Developmental Psychology
1975, Vol. 11, No. 6, 837–842

Effects of Age and Ability on Syllogistic Reasoning in Early Adolescence

DANIEL P. KEATING
University of Minnesota

ALFONSO CARAMAZZA
The Johns Hopkins University

To assess the influence of age and ability on linear syllogistic reasoning in early adolescence, 11- and 13-year-old subjects of two levels of ability (bright and average) were presented with 64 three-term series problems (e.g., "If John is better than Bill, and Bill is better than Tom, then who is the best?"). Results showed that the effect of ability was quite dramatic even with verbal grade equivalent and Standard Progressive Matrices scores as covariates, whereas the effect of age was marginally significant and confounded with both. Comparison to Clark's 1969 adult sample showed that the pattern of errors was essentially similar for adolescents and adults, but more similar for the average than the bright subjects. Implications of the findings for Clark's theory and for a theory of intellectual precocity are discussed.

A subject's success in correctly solving a problem depends on many factors: Some reflect momentary fluctuations in attention due to set and motivation; others reflect such long-term factors as the subject's intellectual ability, developmental level, and specific experiential history with the problem at hand. Our concern in this article is to investigate the effects of age and certain intellectual skills on the subject's syllogistic reasoning ability. In particular, we are concerned with the effects of subjects' reasoning skills at ages that correspond roughly to preformal and formal operational thought (Inhelder & Piaget, 1958) on their ability to solve problems that, at least superficially, require only linguistic processing (Clark, 1969b).

The specific problem types used in the present experiment were three-term series problems which consist of two comparative sentences as premises and a question (e.g., "If John is taller than Bill, and Tom is taller than John, then who is the tallest?"). Previous investigators have studied how children and teenagers (Donaldson, 1963; Hunter, 1957; Piaget, 1923) and adults (Clark, 1969a, 1969b; DeSoto, London, & Handel, 1965; Huttenlocher, 1968) solve these problems.

One important aspect neglected in the more recent studies, however, is the question of individual differences. It is our contention that an analysis of the relative performance and patterns of performance for problem types of individuals of different levels of ability would contribute considerably to our understanding of the processes involved in syllogistic reasoning. That is, a comparison of differential facilitation for problem types as a function of intellectual skills can reveal component processes in the solution of three-term series problems. Thus, in this article we are also concerned with the nature of the processes normally involved in solving linear syllogisms.

Such an investigation contributes also to a better understanding of specific problem-solving abilities that many observers agree develop rather rapidly during the period of early adolescence. By extending the study of the development of cognitive skills beyond a concentration on individual differences in power (the psychometric approach) or of universal cognitive structures (Inhelder & Piaget, 1958), we hope to gain insight into the kinds of cognitive processes that are involved in intellectual development.

This research was partially supported by Grant GS-39775 from the National Science Foundation to the first author. The authors thank Julian C. Stanley for helpful comments on an earlier draft.

Requests for reprints should be sent to Daniel P. Keating, Institute of Child Development, University of Minnesota, Minneapolis, Minnesota 55455.

Method

Subjects

The subjects for this study were 109 male fifth and seventh graders selected in the following manner. From a list of all the fifth- and seventh-grade students in the public schools of a suburban county who scored either at the 98th or 99th percentile of that county's norms on the composite arithmetic scale of the Iowa Test of Basic Skills or between the 45th and 55th percentiles on the same test, 50 male students were randomly selected to form each of these four groups: seventh-grade bright (7B); seventh-grade average (7A); fifth-grade bright (5B); and fifth-grade average (5A). Verbal scores were not used for selection. These 200 students were invited to participate in a study of "reasoning and learning," and were offered a small fee for their participation.

Of the 200 students invited, these 109 came to the test: 31 from the 7B group; 19 from 7A; 37 from 5B; and 22 from 5A. The students were mostly middle-class and white. Less than 5% were from minority groups, and they were not overrepresented in either of the levels of the age and brightness factors.

The ages of the bright and average groups within each grade were not significantly different from each other. The fifth graders' mean age was 11 years 3 months with a standard deviation of 5.5 months, and the seventh graders' mean age was 13 years 1 month with a 4.3-month standard deviation. For further details, see Keating (1975).

