Review of Educational Research
Summer, 1986, Vol. 56, No. 2. Pp. 212–242

The Effects of Adjunct Questions on Prose Learning

Christiaan Hamaker
University of Amsterdam

ABSTRACT. The research literature on the effects of factual and higher order adjunct questions is reviewed. The influence of 13 design variables on the direction and size of adjunct-questions effects was investigated. Adjunct questions of all cognitive levels have a strong facilitative effect on repeated test questions and a weaker effect on test questions related to the adjunct questions. Unrelated test questions are affected negatively by factual prequestions and by factual postquestions when study time is controlled. Factual postquestions have a positive effect on unrelated questions only when no time restrictions are imposed. Effect sizes are found to be related to text length, density of adjunct questions, adjunct-question format, test-question format, and the level of performance in the control group; they are unrelated to subjects' age, the interval between reading task and test, whether or not subjects are allowed to consult the text while answering the adjunct questions, and the average distance between adjunct questions and relevant text information. When compared to factual adjunct questions, higher order adjunct questions lead to improved performance on repeated, related, and unrelated higher order test questions, and possibly also on unrelated factual test questions. This indicates that higher order adjunct questions may have a more general facilitative effect than factual adjunct questions. The analysis of a recent adjunct-questions study illustrates the role that review results can play in (re)interpreting experimental findings.

Since Rothkopf's pioneering work (1965, 1966), a steady flow of research studies on the effects of adjunct questions has appeared. Adjunct questions are questions added to an instructional text to influence what is learned from the text. This article draws together the results of these studies in an integrative review.

In an adjunct-questions experiment, subjects are presented with an instructional text that they are asked to study. One or more groups read a text that includes adjunct questions; they are instructed to answer these as they encounter them in the text. In most cases, a control group studies the same text without adjunct questions. Studying the text and answering the adjunct questions will be referred to as the *experimental task*. After the subjects complete the experimental task, they are given the *criterion test*. The analysis of the criterion test results establishes whether and how much the adjunct questions have influenced what was learned.

The first section of this paper summarizes the many variations on this basic

---

This study is based on a portion of the author's doctoral dissertation submitted to the Department of Social Sciences of the University of Amsterdam.

The author gratefully acknowledges the comments of Harry Vorst, and the assistance of Paul Kirschner in collecting the data.

design found in the research literature. Three design features are generally assumed to relate to qualitative differences in the kind of effect produced by adjunct questions. These features are: the cognitive level of the adjunct questions, the relation between the adjunct questions and the test questions, and the placement of adjunct questions in the text. Other design variables may influence the size of the effect of adjunct questions. These include text length, format of the adjunct questions, format of the test questions, and frequency of adjunct questions. The second section describes how studies were selected for the review and how the data base was prepared. It also includes an extended example to illustrate the method. The third section presents the review results.

Major Research Design Features

Cognitive Level of Adjunct Questions

In a majority of adjunct-questions studies, the inserted questions are *factual questions*. They "ask a learner to repeat or recognize some information exactly as it was presented in instruction" (Andre, 1979, p. 282). Other studies use *higher order questions* as adjunct questions. Higher order questions "ask the student to mentally manipulate bits of information previously learned to create an answer, or to support an answer with logically reasoned evidence" (Winne, 1979, p. 14). The category of higher order questions contains a wide variety of question types, as other reviewers have indicated (Andre, 1979; Carrier & Fautsch-Patridge, 1981). The effects of factual adjunct questions and higher order adjunct questions will be reviewed separately.

Position of Adjunct Questions

Four sequential arrangements of experimental text and adjunct questions can be distinguished: (a) *massed prequestions*, in which all adjunct questions occur at the beginning of the text; (b) *inserted prequestions*, in which the adjunct questions are inserted into the text at a number of points, always preceding the text passage containing the information needed to answer them; (c) *inserted postquestions*, in which the adjunct questions are inserted into the text at a number of points, always following the text passage containing the information needed to answer them; (d) *massed postquestions*, in which all adjunct questions are placed together at the end of the text.

Subtests of the Criterion Test

A distinction is generally made between repeated and new test questions (Anderson & Biddle, 1975). Other words, such as *relevant* and *incidental* have been used in the literature to refer to the same dichotomy. *Repeated questions* are test questions encountered previously by the subjects as adjunct questions. New questions are test questions that the subject has not seen prior to their appearance in the test. However, the dichotomy of repeated and new test questions is too simple. In this review, the category of new questions is split into two subcategories, *related* questions and *unrelated* questions. Test questions are classified as related questions whenever a relationship between adjunct questions and test questions can be specified so that the learning activities necessary to answer the adjunct questions

can reasonably be expected to influence positively the learning of information needed to answer the test questions. Five kinds of relationships satisfying this definition have been found in the research literature. They are as follows:

1. The adjunct questions concern a restricted category of text information (e.g., proper names) and the test contains new questions about items from the same category of information (e.g., Rothkopf & Bisbicos, 1967).
2. The test question covers information not directly needed to answer the adjunct questions, but supposedly reviewed while searching for an answer to a postquestion. This may be the case in adjunct inference questions (Frase, 1969, 1970, 1971) or when the test question is drawn from the same sentence as an adjunct question (McGaw & Grotelueschen, 1972).
3. The text contains definitions of concepts or statements of general principles, and adjunct questions require the subjects to identify examples of these concepts, whereas the new test questions require the identification of new examples of the same concepts or principles (e.g., Watts & Anderson, 1971).
4. The adjunct questions are factual questions about specific information and the test contains higher order questions, the answers to which use the information covered by the factual adjunct question(s) (e.g., Andre, Mueller, Womack, Smid, & Tuttle, 1980, especially their Experiment 4). This relation can also occur the other way around, when higher order adjunct questions are followed by factual test questions on the same information.
5. The test question is a paraphrase of the adjunct question rather than a verbatim repetition (e.g., Anderson & Biddle, 1975). The usual method for constructing paraphrase questions is to replace all substantive words (nouns, verbs, adjectives) in the original with synonyms.

Note that questions are not classified as related when their only similarity is cognitive level. For instance, when a text presents concept definitions, and adjunct questions require the learner to identify examples of concept A, then test items would be called related when they require the identification of new examples of A, but called unrelated when they require identification of examples of other concepts.

When none of the four relationships holds between adjunct questions and new test questions, the test questions are said to be unrelated. The distinction between these two subcategories of new test questions is considered a major improvement over previous reviews.

We now turn to a brief review of adjunct-questions effects as they have issued from various qualitative reviews. No detailed theoretical analysis is presented here. The purpose of this section is to set the stage for the results of the integrative review presented in the final section.

Effects of Factual Postquestions

Rickards (1979) reviewed the effects of factual postquestions. He found evidence for four effect types. Two of these are due to "backward processing," i.e., mental processing of text information read before the encounter with the adjunct question. The *specific backward effect* of postquestions is caused by direct review of material actually questioned; the *general backward effect* is due to "the mental review of material adjacent to and/or thematically related to the questioned material" (Rickards, 1979, p. 183). In massed postquestions treatments, facilitative effects of

adjunct questions can only be backward effects. The specific backward effect shows up as improved performance on repeated test questions, whereas the general backward effect would be visible in related test questions.

