

Development of inferences over elementary-school grades: III. Verbatim and forward-consequence inferential errors made by regular and gifted students

MELVIN H. MARX

Western Carolina University, Cullowhee, North Carolina
and Florida Institute of Technology, Melbourne, Florida

Students in second, fourth, and sixth grade regular and gifted classes were read a series of short stories by their teachers. They were then asked to identify two statements about each story as true or false. Half of the statements were factual (verbatim); of these, half were true, half false. The other half of the test statements were all false but were arranged so that if a forward-consequence inference implicit in the story were made they would be true. Susceptibility to inferential bias in memory showed no improvement over grades for the regular students and improved only for the sixth-grade gifted students. These results contrast sharply with those for verbatim items, on which the usual improvement occurred over grades.

This experiment is the third in a series of exploratory studies conducted to examine the course of development of inferences over the elementary-school grades. The first two consisted of comparisons of memory for implicit and explicit words, the former presumably requiring inferences for retrieval (Marx, 1991, 1992).

The present experiment was designed to measure changes over the second through sixth grades in susceptibility to bias from inferences in memory in normal and gifted children; one class of each type of student was tested in the second, fourth, and sixth grades. Short stories were read to the students by their teachers, and then verbatim and inferential test items were presented, which were to be marked "true" or "false." Marking inferential statements "true" was assumed to indicate the acceptance of a plausible but not logically required inference. The inferences thus suggested were all of a type that has been labeled *forward-consequence* (Singer, 1988; Thorndyke, 1976). This type of inference implies that some particular event has occurred, although it has not specifically been described. The error score on inferential items (calling them "true") was thus a measure of the tendency to generate such inferences.

Thanks are due Ben Franklin, Assistant Superintendent, St. Lucie County Florida Board of Education; Lois Watts, Principal, Debby Ruscum, Coordinator, and the participating teachers, Lawnwood Elementary School, Fort Pierce, FL, for their cooperation in the collection of the data; and Yung Che Kim of Keimyung University, Daegu, South Korea, for providing the analysis of variance. Address correspondence to M. H. Marx, Department of Psychology, Western Carolina University, Cullowhee, NC 28723.

METHOD

Subjects

A total of 120 students in regular and gifted classes of the second, fourth, and sixth grades in a public school served as subjects. There were eight invalidated answer sheets (no answers, all "true" answers, etc.). The *ns* and age means (with *SDs*) for the six classes were: second grade, regular, *n* = 22 (11 female, 11 male), 8.2 years (1.68); second grade, gifted, *n* = 23 (10 female, 13 male), 7.7 years (.86); fourth grade, regular, *n* = 26 (13 female, 13 male), 10.6 years (.77); fourth grade, gifted, *n* = 17 (10 female, 7 male), 10 years (.75); sixth grade, regular, *n* = 14 (7 female, 7 male), 12.4 years (.82); and sixth grade, gifted, *n* = 18 (11 female, 7 male), 13.2 years (1.60).

Materials

There were 10 short (three-sentence) stories, the first and last of which were included as primacy and recency buffers, and at test there were two statements, one verbatim and the other inferential, for each of the remaining 8 stories, with "true" or "false" answers requested.

The following example illustrates the stories and test items used. *Story*: "A lioness waked up from a long sleep and was very hungry. She went hunting and found a small herd of antelopes. They were feeding in very tall grass." *Verbatim true* test item: "The lioness waked up hungry and went hunting." *Verbatim false*: "The lioness waked up her sleeping mate to go hunting." *Inferential simple*: "The lioness killed and ate an antelope." *Verbatim complex*: "The lioness killed and ate an antelope after she found a small herd feeding in very tall grass."

Design

The experimental design was a straightforward analysis of variance (ANOVA), with school grade and student intelligence rating as the main factors. There were two types of verbatim test items, true and false, and two types of inferential test items, a simple statement requiring the use of an inference to be marked "true" and a more complex statement incorporating the same inference but embroidered with additional, accurate details from the story.