Procedure

The subjects were presented with 64 three-term series problems of the kind described by Clark (1969a). There were 2 problems for each of the 32 different formats of the three-term series problems: regular versus negative equatives; best versus worst; marked versus unmarked; mixed versus unmixed; and order of mixed and unmixed. An example of such a problem is, "If Bill is better than Joe, and Joe is better than Al, then who is best, Bill, Al, or Joe?" Table 1 lists the 32 different kinds of problems.

The 16 names used were selected from Battig and Montague's (1969) longer list of the most common male names. The names used in each problem were randomly selected from this shorter list of 16, as were their order in the problem and in the choices.

The 64 problems were presented in random order using a slide projector and screen, with the subjects being tested in two large groups. Assignment to session was random, and conditions were the same for both groups of subjects. The problem was flashed on the screen and left there for 10 sec while the experimenter read the problem aloud. After the problem was removed from the screen, the subjects had an additional 3 sec to mark

Table 1: Types of Problems Used in the Experiment

Problem type	Form of problem	Form of question
1	If A is better than B, and B is better than C,	then who is best
	If A is better than B, and B is better than C,	then who is worst
	If B is better than C, and A is better than B,	then who is best
	If B is better than C, and A is better than B,	then who is worst
2	If C is worse than B, and B is worse than A,	then who is worst
	If C is worse than B, and B is worse than A,	then who is best
	If B is worse than A, and C is worse than B,	then who is worst
	If B is worse than A, and C is worse than B,	then who is best
3	If A is better than B, and C is worse than B,	then who is best
	If A is better than B, and C is worse than B,	then who is worst
	If C is worse than B, and A is better than B,	then who is best
	If C is worse than B, and A is better than B,	then who is worst
4	If B is worse than A, and B is better than C,	then who is best
	If B is worse than A, and B is better than C,	then who is worst
	If B is better than C, and B is worse than A,	then who is best
	If B is better than C, and B is worse than A,	then who is worst
1′	If A is not as bad as B, and B is not as bad as C,	then who is best
	If A is not as bad as B, and B is not as bad as C,	then who is worst
	If B is not as bad as C, and A is not as bad as B,	then who is best
	If B is not as bad as C, and A is not as bad as B,	then who is worst
2′	If C is not as good as B, and B is not as good as A,	then who is best
	If C is not as good as B, and B is not as good as A,	then who is worst
	If B is not as good as A, and C is not as good as B,	then who is best
	If B is not as good as A, and C is not as good as B,	then who is worst
3′	If A is not as bad as B, and C is not as good as B,	then who is best
	If A is not as bad as B, and C is not as good as B,	then who is worst
	If C is not as good as B, and A is not as bad as B,	then who is best
	If C is not as good as B, and A is not as bad as B,	then who is worst
4′	If B is not as good as A, and B is not as bad as C,	then who is best
	If B is not as good as A, and B is not as bad as C,	then who is worst
	If B is not as bad as C, and B is not as good as A,	then who is best
	If B is not as bad as C, and B is not as good as A,	then who is worst

their answers, which were presented in multiple-choice format. The subject had to circle the correct name from among the three choices. The subjects were given "do guess" instructions, that is, they were required to indicate an answer for each problem whether or not they were sure of their selection. Before the actual experimental material was presented, subjects were given 8 practice problems to work on. All subjects understood the instructions and could perform the task.

During the same session, the subjects also took Raven's (1960) Standard Progressive Matrices as a better measure of nonverbal reasoning ability.

Results

Pattern of Percentage Errors

The first question was whether there were obvious differences in the way early adolescents solved these problems when compared to an adult sample. In Table 2, the percentage errors for each of the eight major types of problems (see Table 1) made by each of the four groups are shown and compared with the same data from Clark's (1969b) sample of college students. The values in Table 2 reveal that the relative difficulty of the different problem types was quite similar for all the groups, even though the differences between groups were of considerable magnitude, as described further below.

To study the similarity to the adult sample more carefully, the percentages of errors for each of the four groups on all 32 problems were correlated with the same data from Clark's (1969b) sample. These correlations are shown in Table 3, separately for "regular" problems (the 16 problems under Types 1–4 in Table 1) and for negative equatives (Types 1′–4′), as well as for all problems combined. Three things are readily apparent from Table 3. First, the similarity in the pattern of errors between an adolescent and an adult sample was quite high.