Inserted postquestions produce backward effects by the same processes as massed postquestions, but in addition they may act forwardly, affecting the processing of text information read after the encounter with the adjunct question. In the *specific forward effect*, "the learner develops a set to attend to the particular type of information being questioned" (Rickards, 1979, p. 183). The specific forward effect will show up in repeated questions and in some types of related questions, specifically the type used by Rothkopf and Bisbicos (1967). The *general forward effect* is due to an increase in the overall level of attention to the experimental task. This general stimulatory process will influence performance on all types of test questions, including test questions totally unrelated to the adjunct questions.

Effects of Factual Prequestions

Prequestions are supposed to have forward effects only. The distinction between massed and inserted prequestion designs is immaterial to the theoretical process through which prequestions produce their effect. Factual prequestions may narrow the experimental task to a search task. The text is processed only as much as is necessary to locate information relevant to the adjunct questions; question-relevant information will be further processed once it has been located; and processing of all other information ceases as soon as its irrelevance for the adjunct questions has been established. This argument would lead one to predict a positive effect of prequestions on repeated test questions; a smaller positive effect on related test questions for which the answers probably have been processed in the search for question-relevant information; and a negative effect on unrelated questions in the test.

Effects of Higher Order Adjunct Questions

The use of higher order adjunct questions adds a new dimension to the discussion: the level of the effect. Most higher order adjunct-questions research primarily concerns *higher order effects*, i.e., the influence of adjunct questions on the learner's ability to use text information in a variety of higher order criterion tests, such as inference and application questions. In addition, however, higher order adjunct questions are often supposed to affect the retention of text information. This will be referred to as the *lower order effect* of higher order adjunct questions. Andre (1979) presented the most elaborate model to date of higher order adjunct-questions effects.

Higher order effects. The assumption common to all higher order adjunct-questions research is that this type of question induces in the learner higher order processing activities, such as integrating and elaborating. The basic characteristic of higher order processing activities is that they cause the learner to produce new information. This information is available for subsequent use in answering criterion test questions. It is highly likely that performance on repeated questions will profit from the availability of this information. Moreover, it may facilitate performance on related test questions whenever some of the information produced contributes to the answer of the related questions. An example adapted from Frase (1969)

illustrates this idea. His text presented a series of four inclusion relations of the form *All A are B*, *All B are C*, *All C are D*, and *All D are E*. An adjunct question might be "Is it true that all A are E?" In the process of answering this question, inferences are produced, such as *All A are C*, *All A are D*, and so forth. Therefore, we may expect facilitation of the related criterion test question "Is it true that all A are D?"

Finally, inserted higher order questions can be supposed to have the same general stimulatory effect as inserted factual postquestions. This effect would show up in unrelated higher order test questions. Note, however, that Andre (1979) concluded that there is no empirical support for this general effect.

Lower order effects. Higher order adjunct questions may or may not have positive effects on factual test questions as well. Whenever factual information has to be reviewed to answer a higher order adjunct question, improved performance on related factual test questions is to be expected. For example, when a higher order question requires the application of a concept definition, performance on a test question requiring recall of the definition may improve.

Other Design Variables

Apart from the three major design features, other variables in adjunct-questions research may influence the size of adjunct-questions effects. The variables that have been investigated in the review are presented here.

1. *Format of the adjunct questions.* Some studies use short-answer questions as adjunct questions; others use the multiple-choice format. Anderson and Biddle (1975) provided evidence that short-answer adjunct questions produce effects approximately two and a half times as large as those produced by adjunct questions in multiple-choice format.

2. *Test question format.* Three formats occur regularly in the research literature: multiple-choice, short-answer, and free-recall.

3. *Control performance.* This is the mean score in the control group expressed as a percentage of the maximum possible score. This variable was included in the review because low control performance indicates that the test leaves a wide margin for improvement through adjunct aids.

4. *Age level of subjects.* This variable takes three values: *elementary school*, *high school*, and *college*. It has been suggested (Rickards & Denner, 1979) that adjunct questions are less effective with, or even a hindrance to, young children.

5. *Lookbacks.* It is likely to make a difference whether or not subjects are allowed to reread the text while answering adjunct questions. The variable takes two values: *lookbacks allowed* and *lookbacks not allowed*. It has been said to demarcate the difference between ecologically valid (lookbacks allowed) and invalid (lookbacks not allowed) adjunct-questions studies (Duchastel, 1983).

6. *Text length.* Text length is expressed as the number of words in the text. This variable is correlated with the general nature of the study. For example, the longer texts are all taken from existing sources, such as textbooks, whereas only the shorter texts are taken from fictitious material.

7. *Coverage.* The proportion of text information covered by adjunct questions varies from one study to the next. No variable reflecting this feature is directly available in the reports, so we had to compute one. We divided text length by the

number of adjunct questions. This variable is admittedly a rough indication of the amount of text information covered, but we could not conceive of a better alternative. High values of this variable indicate that a small proportion of the text is covered by adjunct questions.

8. *Distance.* The time elapsed between reading relevant material and encountering an adjunct question is also likely to influence the strength of the effect. Again, we had to compute a variable from available information. We used text length divided by twice the number of times the text is interrupted by adjunct questions. To illustrate, Frase (1967) used a 2000-word passage containing 20 paragraphs and 20 adjunct questions. The questions were inserted one at a time after every paragraph, two at a time after every two paragraphs, or four at a time after every four paragraphs. The respective values of the variable distance are 50, 100 and 200.

9. *Time of testing.* This is the time elapsed between the end of the experimental study task and the beginning of the criterion test task. This variable is coded on a seven-point scale ranging from *immediate testing* to *a delay of more than one week.* Most studies employ a single criterion test, but in 14 experiments subjects were tested twice. In these 14 studies, the size of adjunct-question effects in the immediate tests was of the same magnitude as the effects in delayed tests. An interaction found by Natkin and Stahler (1969) seemed to be of interest, but a systematic effort by Wyatts (1973) to replicate it failed completely. Other studies came up with no interactions, or significant but uninterpretable interactions (Rickards, 1976a; Sanders, 1973; Swenson & Kulhavy, 1974). Therefore, we felt justified in combining the data from the two tests in a single effect size. This was done because we wanted to include no more than one effect size of a specific type for each experimental treatment to have independent estimates of any given effect type.

10. *Time control.* In approximately half of the studies, students are allotted a fixed amount of time to spend on the experimental task; in the other half, time on the experimental task is not controlled by the experimenter.

Method

Selection of Studies and Outcomes

The group of adjunct-questions studies is remarkably closed. Though some studies with adjunct questions were conducted earlier, the first study by Rothkopf (1966) is the starting point here. Adjunct-questions studies can generally be identified by their reference to Rothkopf's study, or to other studies by Rothkopf and Frase published shortly after 1966. Despite this exclusivity, borderline cases do, of course, occur. This section gives the criteria used to include studies for the review.

