problem only had the Free question.

Because most measures were ordinal rather than interval, nonparametric statistics were used for data analysis. Differences between groups were analyzed using a z score determined from Kendall’s τ as a measure of correlation or by using the sign test to compare “success” on each question. Success meant that the subject understood that all but one variable must be controlled before the effects of that variable could be assessed. All significance levels are reported for two-tailed tests.

In this section we will report: (1) results for each task, (2) age differences, (3) sex differences, (4) comparisons between tasks for each of the three questions; and (5) the relationship between the three questions.

Results for Each Task

Success: Ramp. Figure 3 shows the percent of subjects at each age who were successful on each question for Ramp. The MC question was the easiest, Free and Screen were equally difficult for 12- and 14-year olds but Screen was easier than Free for 16-year olds.

Success: Seeds. As can be seen in Figure 3, all but four subjects correctly solved the Seeds Free question.

Age Differences

Free. As shown in Figure 3, there were no age differences for Seeds (nearly everyone

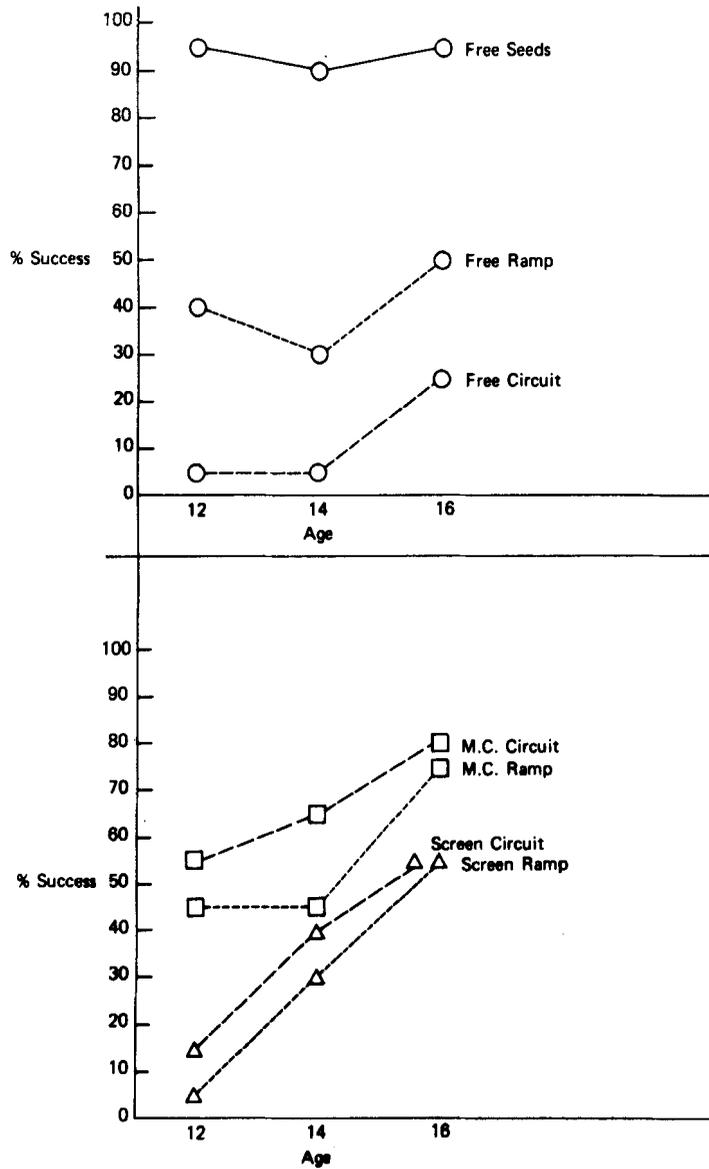


Figure 3. Percent of subjects succeeding on each question for each problem.

succeeded) but older children did better than younger on Free for both Circuit and Ramp: Ramp $z = 2.79$ ($p < 0.01$); Circuit $z = 2.50$ ($p < 0.05$).