Table 2: Percent Error on Eight Problem Types for Four Adolescent Groups and Clark's (1969a) Adult Sample

Problem type	Adults	5A	5B	7A	7B
1	17	28	7	28	6
2	33	43	16	35	9
3	10	18	5	20	2
4	32	50	12	32	5
1′	31	42	21	41	17
2′	26	39	13	28	11
3′	40	40	15	32	9
4′	28	39	10	21	4

Note. 5A = fifth-grade average, 5B = fifth-grade bright, 7A = seventh-grade average, and 7B = seventh-grade bright.

Table 3: Correlations of Percent Errors for the Three-Term Series Problems of Four Adolescent Groups with Clark's (1969a) Adult Sample[a] Problem Types

Group[b]	Regular	Negative equative	Overall
5A	.87**(.93**)	.44 (.49*)	.78**
5B	.70**(.73**)	.32 (.39)	.60**
7A	.75**(.81**)	.58**(.64**)	.65**
7B	.63**(.72**)	.41 (.46)	.52**

[a] Figures in parentheses are correlations corrected for measurement unreliability of the adolescent groups.
[b] 5A = fifth-grade average, 5B = fifth-grade bright, 7A = seventh-grade average, and 7B = seventh-grade bright.
* $p < .05$.
** $p < .01$.

Second, the similarity was more striking for regular three-term series problems than for negative equatives. This latter result suggests that the strategies used by early adolescents in solving negative equatives are overall quite different from those used by adults and that for these problem types there is still considerable room for development. Third, the correlations between the average subjects in our sample and Clark's groups were consistently higher than that between our bright subjects and Clark's groups. These differences were not significant but are in the opposite direction from the results that would be expected from a model that suggests that the bright groups are essentially closer to obtaining the structures or employing the strategies used by the adult group. That is, these results contradict the simple model that would have precocity interpreted merely as earlier acquisition of the structures or strategies of the older groups.

These differences in pattern of errors could be due to differences in measurement reliability of the three-term series problems. Split-half reliability coefficients (with random assignment of paired items each of the 32 formats in Table 1 to each half) were calculated for each age–ability group separately for regular and negative equative problems. These coefficients ranged from .66 to .92, with a median figure of .82. The median for the regular problem types across all groups is .88, and .79 for the negative equatives. For the bright groups (5B and 7B) the median reliability coefficient is .77, and .88 for the average groups (5A and 7A). The correlations corrected for unreliability of measurement for these groups are shown in parentheses in Table 3. The pattern differences noted above remain in the corrected correlations, and it is thus unlikely that they are primarily attributable to greater unreliability of measurement of the negative equatives or of the bright group.

Age and Ability Differences

One of the initial purposes of this investigation was to examine whether there is growth in the ability to do this type of syllogistic reasoning during the age period under consideration, since this type of reasoning has been linked to formal operational structures (Piaget, 1923) and these ages occur in the transitional period from concrete to formal operations. Another purpose was to examine the effect of levels of ability on these syllogistic reasoning problems.

Univariate analyses of variance (ANOVAs) for each of the eight problem types were carried out using the two levels of age (fifth grade, 11 years old; seventh grade, 13 years old) and two levels of ability (average and bright). Additionally, a multivariate analysis of variance (MANOVA) with the different problem types as criteria was also performed for the same levels of age and ability. The results of these analyses are shown in Table 4, where the *F*s and significance levels are listed for the eight univariate ANOVAs and the MANOVA.

The pattern of results is quite striking. Clearly, ability was the major factor in accounting for differences in individual performance on syllogistic reasoning problems. Although there were only two problem types where age was a significant main effect, the MANOVA *F* for age was significant at the $p < .05$ level. Thus, there did appear to be some development due to age independent of level of psychometric ability in the solution of the three-term series problems. None of the Age × Ability interactions were significant.

These results are depicted in Figures 1 and 2 for further clarification. It is clear from Figure 2 that the MANOVA *F* for age was significant, while only two of the problem types showed a significant main effect because of the older groups' consistent superiority within ability levels.