The review is limited to published sources from 1966–1982. Conference papers, technical reports, and dissertations are not included. This limitation could lead to overestimation of effects, because unpublished studies generally are associated with smaller effect sizes, if not null effects (Glass, McGaw, & Smith, 1981). In the present study, review results can be compared directly with those of an earlier review (Anderson & Biddle, 1975) that included published and unpublished sources. The results of this comparison will be presented in the Results section. They show no evidence of a general bias in favor of positive results in published studies.

Studies are included according to three criteria. First, the task must be a prose-learning task. Experiments that use paired-associate learning or the learning of

isolated sentences are excluded. The criterion also excludes programmed instruction. Second, experiments in which detailed informational feedback was given to the subjects who answered adjunct questions were also excluded. (Students in feedback treatments may adopt a very passive attitude toward the adjunct questions, because they know that the correct answer will be presented to them. For that reason, the inclusion of studies with informational feedback might have caused an underestimate of adjunct-questions effects.) Third, the stated instruction to the subjects must include directions to study and learn the passage as a whole. Research in which the subject's only task is to check the truth of a given statement are excluded from the review. (In such an experiment, the subject's task is a search task rather than a prose-learning task. If search tasks had been included in the review, the effect of adjunct questions on unrelated criterion test questions might have been underestimated.)

Two methods of locating relevant research reports were used in combination. One is the ancestry approach (Cooper, 1982). Earlier adjunct-question studies were located by checking reference lists in reports. The other method is computer search. The ERIC system and *Psychological Abstracts* were searched with a combination of keywords commonly used as synonyms or subcategories of adjunct questions in the literature. These included *inserted questions*, *adjunct questions*, *test-like events*, *prequestions*, *postquestions*, *verbatim questions*, and *higher order questions*.

A total of 77 publications reporting on adjunct-questions research were located. Some of these studies did not present quantitative information about learning, because their primary purpose was either to study learning in a qualitative way, or to study some noncognitive aspect of adjunct questions. Others were not used because of their atypical design. These include Memory (1981), who compared a postquestions treatment with a treatment in which subjects answered prequestions and postquestions, and Grant, Keenan, and Hursh (1980) who used a different text for each treatment.

A considerable number of experiments in the pool have designs in which the adjunct-questions variable is combined with one or more other treatment variables. Combinations can occur either in a design in which adjunct questions are compared with some other treatment or treatments, or in a design in which the adjunct questions are combined factorially with other treatments. From comparative designs, only the adjunct-questions treatments and the control groups were used for the review; information from other treatments was discarded. These include some adjunct-questions treatments that cannot be classified unambiguously by means of the three major design variables. Examples are the "structural questions" used by McConkie, Rayner, and Wilsons (1973), which cannot be unambiguously classified as either factual or higher order questions, and the treatment groups from Boyd's (1973) study that received adjunct questions before and after reading relevant text passages. In factorial designs, data from groups receiving the same adjunct-questions treatment but different treatments on the other variables, were generally combined and treated in the review as a single experimental group, yielding one outcome and one effect size.

Methodological Quality of the Studies

On the whole, the methodological quality of adjunct-questions research is quite good. The field has a tradition of straightforward experimentation. This includes

random assignment of subjects to treatments and adequate data analysis, mostly analysis of variance.

In a small number of studies, the research method was inadequate. The most frequent defect was insufficient differentiation between repeated, related, and unrelated questions in the criterion test. For this reason five studies were dropped: Ackerman (1981), Adejumo (1980), Prosser (1974), Santiesteban and Koran (1977), and Wiesendanger and Wollenberg (1978).

In many reports, standard deviations of test scores are not reported. In our review this was not considered a serious omission, because the effect size statistic used does not require the availability of standard deviations (see following section). Even so, the publication of standard deviations should be standard practice to allow computation of other effect size statistics when the need arises.

In some reports information about means is incomplete (Frase, 1967). In others it had to be reconstructed, either by combining scattered pieces of information (e.g., in Rickards, Anderson, & McCormick, 1976), or by reading off the information from graphs (e.g., Frase, Patrick, & Schumer, 1970; Watts, 1973). Four experiments could not be used in the review, because the reports contained insufficient information about criterion test means. These are in Andre et al. (1980, Experiments 6 and 7), Tergan (1979), and Wilson (1979).

Calculating Effect Sizes

An important choice in conducting an integrative review is the method for quantifying the findings of the studies. The alternatives are well known (Glass, McGaw, & Smith, 1981; Green & Hall, 1984), but the criteria for making a choice are less clear. Green and Hall discuss the standardized mean difference, d, and then conclude, "When outcome measures are clearly identifiable and comparable across studies, these values can be used directly rather than resorting to an index like d" (p. 44). This criterion can be paraphrased as: Standardization is unnecessary when a natural common scale exists on which experimental outcomes can be compared.

In the case of adjunct-questions studies, an interpretable common scale is available, because the result of any experimental treatment can be expressed as the difference between (a) the percentage of criterion test questions answered correctly by experimental subjects, and (b) the same percentage in the control group. The idea of a common scale is supported by the narrow range of reported standard deviations. Approximately 25% of the studies in the review reported standard deviations; the mean standard deviation of the proportion correct in these studies is .16, and 70% of the standard deviations are between .12 and .19. An additional advantage of using the simple effect size statistic is the direct comparability of our review results with those obtained by Anderson and Biddle (1975), who used the same effect size statistic.

The formula used for the factual adjunct-questions review follows directly from the preceding discussion:

$$ES = p(T) - p(C),$$

in which ES is effect size, $p(T)$ is the proportion of criterion test items correct in the treatment group, and $p(C)$ is the proportion correct in the control group.

The formula for the higher order review is different because many higher order

adjunct-questions studies do not compare performance of a higher order adjunct-questions group with that of a read-only control group, but with that of a group receiving factual adjunct questions. In these studies, the factual adjunct-questions treatment is viewed as a control treatment rather than as an experimental treatment. The formula used is:

$$ES = p(H) - p(L),$$

in which $p(H)$ is the proportion correct in the group receiving higher order adjunct questions, and $p(L)$ is the proportion correct in the factual adjunct-questions group.

It was noted previously that, in the studies reviewed, the mean value of the standard deviation of the proportion correct was equal to .16; therefore, an effect size of .16 in this review represents a one-sigma effect on Bloom's (1984) scale, and corresponds to an effect size of 1.00 in reviews that use the standardized means formula. In power analysis (Cohen, 1977), an effect size of one fifth of a standard deviation is considered a small effect. This corresponds to an effect size of .03 in this review. Similarly, effect sizes of about .08 and .13 in this review would correspond to medium and large effect sizes by Cohen's standards.

Mean effect sizes can be computed in a number of different ways, the following two of which have been used:

1. *The simple mean.* When the effect sizes are averaged without weighting, treatments are used as the unit of analysis. A rationale for this use of treatments is that the review should indicate whether or not an effect occurs across different operational realizations of the *theoretical treatment*. The term *theoretical treatment* here refers to a type of treatment, such as *inserted factual postquestions*. The specific questions inserted with a specific frequency into a specific text used in an experiment constitute an operational realization of this theoretical treatment.