Procedure

The students were tested by their teachers in the normal classroom setting. They were told that they would be read some very short stories

and that they should "try to remember them because when I finish reading them I will ask you some questions about them." There were 4- to 5-sec pauses between stories and a 1-min rest interval before testing, during which the teacher passed out the answer sheets.

For the test, the teacher said that some statements would be read about the stories:

Some of these are true, some are false. Listen carefully to each statement. If you think it is true—it was that way in the story—circle the T after the same number on the answer sheet as was called out for that statement. If you think it is false—it was not that way in the story—circle the F after the number.

Two practice statements, both verbatim, were then read, one taken from each of the two buffer stories. One statement was clearly true, the other false, and the answers were discussed with the class. No practice on inferential items was given. The two test items were then read for each of the remaining eight stories, with the story related to each set of two statements identified by its title (e.g., "The Lioness").

RESULTS

Inferential Items

The major results are plotted in Figure 1. The most interesting result comes from the comparison of the regular and gifted students with respect to their susceptibility to inferential bias—that is, their tendency to make the implicit inferences, as measured by marking such items "true" and thereby indicating that they thought the statement was in accord ("that way," in the instructions) with the story as it was read. That the making of implied forward-consequence inferences was persistent over grades is indicated by the essentially flat error curve for the regular students, whereas the gifted students did show a marked improvement, albeit only in the sixth grade. The ANOVA revealed no effect of grade [$F(2, 102) = 1.412, p > .05$], but there was a marginal effect of intelligence

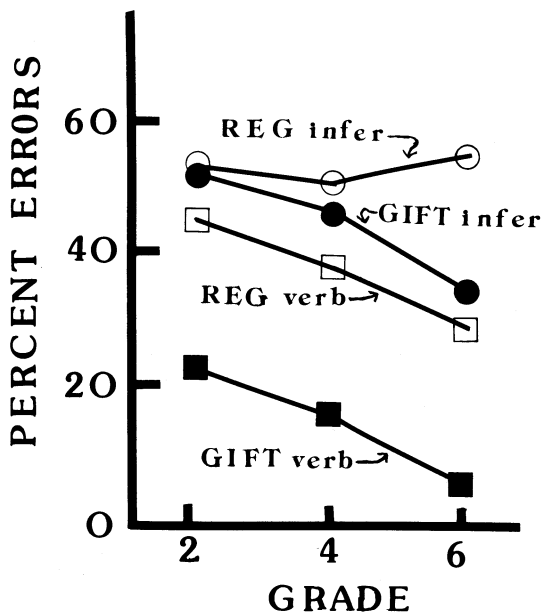


Figure 1. Mean percentages of errors on verbatim and inferential items made by regular and gifted students in the second, fourth, and sixth grades.

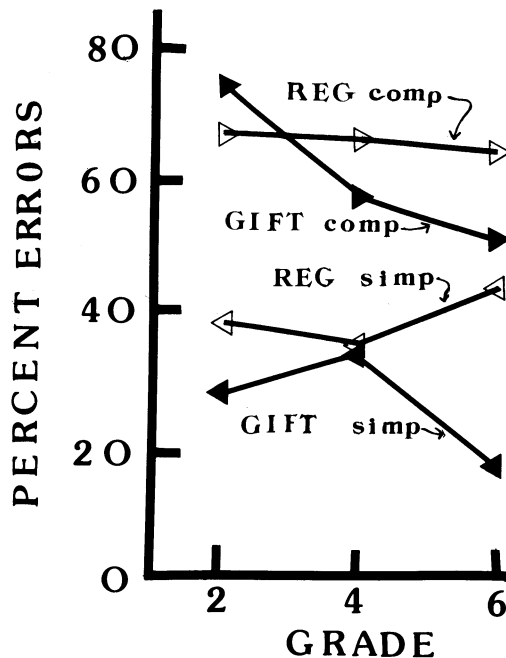


Figure 2. Mean percentages of errors on simple and complex inferential items made by regular and gifted students in the second, fourth, and sixth grades.