MC and Screen. There were no significant age differences for MC although there was a trend toward increased success with age as shown in Figure 3: Circuit $z = 1.05$; Ramp $z = 1.78$. There were striking increases with age for Screen: Circuit $z = 3.29$ ($p < 0.001$); Ramp $z = 3.18$ ($p < 0.01$).

Sex Differences and Knowledge of Variables

There were no consistent sex differences across problems for any question. Within sexes, 67% answered all variables questions correctly; boys of all ages answered the variables questions equally well, while older girls did better than younger girls. More complete knowledge of the variables demonstrated by boys was not associated with greater success for boys on any Ramp question either by age, or overall.

On the Circuit, boys did better than girls for Free and MC: Free $z = 2.22$ ($p < 0.05$); MC $z = 2.21$ ($p < 0.05$). For the Ramp, the only sex difference was for variables, where boys did better than girls: $z = 2.21$ ($p < 0.05$). The sex differences for Circuit interact with age, since only 16-year-old boys are better than girls. One could hypothesize that this is due to some type of knowledge the older boys have. This information, however, does not also lead boys to be superior to girls on the Screen. Since for Ramp, additional information about variables did not help boys, it appears that either the information helping older boys in Circuit is not similar to the variables information in Ramp or that some other explanation must be found for the success of 16-year-old boys on Circuit.

Comparisons between Tasks

In this section we compare results for Circuit, Ramp, and Seeds.

Free question. On the Free question, there are significant differences between success rates for each problem: Circuit was solved by 7 of 60 (10.7%), Ramp by 24 of 60 (40.0%), and Seeds by 56 of 60 subjects (93.3%). As shown in Figure 3 these differences are consistent across ages. Comparing success rates: Circuit vs. Ramp, $z = 3.32$ ($p < 0.001$), Ramp vs. Seeds, $z = 5.98$ ($p < 0.0001$).

The tasks were chosen because they differed in familiarity. Almost every 12-year-old child in a British school has grown plants and is familiar with the variables in the Seeds problem. Many, perhaps most, children are familiar with inclined planes and marbles; 67% answered questions about the variables correctly. It is very unlikely that children have previous experiences with the Circuit problem, although they might try to apply information that they have about electricity.

Although familiarity is a logical explanation for the differences in performance in Free, there are other possible explanations. First, perhaps the method used for presenting the task gave more information about the variables for one task than for another.

Second, one might suppose that the apparatus as a concrete referent could possibly facilitate performance, but in this case the easiest problem, Seeds, had no apparatus.

Third, the Seeds task is similar to questions asked at school and might suggest a previously learned response.

Fourth, Seeds was a concurrent comparison rather than a sequential comparison since two flower pots were used. This would make the seeds Free question more like the Circuit or Ramp MC question. Seeds is still easier than MC for Circuit or Ramp.

As one would expect, results for Free support the hypothesis that subjects try to solve new problems by drawing upon apparently relevant past experiences.

MC and Screen questions. On MC and Screen, there were no differences between Circuit and Ramp as can be seen in Figure 3 (MC, $z = 1.12$, Screen $z = 0.57$). Thus while success on the Free question was related to the familiarity of the variables involved, success on MC and Screen was about the same for both Ramp and Circuit; this relationship will be discussed in detail below. It is interesting to note that while Free was significantly more

difficult for Circuit across ages, MC and Screen were slightly easier or equally difficult for Circuit as compared with Ramp at all ages.

Relationship between Questions

In this section we will first compare the relative success rates for Free, MC, and Screen. Secondly, we will report and discuss the correlation between scores on Free, MC, and Screen for each task.