2. *A weighted mean.* With all else constant, the reliability of an observation (i.e., an effect size) obtained from a study will increase as the number of subjects and/or the number of criterion test questions employed in the study increases. To account for these differences in reliability of observations, a weighted mean was computed in which each effect size was weighted by the number of criterion test questions and the number of subjects on which it is based. In this review, simple means will be reported generally; weighted means will be given occasionally as additional grounds for specific conclusions.

An Extended Example

The method of encoding research studies and computing effect sizes will be illustrated by an extended example, using the Watts and Anderson (1971) study. This is a fairly complex study, employing factual and higher order questions as adjunct questions and in the criterion test. Because of this complexity, it is a good illustration of the difficulties encountered in preparing the data, and the way these difficulties were solved.

Watts and Anderson used a text explaining a number of psychological concepts. Each section of the text gave a definition and two examples of a concept, and mentioned a famous psychologist whose name is linked with the concept. They used three types of adjunct questions: name questions, repeated example questions, and new example questions. One group received name questions with each section

of the passage. The name questions required the subject to identify from four alternatives the name of the psychologist associated with a given concept. This group is indicated by the letter N. Another group received one repeated example question with each section. This group is identified as RE1. A repeated example question asked the subject to identify an example presented in the text from four alternatives. Because there were two examples of each concept in the passage, it was possible to compose two non-overlapping sets of repeated example questions. The group that received the second set of repeated example questions is identified as RE2. Similarly, two sets of new example questions were constructed. The new example questions asked the subject to identify a new example of a given concept from four alternatives. The two groups receiving either set of new example questions are identified as A1 and A2. Finally, the design included a group who read the passage without any adjunct questions (C group).

For this review, the three adjunct-question types first had to be classified as either higher order or factual questions. In the name questions and the repeated example questions, subjects were asked to identify a piece of information they had seen in the passage. Therefore, the task is a simple recognition task, and these question types were classified as factual adjunct questions. To answer new example questions, more complex mental processing seems necessary. Subjects have to check the presence of defining characteristics in the alternatives, or they must compare each alternative with a prototype of the concept, or some such process. For this reason, new example questions are classified as higher order adjunct questions. The criterion test consisted of all the questions used as adjunct questions. The results as presented in the Watts and Anderson (1971) paper are given in Table I.

For the factual adjunct-questions review, no data from either higher order adjunct-questions groups or from higher order criterion test questions were used. For the remaining treatment groups and test questions, the relation of the adjunct questions to the criterion test items had to be classified as repeated, related or unrelated. Name questions were repeated questions for the N group, and unrelated questions for the RE groups. Half of the RE test questions were repeated questions for the RE groups; the other half of the RE questions were related questions. Finally, the name questions in the test were unrelated questions for both RE groups.

TABLE I

Percentage Correct for Six Experimental Groups on Five Criterion Subtests (Data from Watts and Anderson, 1971)

Corresponding criterion subtest	Treatment Groups					
	RE1[a]	RE2[b]	A1[c]	A2[c]	N[d]	Control
RE1	88.0	84.4	88.4	90.0	79.6	88.0
RE2	75.6	84.0	85.6	84.8	79.2	81.6
A1	48.0	50.4	70.4	63.6	42.8	48.0
A2	58.8	57.6	70.4	71.6	47.6	52.8
N	36.4	46.8	48.8	50.5	52.8	50.4

[a] Group that received one repeated example adjunct question.
[b] Group that received two repeated example adjunct questions.
[c] Group that received one of two different sets of new example questions.
[d] Group that received name adjunct questions.

TABLE II
Percentages Correct and Effect Sizes for Combined Groups and Measurements from Table I

Effect	Treatment			Effect size
	Repeated example	Name	Control	
		Name questions subtest		
Repeated		52.8	50.4	+.02
Unrelated	41.6		50.4	−.09
		Repeated example subtest		
Repeated	86.0		84.8	+.01
Related	80.0		84.8	−.05
Unrelated		79.4	84.8	−.05

Treatments RE1 and RE2 are considered essentially identical operational realizations of inserted factual postquestions. Therefore, the results of these groups have been collapsed. As a result, five effects were calculated for the factual adjunct-questions review: a repeated and an unrelated effect for the N group, and repeated, related and unrelated effects for the combined RE groups. For the Watts and Anderson data, the computation of the effect sizes is straightforward, because the data are presented as percentages. From Table I the percentages presented in Table II are computed by averaging data from RE groups and RE test questions.

For the higher order adjunct-questions review, two comparisons between treatments were used. The first was between new example and repeated example groups; the second, between new example and name groups. Recall that comparisons for the higher order review are always between higher order and factual adjunct-questions groups, rather than between higher order adjunct-questions groups and a read-only control group. In all comparisons the treatment subgroups (A1, A2, RE1, and RE2) were combined as in the factual adjunct questions review.

The higher order review used six possible types of effects: repeated, related and unrelated effects on lower order test questions, and repeated, related and unrelated effects on higher level test questions. The classification of the Watts and Anderson data is given in Table III. Note that this particular study contributes data on five of the six possible effect types; it provides no information on the effects of higher order adjunct questions on unrelated higher order test questions.

Nine effect sizes were computed from these comparisons as indicated in Table IV. As can be seen in this table, the results from the parallel A and RE groups were averaged to yield a single effect size of each type. However, the results from the A–RE and A–N comparisons were not combined. This reflects the general rule that a separate effect size is computed for each separate operational realization of a theoretical treatment. A disadvantage of this rule is that the two effect sizes are not independent, because they have one group in common.

The Studies Used in the Review

After the elimination of all those studies that for one reason or another could not be used in the review, 61 published experiments remained. Fifty experiments, contributing 204 outcomes, were used in reviewing the effects of factual adjunct

TABLE III
Classification of Effect Types for the Higher Order Review (Data from Watts and Anderson, 1971)

Corresponding criterion subtest	Treatment comparison					
	A1–RE1	A1–RE2	A2–RE1	A2–RE2	A1–N	A2–N
A1	Hrep	Hrep	Hrel	Hrel	Hrep	Hrel
A2	Hrel	Hrel	Hrep	Hrep	Hrel	Hrep
RE1	Lrep	Lrel	Lrep	Lrel	Lunr	Lunr
RE2	Lrel	Lrep	Lrel	Lrep	Lunr	Lunr
N	Lunr	Lunr	Lunr	Lunr	Lrep	Lrep

Note. The meaning of Treatment and Subtest abbreviations is given in Table I. Hrep = effect on repeated higher order questions; Hrel = effect on related higher order questions; Lrep = effect on repeated lower order questions; Lrel = effect on related lower order questions; Lunr = effect on unrelated lower order questions.

TABLE IV
Effects Computed for Higher Order Review from Watts and Anderson (1971) Data

Criterion subtest	Treatment comparison	
	New example–repeated examples	New example–name questions
New example	Hrep	Hrep
	Hrel	Hrel
Repeated example	Lrep	Lunr
	Lrel	
Name questions	Lunr	Lrep

Note. Abbreviations are explained in Table III.

questions; 21 experiments, contributing 100 outcomes, were used for the review of higher-order adjunct questions. Table V lists the 61 experiments. Some of the text lengths given in Table V have been estimated from information in the article; a question mark is entered in Table V only when no reliable estimate could be made.