[$F(1, 102) = 3.910, p = .05$]; no interaction occurred. Also, chi-square analysis of the proportion of students in each sixth-grade class who made more inferential errors than the mean score for their grade suggested the reliability of the difference between the regular and the gifted proportions [11/14 for regular and 5/18 for gifted sixth graders; $\chi^2(1) = 8.14, p < .01$].

Verbatim Items

The results plotted in Figure 1 for verbatim items showed the expected improvement over grades, for both types of student, as well as the anticipated superior performance of the gifted students. The statistical reliability of these effects is indicated by the ANOVA [for grade, $F(2, 102) = 7.540, p < .01$, and for intelligence, $F(1, 102) = 45.559, p < .01$]. There were no reliable interactions.

Simple Versus Complex Inferential Items

Figure 2 shows the mean error rates for the two types of inferential items. Embellishing this kind of item by adding veridical detail apparently had the anticipated facilitatory effect on error rate for both regular and gifted students. The ANOVA showed no differences for the grade variable ($F < 1.00$) for both simple and complex items. The gifted students were reliably superior on simple items [$F(1, 102) = 6.383, p < .01$], but not on complex items ($F < 1.00$). Apparently the embellishment of the inferential items made them more attractive, and thus more susceptible to inferential bias, for gifted as well as for regular students.

The mean/overall differences between error rates for simple and complex items were reliable by chi-square test.

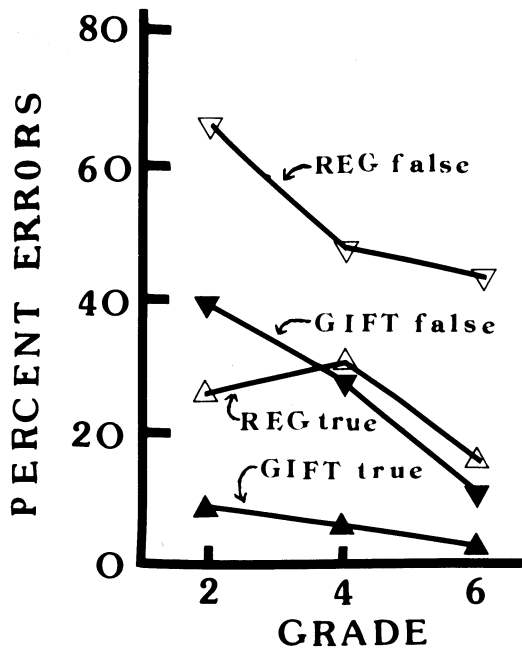


Figure 3. Mean percentages of errors on true and false verbatim items made by regular and gifted students in the second, fourth, and sixth grades.

The proportion of regular students who made more errors on complex than on simple inferential items, with ties disregarded, was 37/45 (82%), reliably greater than a 50-50, or chance, expectation [$\chi^2(1) = 18.69, p < .01$]. The comparable figures for the gifted students were very similar: 36/39 (92%) [$\chi^2(1) = 27.92, p < .01$].

True Versus False Verbatim Items

Figure 3 shows the error rates for true and false verbatim items. It is clear that calling such false items "true" was much more common than calling true items "false." An ANOVA revealed no grade effect for the verbatim true items [$F(2, 102) = 2.705, p > .05$], but there was a reliable grade effect for verbatim false items [$F(2, 102) = 8.885, p < .01$]. The intelligence variable was also reliable [$F(1, 102) = 30.891$ for true items and $F(1, 102) = 27.615$ for false items, both $ps < .01$].

