This study examined effects of inference quizzes on inferential thinking. A total of 213 grade 9 history students were randomly assigned to treatments within classrooms. Group P took eight weekly quizzes requiring recall of stated facts. Group I took 9 quizzes requiring students to draw inferences about interest groups. Teachers had no knowledge of quiz content or treatment membership. No between treatment differences were found on ability to recall Random Facts or Inference-Relevant Facts. Group I scored higher than P on a test requiring new inferences about the treatment interest groups. There were no differences in inferences about a new subject. (Author)
The idea that students adapt their study behavior to their expectations of test requirements is not new (cf. Meyer, 1934, 1935). Indeed the educational power of "testing effects" has been recognized as a variable that may confound studies of other variables (Campbell and Stanley, 1963). And Rothkopf (1968) suggests that test-like events can be used to reinforce, hence develop, desired study behaviors. Rothkopf (1966) and Rothkopf and Bisticos (1969) have used test-like events interspersed in prose to maintain careful reading of successive passages, to improve recall of information not treated directly by quizzes, and to focus attention on particular types of information.

In stating that the kinds of questions teachers ask limit the kinds of thinking students learn, Taba (1966) suggests that test-like events may affect higher order thinking as well, and considerable effort has been invested in training teachers to ask thinking questions. Yet the only studies this author has seen to date on using written questions to develop reasoning skills (Cooper, 1967; Hunkins, 1969) report no effects. Both of these studies required students to identify and adapt to several levels of questions from Bloom's Taxonomy, which would seem to be tremendously complex.

This study attempts to determine if one particular kind of
reasoning can be influenced by administering weekly quizzes which require students to draw a consistent form of inference about interest groups. It is hypothesized that when scores of students given quizzes requiring simple recall of stated facts (group F) are compared with scores of students quizzed on inferences about interest groups (group I) on each of four final tests:

- $F > I$ on recall of Random Stated Facts,
- $I > F$ on recall of Interest Group Facts,
- $I > F$ on drawing new inferences about previously treated interest groups,
- $I > F$ on drawing inferences about new interest groups in a new subject.

A roughly stratified sample of 213 middle- to upper-ability grade 8 U.S. history students (Class averages on D.A.T. Verbal Reasoning scores ranged from 50.23 to 93.60 percentile with a grand mean of 68.6%) served as the study population. Subjects were randomly assigned to one of the two treatment groups within each classroom. Thus both treatments were nested within each classroom and student abilities and teacher behavior equated for the treatments.

Teachers were not told the nature of the treatment quizzes or which students were assigned to each treatment. Students were told only that E, who administered quizzes, was "a student of teaching" who assisted the teacher by giving quizzes, correcting them, and recording scores in the grade book. Otherwise quizzes were not discussed in class.

Each week E distributed both types of quizzes directly to students, according to treatment membership, in a manner intended to obscure the fact that different types of quizzes were being administered. The
previous week's quiz was returned to individuals for a brief inspection as individuals finished. Then all papers were collected. This procedure was repeated for eight weeks, once for each of the chapters III through X of *Land of the Free*.

Inference questions were multiple choice questions stated so that response terms were always four of six economic groups, four of five religious groups or the four geographic divisions of colonial America. These items required students to extrapolate from statements in the text to decide which of the stated interest groups was most probably involved in an event. Appended to each objective item was the question, "Why?" requiring the student to explain his reasoning. For example, given the statement that the Boston Port Bill closed the Port of Boston, and inference item might ask:

The Boston Port Bill most directly affected business of
A) farmers  B) merchants  C) laborers  D) planters

Why?

Factual questions given to group F were common multiple choice items requiring simple recall of stated information. The only change from the type of multiple choice item so common in social studies was the addition of "Why?" to match the "Why" part of the inference items. Thus the five-item factual quizzes appeared quite similar to the five item inference quizzes.

Four final tests were constructed and analyzed for validity and reliability. The first test, Recall of Random Facts (RRF) was composed of items comparable to Factual treatment quiz items but did not repeat treatment items. The second test, the History Inference Test (HIT), was composed of items comparable to (but not repetitions of) the inference
treatment quizzes. These items required students to draw new inferences about the same interest groups as previously treated. The third test, Recall of Interest Group Facts (RIGF), was constructed by writing items to test recall of the specific facts from which HIT items were drawn. An RIGF item might ask students what the Boston Port Bill did, for example. The fourth test, the Inference Transfer Test (ITT) was composed of inference items similar in logical form to HIT items, but involving a new set of interest groups and a new text about a nonexistent country to which students could refer during the examination.

In the ninth week, students were given a sixty item recall test composed of mixed RRF and RIGF items. When papers had been collected, E distributed this same sixty item test with correct answers marked. Students were told to use this test as feedback and as a study guide for the test to be given the following day, as a means of rematching groups on knowledge of information required in answering HIT items. The ITT was administered a week after the History Inference Test.

Results on these four tests are reported in Table I. Hotelling's

<table>
<thead>
<tr>
<th>Table I</th>
<th>Observed Means and Standard Effects on the Four Dependent Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RRF</td>
</tr>
<tr>
<td>Inference</td>
<td>10.289</td>
</tr>
<tr>
<td>Factual</td>
<td>11.163</td>
</tr>
<tr>
<td>Effecta</td>
<td>-.676</td>
</tr>
</tbody>
</table>

a. standardized by dividing differences in means by SE for each test.

$T^2$ was computed on all four tests which revealed that at least one difference could be found on the possible comparisons. Post hoc contrasts on the individual tests indicated that groups differed only on the HIT.
It was concluded that the inference treatment quizzes did, in fact, facilitate the drawing of new inferences within the familiar subject matter, but there were no other reliable effects. Judging from the RGIF scores, this effect cannot be attributed to differential knowledge of facts and seems to imply that the quizzes improved reasoning. If similar results are obtained in subsequent studies, it would appear that the kind of thinking required by classroom quizzes may influence the kinds of thinking students learn.
References


Cooper, J. *Two Types of Social Studies Examinations and Their Effects on Student Learning.* (Doctorial Dissertation, Stanford University, 1967) Ann Arbor: University Microfilms No. 67-17, 404