Results

Effects of Factual Adjunct Questions

The average effect sizes for prequestions and postquestions on the three types of criterion test questions are discussed first. Table VI presents the results.

Repeated test questions. Prequestions as well as postquestions have a sizable facilitative effect on repeated questions: $t(22) = 4.55$, $p < .01$, and $t(58) = 12.21$, $p < .01$, respectively. This is no surprise. Both effects are well-established in the literature. The effect found for prequestions in the present review is larger than that found by Anderson and Biddle (1975). They found a mean effect of .108. The difference is caused by the studies published since 1975. These have a larger average effect ($M = .21$) than the studies published before 1975 ($M = .11$), $t(21) = 1.58$, $p > .10$. There is better agreement on the average effect of postquestions on repeated test questions. The present review found a somewhat larger average effect of studies

TABLE V

Studies and Length of Texts Used in Reviewing Adjunct-Question Effects

Experiment	Text length in words	Adjunct questions	
		Factual	Higher order
Allen, 1970	5,500		×
Anderson and Biddle, 1975, I	550	×	
Anderson and Biddle, 1975, II	550	×	
Anderson and Biddle, 1975, III	600	×	
Andre and Womack, 1978	5,000	×	
Andre et al., 1980, I	2,750		×
Andre et al., 1980, II	1,375		×
Andre et al., 1980, III	825		×
Andre et al., 1980, IV	?		×
Andre et al., 1980, V	1,650		×
Bing, 1982	1,700	×	×
Boker, 1974	2,500	×	
Boyd, 1973	2,000	×	
Bull and Dizney, 1973	?	×	×
Duchastel, 1979	750	×	
Duchastel, 1981	1,700	×	
Eischens, Gaite, and Kumar, 1972	1,354	×	
Ellis, Wulfeck, and Montague, 1980	?	×	
Felker and Dapra, 1975	3,000	×	×
Frase, 1967	2,000	×	
Frase, 1971	258		×
Frase et al., 1970	2,000	×	
Frase and Schwartz, 1975	1,218	×	
Friedman and Rickards, 1981	?	×	×
Gagne and Nemory, 1978	350	×	×
Graves and Clark, 1981	2,000	×	
Gustafson and Toole, 1970	23,000	×	
Hiller, 1974	1,800	×	
Hunkins, 1969	?		×
Koran and Koran, 1975	5,000	×	
Laporte and Voss, 1975	1,500	×	
Mayer, 1975	?	×	×
McGaw and Grotelueschen, 1972, I	5,460	×	
McGaw and Grotelueschen, 1972, II	5,460	×	
McKenzie, 1972	?		×
Natkin and Stahler, 1969	2,500	×	
Peeck, 1970	3,000	×	
Reynolds and Anderson, 1982	9,500	×	
Reynolds, Standiford, and Anderson, 1979	9,500	×	
Rickards, 1976a	800	×	×
Rickards, 1976b	2,100	×	
Rickards et al., 1976	640	×	
Rickards and Denner, 1979	800	×	
Rickards and Di Vesta, 1974	800		×
Rickards and Hatcher, 1977–1978	800	×	×

TABLE V—*Continued*

Experiment	Text length in words	Adjunct questions	
		Factual	Higher order
Rothkopf, 1966	5,200	×	
Rothkopf, 1972	14,200	×	
Rothkopf and Bisbicos, 1967	9,000	×	
Rothkopf and Bloom, 1970	16,200	×	
Sagaria and Di Vesta, 1978	800	×	
Samuels, 1969	?	×	
Sanders, 1973	2,000	×	
Snowman and Cunningham, 1975	2,189	×	
Swenson and Kulhavy, 1974	1,320	×	
Walker, 1974, I	2,240	×	
Walker, 1974, II	2,240	×	
Watts, 1973	700	×	
Watts, 1974	135	×	×
Watts, 1975	1,000	×	
Watts and Anderson, 1971	2,250	×	×
Yost, Avila, and Vexler, 1977	?		×

TABLE VI

Mean Effect Size, Standard Deviations and Number of Outcomes of Factual Adjunct Questions by Position and Criterion Test Type

Question position	Criterion test type								
	Repeated			Related			Unrelated		
	M	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
Prequestions	.15	.15	23	.09	.11	5	−.05	.08	25
Postquestions	.16	.10	59	.07	.07	34	.01	.07	50

published before 1975 ($M = .17$) than Anderson and Biddle ($M = .132$). The studies published after 1975 continue to find a clear direct effect of postquestions ($M = .15$).

The differences between the results of the two reviews with respect to studies published before 1975 may be due to differences in selection strategy. The number of outcomes on which Anderson and Biddle based their conclusions is larger than the number of outcomes from pre-1975 studies in the present review. It must be assumed that Anderson and Biddle included unpublished studies, but they do not describe their data base in sufficient detail to be certain.

The effects of prequestions and postquestions on repeated test questions are important, even in terms of educational practice. The average performance of the control groups in the experiments reviewed is .40. An average effect size of .15 represents an increase of 37.5% over control performance. By reference to still another effect size scale (Bloom, 1984), the direct effect of adjunct questions has a size of approximately 1 sigma. This means that adjunct questions are slightly less

effective than reinforcement, and equally effective as such instructional strategies as mastery learning, student time on task, and training of study skills.

Related test questions. On related test questions, prequestions as well as postquestions have a positive effect that is approximately half as large as their effect on repeated test questions. These effect sizes cannot be compared with those reported by Anderson and Biddle, who did not distinguish between related and unrelated test questions. The effect of prequestions ($M = .09$) is not significant, $t(4) = 1.80$, $p > .05$, one-tailed, but note the very small number of observations. The effect of postquestions ($M = .07$) is smaller, but highly significant, $t(33) = 5.50$, $p < .01$. From the present review, it can be concluded that an effect of factual adjunct questions on related criterion test questions stands on firm empirical ground.

Unrelated test questions. Prequestions have a negative effect on unrelated test questions ($M = -.05$, $t(24) = -2.91$, $p < .01$). This agrees with the idea that prequestions reduce the task to a search task, and that consequently text information not relevant to the adjunct questions is processed less extensively than it is in the absence of prequestions. Yet, this negative effect is not very large; by power analysis standards it would be considered a small to medium effect.

An interesting result of this review is the absence of an effect of postquestions on unrelated test questions [$M = .01$, $t(49) = .75$, $p > .20$]. Reviewers and experimenters have concluded that postquestions have a generally facilitative effect on prose learning. Frase (1968b) was the first to make this general conclusion from adjunct-questions studies: "The results . . . indicate that both pre-questioning and review questions have a facilitative effect upon the retention of the question relevant information, but that review questions can be generally facilitating" (p. 324). Anderson and Biddle (1975) concluded: "Adjunct questions after sections of text significantly more frequently have a positive than a negative effect on unrelated test items" (p. 92). Recently, Duchastel (1983) wrote: "It is mainly the effects of post-questions which create the most interest, for it is they that enhance both relevant (question-related) learning and incidental (unrelated) learning" (p. 1).