The differences between the bright and average groups are presumably present also in ability components other than the selection variable (arithmetic composite). Scores

Table 4: ANOVA and MANOVA *F*s for Eight Problem Types

Problem type	Grade (G)	Ability (A)	G × A
ANOVA			
1	.073	47.468***	.131
2	3.751	54.021***	.007
3	.045	35.625***	1.125
4	9.074**	78.591***	2.556
1′	.360	29.496***	.133
2′	1.123	31.021***	.506
3′	2.362	39.257***	.108
4′	8.886**	49.098***	3.743
MANOVA	2.118*	11.630***	1.547

* $p < .05$.
** $p < .01$.
*** $p < .001$.

of the same subjects on Raven's Standard Progressive Matrices were used as covariates in a reanalysis of the data, as were their vocabulary grade equivalents and age in months. The pattern of multivariate analysis of covariance (MANCOVA) *F*s was the same as the MANOVA Fs across all combinations of covariates, with the exception of grade (age). With Standard Progressive Matrices and vocabulary grade equivalents covaried, the grade effect was no longer significant ($p > .06$), but the ability effect was still highly significant ($p < .001$). This should not be overinterpreted because of the small differences *between* MANCOVA and MANOVA *F*s for grade. Square roots of percentage errors also yielded similar results. This provides further confirmation of the dominant ability effect and the less dramatic age effect.

The correlations of total score on the 64

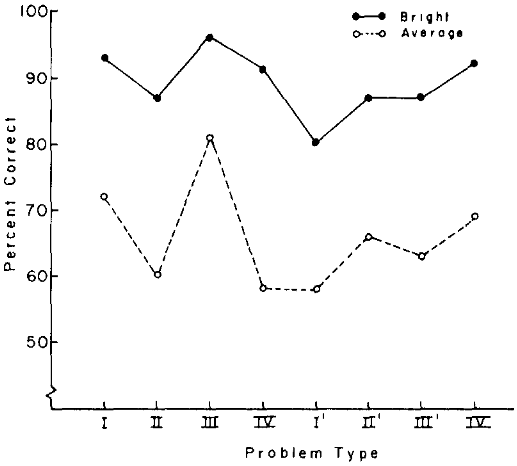

Figure 1. Percent correct responses by problem type for bright and average students.

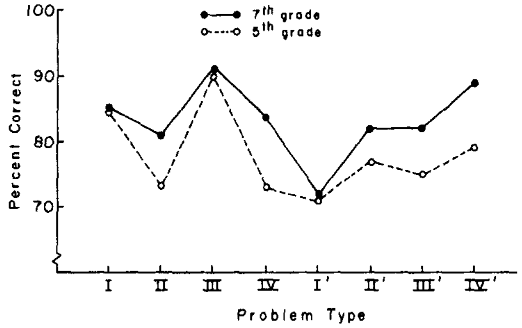

Figure 2. Percent correct responses by problem type for fifth and seventh graders.

three-term series problems with the Standard Progressive Matrices and vocabulary grade equivalents were .20 ($p < .05$) and .32 ($p < .01$), respectively. These were not significantly different from each other, $t(106) < 1$, $p > .10$, suggesting that both represent ability domains that contribute to facility in solving syllogistic reasoning problems. (The correlation between them is .11, $p > .10$). Unless the contribution of the Standard Progressive Matrices is entirely due to a close relationship with language ability, this suggests that the solution of such problems is not solely a linguistic process and that abstract nonverbal reasoning may play a significant role.

Type of Errors

Piaget (1923) has suggested that younger children misconstrue the nature of this problem as classificatory rather than serial, and thus specific types of errors are made by them. For example, given the problem "John is better than Bill. John is worse than Tom. Who is the worst?", Piaget has suggested that preformal operational children interpret the comparative sentences as strict predications of the nouns. That is, in "John is better than Bill," both John and Bill are *good*, and, in "John is worse than Tom," both John and Tom are *bad*. Thus when the child is asked to say who is worst he will base his choice on the specific value predicated of the noun. By such reasoning in the present example, Bill is best since it has been predicated that he is good. John is intermediate because he is both good and bad. Tom is the worst since he has been predicated only the value bad. In this way specific predictions can be made about type of errors the child is likely to make if he interprets the comparative constructions as strict cases of predication.

Problem Type 4 (see Table 1) is especially susceptible to error using this incorrect strategy, and therefore the type of errors made on this problem were analyzed for the 7A and 5A groups. (The error rate for groups 7B and 5B were not analyzed because of the paucity of errors in those goups.) The results of this analysis are shown in Table 5. The younger average students made fewer classification errors than "other" errors, whereas the older average students made more classification errors. This suggests that development of performance in this type of syllogistic reasoning was not solely due to a substitution of a "better" strategy for the incorrect classification strategy. We should point out that to the extent that our results do not support Piaget's (1923) claim, they also cast doubt on one of Clark's (1969b) arguments for his "linguistic" theory of solving three-term series problems, since Clark has construed the results reported by Piaget as support for his own theory.