Overall, the difference in mean error rate between true and false verbatim items indicated in Figure 3 was statistically confirmed. The proportion of regular students who made more errors on false than on true items (with ties disregarded) was 43/50 (86%); compared with a 50-50 baseline, this proportion was reliable [$\chi^2(1) = 25.92, p < .01$]. The comparable proportion for the gifted students was 35/37 (95%) [$\chi^2(1) = 29.43, p < .01$].

DISCUSSION

The most provocative result in this experiment is the contrast between the error rates for inferential and verbatim items; the predictably steady improvement over grades shown for verbatim items was not matched by inferential items. Why even gifted students showed little improvement on inferential items until the sixth grade raises an especially in-

teresting question. Apparently superior intellectual status, which produced a very great improvement in verbatim tests, was not by itself sufficient, up to the sixth grade, to stem susceptibility to inferential bias. How this kind of susceptibility can best be checked, assuming that such checking is part of an instructional objective, remains to be seen. Specific instruction concerning and practice on identification of certain kinds of inference seems, in general, to offer the most promising line of attack on this problem.

The teachers in this experiment were asked if they had devoted any class time to a discussion of inferences and their necessary role in understanding discourse. None reported such discussion. This issue is important, because of the stress on inferences in some language texts. It is obvious that instructional set, including self-instructions, plays a primary role in how subjects treat inferences of the kind examined in the present research.

The present results provide a picture somewhat different from that in earlier research done with mainly younger children who were questioned about the adequacy of ambiguous information. Beal (1990) has reported that third graders were more capable of separating inferential from factual materials than younger children were. In another study (Beal & Belgrad, 1990), there was a regularly decreasing tendency to over-emphasize the adequacy of ambiguous picture cues from preschool to second grade. It would be interesting to see whether a similar improvement in evaluation of the ambiguous information with which students were provided in the present experiment would result in a similar improvement over grades, in spite of the failure of the older children to answer direct informational questions more accurately (except of course for the gifted sixth graders).

It should be mentioned that when the inferences in the present experiment were made—that is, during comprehension or at retrieval test—cannot be determined by the testing procedure. This problem is crucial for investigators who are primarily concerned with interpreting how discourse is understood (cf. Keenan, Potts, Golding, & Jennings, 1990; Singer, 1988), but it is less important when the focus is on development. That is to say, the primary concern in the present study was the development of the generation of inferences and their biasing effect on memory: it is left to future research to examine the locus of the inference generation, with the use of on-line tests.

Secondary interest attaches to the other variables in the experiment. The anticipated facilitation of errors resulting from the embellishment of simple inferential statements was found, as was the greater difficulty in correctly identifying false verbatim statements as opposed to true ones, presumably reflecting a general tendency to mark answers "true" rather than "false."

REFERENCES

- BEAL, C. R. (1990). Development of knowledge about the role of inference in text comprehension. *Child Development*, *61*, 1011-1023.
- BEAL, C. R., & BELGRAD, S. L. (1990). The development of message evaluation skills in young children. *Child Development*, *61*, 705-712.
- KEENAN, J. M., POTTS, G. R., GOLDING, J. M., & JENNINGS, T. M. (1990). Which elaborative inferences are drawn during reading? A question of methodologies. In D. A. Balota, G. B. F. d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 377-402). Hillsdale, NJ: Erlbaum.
- MARX, M. H. (1991). Development of inferences over elementary-school grades: I. Recall and association of implicit words. *Bulletin of the Psychonomic Society*, *29*, 460-462.
- MARX, M. H. (1992). Development of inferences over elementary-school grades: II. Retention of explicit and implicit words. *Bulletin of the Psychonomic Society*, *30*, 167-169.
- SINGER, M. (1988). Inferences in reading comprehension. In M. Daneman, G. MacKinnon, & T. Weller (Eds.), *Reading Research: Advances in theory and practice* (Vol. 6, pp. 177-219). New York: Academic Press.
- THORNDYKE, P. W. (1976). The role of inferences in discourse comprehension. *Journal of Verbal Learning & Verbal Behavior*, *15*, 437-446.