Statements about the effect of adjunct questions have found their way into the *Annual Review of Psychology* ever since the adjunct-questions research got underway in 1966. Anderson (1967), Gagne and Rohwer (1969), Glaser and Resnick (1972), and McKeachie (1974) all concluded that postquestions could have general facilitative effects. Wittrock and Lumsdaine (1977) are the only reviewers to have been skeptical about this general effect. Resnick (1981) abstained from generalizations concerning adjunct-questions effects.

The data in Table VI show that the conclusion of a general facilitative effect of postquestions has been overoptimistic. As a check on the negative conclusion reached here, weighted average effects were calculated. When this is done, the average size of the effect of postquestions on unrelated criterion test questions changes sign ($M = -.01$ instead of $M = .01$ in Table VI). Therefore the absence of a general indirect effect of postquestions in the results in Table VI cannot be ascribed to the presence of a large number of negative studies using few subjects and criterion test questions in the data base.

Although most generalizations about the presumed facilitative effect of postquestions have not commented on the difference between inserted and massed postquestions, it was argued above that a facilitative effect on unrelated test questions is only to be expected when questions are inserted. Adjunct questions

can have their general stimulatory effect only when learners are reminded from time to time that they are expected to be able to answer questions about what they are reading. For this reason, separate mean effect sizes were computed for the two positions of postquestions. The average effect of massed postquestions on unrelated test questions is −.02; the average effect of inserted postquestions is .01. This difference is not significant, $t(48) = .90$, $p > .20$. When weighted means are computed the same values appear. In conclusion, there is no confirmation in the data for the hypothesis that inserted postquestions have a facilitative effect on the learning of unrelated test questions. However, this does not rule out the possibility that *under specific conditions* postquestions do produce a positive effect on unrelated test questions. This possibility is explored here.

Format of adjunct questions and test questions. Anderson and Biddle (1975) found a strong influence of the format of adjunct questions on effect size. Short-answer adjunct questions were found to have an average effect two and a half times as large as multiple-choice questions. The same analysis was conducted in the present review. The results of this analysis are presented in Table VII.

Most entries in Table VII are computed from a small number of outcomes. This is particularly true for the entries corresponding to studies employing different formats for adjunct questions and test questions. The following conclusions must, therefore, be considered with caution. The Total column confirms Anderson and

TABLE VII
Mean Effect Sizes for Adjunct-Question and Test-Question Formats

Adjunct-question format	Test format: Multiple choice		Test format: Short answer or free recall		Total	
	M	*n*	*M*	*n*	*M*	*n*
Prequestions						
Multiple-choice						
Repeated	.06	7	—	—	.06	7
Related	.07	1	—	—	.07	1
Unrelated	−.04	8	—	—	−.04	8
Short-answer						
Repeated	.17	3	.20	10	.19	13
Related	.15	1	.07	3	.09	4
Unrelated	−.04	3	−.07	11	−.06	14
Postquestions						
Multiple-choice						
Repeated	.10	11	.12	4	.10	15
Related	−.05	1	.03	3	.01	4
Unrelated	.02	12	−.02	3	.01	15
Short-answer						
Repeated	.11	8	.17	26	.16	34
Related	.05	8	.08	22	.07	30
Unrelated	—	—	.02	31	.02	31

Biddle's (1975) finding that short-answer adjunct questions produce larger effects on repeated questions than multiple-choice adjunct questions, $t(18) = 7.44$, $p < .01$. The difference is smaller than that reported by Anderson and Biddle, however, and when the weighted means are considered the difference is even smaller ($M = .16$ for short-answer postquestions, and $M = .11$ for multiple-choice postquestions). The superiority of short-answer adjunct questions is also visible in the effects of postquestions on related test questions, $t(32) = 1.85$, $p < .05$, one-tailed. Adjunct-question format does not affect the size of effects on unrelated questions.

Another issue concerning adjunct-question format must be raised. Because adjunct-question format and test-question format are highly correlated, it might be suggested that the superiority of short-answer adjunct questions over multiple-choice adjunct questions is not caused by the adjunct questions themselves, but by the different methods of measuring the effect. The format of the test questions might influence effect size in a way similar to the ceiling effect. In the research reviewed, multiple-choice tests result in higher control performance ($M = .59$) than do short-answer tests ($M = .31$), and a high control performance narrows the margin for improvement through adjunct questions. The data in Table VII confirm the influence of test format, but they also show that when experiments using the same test format are compared, short-answer adjunct questions still produce larger effect sizes than do multiple-choice adjunct questions. In conclusion, from the comparison of short-answer and multiple-choice adjunct questions, the short-answer format emerges as superior, but by a smaller margin than in Anderson and Biddle's review.

Subject populations. The majority of adjunct-questions studies have been conducted with college students as subjects. In the review data, 17 outcomes out of a total of 204 are from studies with elementary school children, and 48 are from studies with high school students. The rest are from studies on college students. For elementary school and high school studies, no separate mean effect sizes for prequestions studies will be reported here, because very few observations are available for each effect type (no n is larger than 3).

Elementary school children show no significant positive effect of postquestions on repeated test questions, $M = .07$, $t(5) = 1.39$, $p > .10$, but the number of observations is small; the effect of postquestions on unrelated test questions is similarly nonsignificant in this group, $M = .04$, $t(3) = 1.48$, $p > .10$. For related questions, only a single observation was available.

The pattern of postquestions results of high school studies is similar to the overall pattern. We found significant effects on repeated questions, $M = .12$, $t(18) = 7.95$, $p < .01$, and on related questions, $M = .04$, $t(12) = 2.29$, $p < .05$, but not on unrelated questions, $M = -.01$, $t(9) = .43$, $p > .20$. The review sheds no new light on the issue of the effectiveness of adjunct questions for young children.

Lookbacks. In the vast majority of adjunct-questions experiments, subjects were not allowed to review the text at the moment adjunct questions were encountered. Only 13 outcomes in the review data are derived from studies in which lookbacks were permitted (or, in the case of prequestions, the consultation of the questions during reading).

It could be argued that the absence of an effect of postquestions on unrelated test questions is due to the inclusion of lookback studies in the data. When lookbacks are allowed, learners are likely to be less motivated to study the whole

text; they can process the text superficially, waiting for the adjunct question to appear, and then search the text for the answer. Therefore, a stronger effect of postquestions on unrelated test questions might be expected when mean effect sizes are computed from nonlookback studies only. This expectation is not supported by the facts. When the three lookback outcomes ($M = -.02$) are removed, the remaining 44 nonlookback outcomes still have a mean of .01. The difference is not significant, $t(45) = .61$, $p > .20$.

Andre (1981) directly compared a lookback condition with a nonlookback condition without using a control group. He found that these groups perform on an equal level on related test questions, whereas the lookback group did slightly better than the nonlookback group on unrelated test questions. This last result is the opposite of what was expected. Allowing subjects to review the text when answering adjunct questions has not been proven to be an important research variable. This may be due, however, to the scarcity of lookback studies.