On the other hand, it is noteworthy that the older average students' ratio of classification to other errors, 22% to 12%, is more similar to Clark's (1969a) adults' ratio, 6% to 3%, than is the younger average students' ratio, 21% to 29%. One might conclude that the "hold" on adolescents' thinking of the classification strategy is greater than that of other incorrect strategies. But it is also noteworthy that the absolute proportion of classification errors remains stable during a period of rapid development of formal operational thinking (Keating, 1975). This suggests that a simple linear model of intellectual precocity will not suffice.

Discussion

It is interesting to note that these findings accord well with a related study of ability, age, and performance on Piagetian tasks of formal operational reasoning (Keating, 1975). In that investigation, using a smaller sample from these same groups, the age difference was significant but overshadowed by the differences associated with ability. Thus it seems that in both syllogistic reasoning and formal operational reasoning, the power differences attributable to level of ability (psychometrically assessed) is of considerable magnitude, while the differences at-

Table 5: Percentage of Types of Errors for Average Fifth-Grade (5A) and Seventh-Grade (7A) Subjects on Problem Type 4

Error type	5A	7A
Correct solution	50	66
Classification error	21	22
Other error	29	12

tributable to age even in this presumably transitional period are significant but of much less magnitude.

As noted earlier, one of the chief advantages of using the three-term series problem is its sensitivity in revealing different cognitive processes (e.g., different solution strategies) that may be associated with age or ability differences. In this investigation the evidence regarding process differences though not totally compelling is quite interesting. There are several issues which ought to be stressed. First, as pointed out earlier, it seems that brighter early adolescents are not using strategies totally like either the other adolescents we tested or the adult group tested by Clark. This contradicts the simplistic approach to precocity which would treat it as merely the earlier acquisition of the cognitive strategies that characterize the adult. A more complicated model would seem to be required to account for our findings.

The greater similarity of our average groups and Clark's groups may be attributable to the characteristics of the latter sample: It may have been more comparable in characteristics salient to performance on these tasks to a 50th-percentile-or-above ability group than to a 98th-percentile group.[1] If a 98th-percentile-or-above ability group of college age completed the same set of tasks, and if the correlation of this group with our bright groups were high, then this would constitute strong evidence for a process difference in the solution of syllogistic reasoning tasks related to level of ability but not to age. That is, bright adolescents may be more similar to bright than to average adults, which is somewhat different from a straightforward mental age model. We should also point out that possible differences between the bright and average adolescents could be masked by the dominant power difference.

A further comment regarding the processes involved in the solution of such problems is in order. Clark (1969a, 1969b) argued that the solution of these problems is primarily but not exclusively dependent on linguistic reasoning processes. The evidence from this investigation, including the contributing factor of nonverbal ability and the absence of clear evidence for the developmentally decreasing dominance of classification errors, suggests that something more than linguistic processes is involved.

[1] The average Scholastic Aptitude Test—Mathematical score for college students at or above the 98th percentile would be approximately 775, as compared to a 50th-percentile score of slightly less than 500. Clark's group almost certainly fell between these two figures, and assumptions of "greater" similarity are necessarily speculative.

REFERENCES

Battig, W. F., & Montague, W. C. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs*, 1969, *80*(3, Pt. 2).

Clark, H. H. Influence of language on solving three-term series problems, *Journal of Experimental Psychology*, 1969, *82*, 205–215. (a)

Clark, H. H. Linguistic processes in deductive reasoning. *Psychological Review*, 1969, *76*, 387–404. (b)

DeSoto, C., London, M., & Handel, S. Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology*, 1965, *2*, 513–521.

Donaldson, M. *A study of children's thinking*. London: Tavistock, 1963.

Hunter, I. M. L. The solving of three-term series problems. *British Journal of Psychology*, 1957, *48*, 286–298.

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

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

Keating, D. P. Precocious cognitive development at the level of formal operations. *Child Development*, 1975, *46*, 276–280.

Piaget, J. Une forme verbale de la comparaison chez l'enfant. *Archives de Psychologie*, 1923, *18*, 141–172.

Raven, J. C. *Guide to the Standard Progressive Matrices*. London: Lewis, 1960.

(Received May 8, 1975)