Free and MC. Comparing success on MC and Free (Fig. 3), we have seen that MC is equally difficult for Circuit and Ramp while Free is harder for Circuit than for Ramp. In both Circuit and Ramp, Free is more difficult than MC: Circuit $z = 5.54$ ($p < 0.0001$); Ramp $z = 1.96$ ($p < 0.05$). Several explanations for this result are possible. It could be that by the time Circuit subjects are asked the MC, they have been given enough information to understand the task and can now answer correctly. This does not explain why there are age differences for Free but not for MC. If the problem were one of comprehension, one would expect age differences for both questions. An alternative explanation is that the MC format organized the information available for solving the Circuit or Ramp so that the task in both cases had the same requirements: the subject only needed to reason about information that had been organized for him already rather than first organizing the information and then reasoning about it.

Free and Screen. Comparing success on Screen to success on Free (Fig. 3), Free is more difficult for Circuit than for Ramp, but Screen is equally difficult. As we noted when comparing MC and Free, the way the information in the question is organized appears to affect the success rate. Both MC and Screen ask the subject to consider more specific information than Free. The MC emphasized the information necessary to set up a controlled experiment so it was easier than Free. The Screen emphasized results; this information is not relevant to setting up a controlled experiment. Thus Screen was as difficult as Free for 12- and 14-year olds but easier for 16-year olds.

Clearly some change in ability to organize or process information on the Screen question takes place between 14 and 16. It appears that the young child, forced to pay attention to results, apparently does not attempt to process information about the procedure used to produce the results, whereas the older child does consider the procedure. Even when the experiment done behind the screen is repeated without the screen, 25% of the children still choose with certainty the marble that wins or the wire that the experimenter mentions when the box buzzes.

In summary, when the subject was presented with the Free question, which required him to develop a plan of attack, it appeared that the familiarity of the task was a variable. But when the subject was presented with MC or Screen, which organize the information either in a relevant or irrelevant manner, then familiarity with the task did not affect success rate. For MC and Screen, the organization of the information affected the success rate.

Correlations between Questions

Further evidence for the idea that MC is a simplification of Free, and that 12- and 14-year olds handle Screen differently from 16-year olds is gained by looking at the correlations between the three questions shown in Table II.

Free and MC. Although Free is harder than MC, scores are significantly correlated

for both tasks: Ramp (Free, MC) $z = 0.35$ ($p < 0.01$); Circuit (Free, MC) $z = 0.35$ ($p < 0.01$). It appears that some of the same abilities are necessary for performance on both Free and MC.

Free and MC versus Screen. Whereas scores on Screen show a dramatic increase with age, scores on Free and MC increase slightly, or not at all, with age. When we look at the relationship between scores on Screen with scores on Free or MC, we see that they are independent for 12- and 14-year olds, but are related for 16-year olds as shown by the correlations in Table II.

Furthermore, for 12-year olds, the number of subjects who fail the Screen but pass the MC far exceeds the number who pass Screen but fail MC. Fourteen- and 16-year olds are just as likely to pass MC and fail Screen as to fail MC and pass Screen. Using McNemar's test for the significance of changes: for 12-year olds, Circuit $\chi^2 = 4.08$ ($p < 0.05$), Ramp $\chi^2 = 6.13$ ($p < 0.02$); for 14- and 16-year olds, there are no significant differences.

In summary, familiarity with the variables is associated with success on Free, but three major results are inconsistent with the hypothesis that familiarity is a major determinant of success. First, the familiarity hypothesis would predict that the same problem would be difficult for MC and Screen as for Free. Instead, the Circuit was most difficult for

TABLE II
Relationship between Free, Multiple Choice, and Screen by Age Using Kendall's τ

		Circuit							
		Multiple Choice				Screen			
Age		12	14	16	All	12	14	16	All
Free		.25	.41**	.26	.35**	.00	.12	.52	.32**
Multiple Choice						.02	.10	.31	.22*
		Ramp							
		Multiple Choice				Screen			
Age		12	14	16	All	12	14	16	All
Free		.36*	.25	.23	.35**	-.10	.19	.38*	.23**
Multiple Choice						-.16	.11	.29	.20*

* $p < 0.05$.
** $p < 0.01$.