These data contradict Duchastel's (1983) suggestion that the results from lookback studies are different from those of nonlookback studies. Duchastel supposed that the preponderance of nonlookback studies in the adjunct-questions literature severely limits the generalizability of the results to educational settings, in which the standard procedure in textbooks and other instructional materials is to allow lookbacks to the students. Whereas this may be true in other respects, the available evidence shows no influence of this design variable with respect to the amount of learning. It must be remembered, however, that this conclusion is weak because so few lookback studies are available. We return to this point in the Discussion section in relation to a recent study by Duchastel and Nungester (1984).

Time of testing. The criterion test was administered either immediately or with a delay of a few minutes to more than 2 weeks. Even a delay of 2 weeks is a short period in comparison with the retention period considered desirable in educational settings. Despite this fact, it is interesting to explore the hypothesis that effect size depends on the time of testing. As was briefly indicated above, those studies that measured the effect of adjunct questions twice, once immediately after the experimental task and again after a delay, did not produce meaningful or consistent interactions of effect size with time of testing.

For these studies, we also computed separate effect sizes for the immediate and the delayed test. Only a small subset of reports provides information on the most important effect types. The effect of prequestions on repeated test questions is slightly larger on the immediate test ($M = .22$; $n = 3$), than on the delayed test ($M = .18$). The same is true for the effect of postquestions on repeated test questions ($M = .22$ on the immediate test; $M = .19$ on the delayed test; $n = 5$). The effect of prequestions on unrelated test questions does not change much ($M = -.04$ on the immediate test; $M = -.03$ on the delayed test; $n = 4$). The effect of postquestions on unrelated test questions is positive in this set of studies and increases over time ($M = .04$ on the immediate test; $M = .07$ on the delayed test; $n = 8$).

Table VIII displays the results of the studies in which the effect of adjunct questions was measured only once. Only data for postquestions studies are shown because the number of prequestions studies with delayed testing was too small to yield meaningful results. Table VIII shows the effect on repeated test questions to be very stable. The effect on unrelated test questions shows a tendency to increase in the first 2 days after studying and to decrease with longer delay, but the studies

are too few to indicate a trend. In conclusion, effect sizes do not seem to depend on the time interval between task and test.

Time control. There has been some discussion about time control in adjunct-questions research (Faw & Waller, 1976). It has been argued that some experimental results might be accounted for by the extra time spent on the experimental task in adjunct-questions treatments. Peeck's (1970) study showed that inducing students to spend more time on the task by means other than adjunct questions caused the same effect as an adjunct-questions treatment. In general, when time is not controlled by the experimenter, adjunct-questions groups spend more time on the experimental task than control groups. Therefore, it may be expected that the positive effects of adjunct questions are somewhat larger in studies that do not impose time control than in studies that do. Table IX displays the data relevant to this hypothesis. Clearly, the expected pattern is found for unrelated questions in the postquestions studies, but nowhere else in Table IX. The effect of prequestions and of postquestions on repeated test questions is greater when study time is fixed than when no time limit is imposed, $t(18) = 1.95$, $p < .10$ for prequestions, and $t(52) = 2.07$, $p < .05$ for postquestions.

The more interesting finding, however, is the effect of postquestions on unrelated questions. Postquestions have a small positive effect on unrelated test questions when time is free, $t(30) = 2.92$, $p < .01$, and a small nonsignificant negative effect when time is fixed, $t(15) = -1.69$, $p > .10$. The difference between the two types of design is also significant, $t(45) = 3.10$, $p < .01$.

TABLE VIII

Mean Effect Sizes in Postquestions Studies for Five Delay Intervals

Delay interval	Criterion test type					
	Repeated		Related		Unrelated	
	M	*n*	*M*	*n*	*M*	*n*
Immediate	.15	26	.07	30	.00	28
Less than 1 day	.14	2	—	—	.07	2
1–2 days	.19	6	.07	2	.03	3
3–7 days	.14	11	—	—	−.01	3
More than 1 week	.16	3	.00	2	−.07	2

TABLE IX

Mean Effect Size in Prequestions and Postquestions Studies with Fixed or Free Time

Study	Criterion test type					
	Repeated		Related		Unrelated	
	M	*n*	*M*	*n*	*M*	*n*
Prequestions						
Time fixed	.25	7	.15	1	−.05	7
Time free	.10	13	.07	4	−.03	15
Postquestions						
Time fixed	.19	20	.04	7	−.03	16
Time free	.14	34	.07	27	.03	31

This is an important result, because we seem to be on the track of experimental conditions that yield a consistent positive effect of postquestions on unrelated test questions. The effectiveness of postquestions when no time control is imposed, and the suggestion of a slight negative effect when there is time control, can be explained in the following way. Assume that adjunct questions have a double effect: a general effect of inciting the students to greater overall effort, and a specific effect of directing the attention of students to the textual materials covered by the adjunct questions. When time is free, both effects can come into play. This leads to an overall improvement on all types of test questions, due to the general effect, and improvement on repeated and related questions, caused by the specific effect. In fixed time designs, the effects would be different. Under the additional assumption that an increase in effort is always time-consuming (cf. Reynolds & Anderson, 1982), fixed time designs prevent the general effect from occurring. Only the specific effect can occur, causing a shift of attention from unrelated to related text information. This would explain positive effect sizes on repeated and related questions, and negative effect sizes on unrelated questions.

Interrelatedness of Design Variables

Before discussing the relationship between effect size and the four quantitative variables, some information concerning the interrelatedness of the design variables will be presented here. For this purpose the correlations between these variables were computed. This presented some difficulties, because the set of design variables includes different types of variables. *Time control* and *lookbacks* are dichotomous variables; *adjunct-question format* and *test question format* are categorical variables; *age level of subjects* and *time of testing* are best viewed as ordinal variables; and *control performance*, *text length*, *coverage*, and *distance* are measured on interval scales, but their distributions are highly skewed.

The following adjustments were made before computing correlations: The two format variables were reduced to dichotomies by using only data points in which adjunct questions and test questions were either in multiple-choice or short-answer format (this involved dropping five studies with mixed or unknown adjunct-question format and six studies with mixed or unknown test format or with free recall tests); the ordinal variables were assigned numerical values; and the logarithm of the interval variables with skewed distributions was used to avoid a disproportional influence of extreme values on results.