Free, while Circuit was slightly easier for MC and Screen. Second, the familiarity hypothesis would predict equal relationships across ages for each question. Instead, success on Screen was strongly related to age, while success on Free and MC was not. Third, the familiarity hypothesis would predict equal relationships across questions. Instead, Screen was independent of Free and MC for 12- and 14-year olds, but was related for 16-year olds. These results suggest that if familiarity is important, it is not the only important factor in this sort of reasoning. Possible explanations range from independent developmental curves for each of the three questions with a limited role for familiarity, to a catholic information processing model.

The results of this study suggest that concrete operations are not a unitary trait. Performance on Screen provides the strong relationship to age that one would expect for a change from concrete to formal thought; performance on MC and Free is not so strongly related to age. The correlations in Table II offer further evidence that Screen and Free or MC are assessing different types of performance in 12- and 14-year olds. For the 16-year olds only, results do suggest a single factor to explain performance on the two tasks. In general, the results of these experiments suggest a change in method of processing information between 12 and 16.

Information processing theorists have at various times suggested ideas about inhibiting salience [7], or increases in computing space [3,8]. One could explain this change by saying that older children had more computing space or inhibited the seemingly salient information in order to process the necessary information and therefore could organize the information in Screen and would be more efficient at Free than younger children.

The role of a concrete presentation for a task clearly depends on the nature of the presentation. Concrete presentations that emphasize results, like Screen, could make it less likely that a subject would use a logical skill. Questions presented completely verbally, like Seeds, are not necessarily more difficult than those presented with many concrete objects.

The findings of this study suggest that programs which aim to teach logical thought will be most successful if they emphasize the recognition and organization of relevant information rather than if they simply emphasize the "all other things equal" schema. Recent studies [9,10] offer support for these findings.

The authors would like to thank Dr. A. R. Jonckheere for providing encouragement and intellectual stimulation for all aspects of this study and A. R. Samaroff and Murry Levine for comments on an earlier draft of the manuscript. Special thanks are due to Miss Muriel Carter, the students, and the staff of Southgate School for their cooperation. This research was carried out while one of the authors was a visitor to the Department of Psychology, University College London, and the other author was a University of Michigan Exchange Fellow at University College London. This material is based upon research supported by the National Science Foundation under grant No. SED74-18950. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation. Partial results of this study were presented at the Jean Piaget Society Meeting, May, 1977.

References

1. Inhelder, B. and J. Piaget, *The Growth of Logical Thinking: From Childhood to Adolescence*, A. Parsons and S. Milgram (Trans.), New York: Basic Books, 1958.
2. Levine, D. I. and M. C. Linn, "Scientific reasoning ability in adolescence: Theoretical

viewpoints and educational implications," *Journal of Research in Science Teaching*, 14, (4): 371-384 (1977).

3. Pascual-Leone, J., "A mathematical model for the transition rule in Piaget's developmental stages," *Acta Psychologica*, 63: 301-345 (1970).

4. Wollman, W. and A. E. Lawson, "Teaching the procedure of controlled experimentation: A Piagetian approach," *Science Education*, 61(1): 57-70 (1977).

5. Wollman, W., "Controlling variables: A neo-Piagetian sequence," *Science Education*, 61(3): 385-391 (1977).

6. Linn, M. C., "Scientific reasoning: Influences on task performance and response categorization," *Science Education*, 61 (3): 357-369 (1977).

7. Hunt, J. McV., *Intelligence and Experience*, New York: Ronald, 1961.

8. McLaughlin, G. H., "Psycho-logic: A possible alternative to Piaget's formulation," *British Journal of Educational Psychology*, 33: 61-67 (1963).

9. Linn, M. C., B. Chen, and H. D. Thier, "Personalization in science: Investigations at the middle school level," *Instructional Science* (1976).

10. Linn, M. C. and H. D. Thier, "The effect of experiential science on the development of logical thinking in children," *Journal of Research in Science Teaching*, 12: 49-62 (1975).

11. Bryant, P. E., *Perception and Understanding in Young Children: An Experimental Approach*, London: Methuen, 1974.

Received January 10, 1977

Revised February 2, 1978