Table X presents the correlations between the design variables. These correlations should be interpreted with caution for two reasons. First, because the variables are of different types, the maximum possible value of some correlations in Table X is less than 1.00. This is particularly true for correlations involving dichotomous variables with an uneven distribution of observations over the two categories. The correlations of lookbacks with the other variables are most strongly affected, because the lookbacks variable has a very uneven distribution. Second, the statistical significance of these correlations cannot be established exactly, because data points derived from a single study are not statistically independent. The number of independent observations in this analysis is approximately equal to the number of studies reviewed (50), but this number is not equal for all variables, because some studies contribute more than one value and in others information on a variable may be missing. With 50 independent observations, a product-moment correlation

TABLE X
Intercorrelations of Ten Design Variables

Variable	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) *Text length*	.81	.57	−.18	.06	.42	.31	−.06	.33	−.15
(2) *Coverage*		.54	−.20	.03	.34	.16	−.16	.19	−.24
(3) *Distance*			−.11	.06	.19	.19	.04	.02	.30
(4) *Control performance*				−.51	−.74	.46	.09	−.07	.03
(5) *Adjunct question format*					.50	.10	−.02	−.17	−.09
(6) *Test questions format*						−.32	−.16	.09	−.09
(7) *Time control*							.01	−.06	−.21
(8) *Lookbacks*								−.11	.01
(9) *Subject age*									−.13
(10) *Time of testing*									

Note. For the calculation of correlations the qualitative variables were coded as follows: *adjunct questions format* and *test question format*, multiple-choice = 1, short-answer = 2; *time control*, time limit imposed = 1, no time limit = 2; *lookbacks*, not allowed = 1, allowed = 2; *subject age*, elementary school = 1, high school = 2, college = 3; *time of testing*, immediate = 1, less than 1 day = 2, 1–2 days = 3, 3–7 days = 4, more than a week = 5.

coefficient of .28 or higher is significant at the .05 level. In the interpretation of Table X, this value will be used as the borderline between significant and nonsignificant correlations.

In Table X, evidence for two subsets of interrelated design variables is found. The first subset consists of three length-related variables: text length, coverage, and distance. This is as expected. In longer texts, the average distance between adjunct questions and related material is longer than in short texts, and coverage is less dense (i.e., the number of words to each adjunct question is larger). The second subset consists of three format-related variables: adjunct-question format, test format, and control performance. The interrelatedness of these three variables was discussed at some length earlier. Of the three format-related variables, only test format is associated with length-related variables: short-answer tests occur more often in studies with long texts and low coverage. Text length is also associated with age of subjects and time control. Naturally, with younger children only short texts are used. This correlation is not higher, because the studies with older subjects include studies with short texts as well as studies with (very) long texts. As for time control, students were allotted a fixed amount of time more often in studies using short texts than in studies using long texts. Time control is also associated with test format and with control performance; in studies without time control, multiple-choice tests are used more frequently and performance of the control groups is higher than in studies with time control. Finally, time of testing and distance are significantly correlated.

Most noteworthy in Table X, however, is that the number of significant correlations is small. It can be inferred from this that it is not possible to distinguish subcategories of studies in the set of studies reviewed. It might be thought, for instance, that some studies were designed primarily for ecological validity, whereas others reflected a concern for experimental control. Ecological validity might be visible in features such as long texts, short-answer adjunct questions and tests, no time control, lookbacks allowed, and delayed testing. The absence of clear patterns

of interrelations in Table X points to the absence of these or other identifiable subcategories of studies. Moreover, it is an indication that, in general, the possible influence of any design variable on effect sizes is not confounded by the correlations of this variable with other design variables. An obvious exception to this statement is the confounding of adjunct-question format, test format, and control performance discussed earlier. A similar confounding may occur between the length-related variables discussed in the next section. But in most other cases, the design variables are independent predictors of effect size.

The Quantitative Variables

To assess the influence of the four quantitative variables, text length, control performance, coverage, and distance, correlation coefficients of these variables with effect size were computed for five of the six effect types defined by adjunct-question position and criterion test type. The number of observations on the sixth type, the effect of prequestions on related test questions, is too small (5) to yield meaningful results. For text length, coverage, and distance, the logarithms rather than the raw scores were used as before. Table XI displays the results.

Text length is negatively related to the effect of prequestions on repeated questions; the effect is larger in studies with short texts than when long texts were used. The obvious explanation of the larger effect is that the distance between the prequestions and relevant text information is generally smaller in short texts than in long texts, the adjunct questions will be less easily forgotten during reading, and relevant text information will be located with more certainty. It is therefore surprising to find that this effect is not significantly correlated with the distance variable. This is an inconsistency in our results for which there is no ready explanation. The negative correlation of this effect type with control performance fits in with the correlation of control performance with adjunct-questions format and test format, and with the influence of these design features on effect size (see Table VII).

TABLE XI

Correlations of Four Quantitative Design Variables with Effect Size

Variable	Effect type		
	Repeated	Related	Unrelated
	Prequestions studies		
	(n = 18)		(n = 20)
Text length	−.66*		.23
Coverage	.17		.22
Distance	−.22		.15
Control performance	−.51*		−.06
	Postquestions Studies		
	(n = 52)	(n = 31)	(n = 43)
Text length	.33*	.30	.48*
Coverage	.32*	.24	.36*
Distance	.20	−.14	.24
Control performance	−.49*	−.38*	.11

* $p < .05$.

There are no reliable correlations of the four quantitative variables with the effect of prequestions on unrelated questions. In a direct comparison, Frase (1968a) found that more frequent adjunct questions led to a sharper differentiation between repeated and unrelated items on the criterion test. The positive effect of prequestions and postquestions on repeated items was stronger when adjunct questions were frequent than when they were infrequent. Similarly, the negative effects of prequestions and postquestions on unrelated test questions were stronger with frequent than with infrequent adjunct questions. If this pattern held across studies, a positive correlation of length-related variables, particularly of distance, with the effect of prequestions on unrelated test questions would be expected. As the results in Table XI show, this prediction is not confirmed by the data.

The effect of postquestions on repeated questions is significantly correlated with text length, coverage, and control performance. The positive effect on repeated questions is larger in studies using long texts than in studies using short texts. This trend is the reverse of that found in prequestions studies, in which long texts were associated with smaller effects on repeated questions.

The effect of postquestions on related test questions is significantly correlated with control performance, but not with length variables.

Text length and coverage are significantly correlated with the effect of postquestions on unrelated test questions. These correlations come as a surprise. It was concluded earlier that there was no reliable general facilitative effect of postquestions, so it could be expected that effect sizes of postquestions on unrelated test questions would be randomly distributed around zero. And, of course, they should not be significantly correlated with any predictor. Even more surprising are the signs of the correlations; these indicate that the larger (positive) effect sizes occur with long texts and relatively few postquestions (high values of coverage indicate a small proportion of text information covered by postquestions). This result can be interpreted in terms of nonselective attention. With relatively few adjunct questions, learners are less likely to consider the task "done" when they have been able to answer the questions. They may perceive the adjunct questions as checks on their studying behavior, and the questions may be an incentive to study all of the text more intensively than they would have done otherwise. That is to say, the postquestions raise the overall level of processing activities (nonselective attention).

A difficulty with this interpretation is that text length is related to other design features. One of these is the time control variable whose influence was established in the previous section. The mean length of texts used in time control studies is less than half the mean text length in studies imposing no time limit. This means that the contributions of text length and time control cannot be disentangled.

Effects of Higher Order Adjunct Questions

Combining outcomes of higher order adjunct-questions studies is a much more doubtful enterprise than combining outcomes of factual adjunct-questions studies. The texts and the adjunct questions used in higher order studies are more heterogeneous than those used in factual adjunct-questions research. This was clear from the beginning, but it became even more apparent as research studies were analyzed and prepared for the review. Some higher order studies use integrative, or *comprehension*, adjunct questions. Other studies use application questions requiring the subject to recognize new examples of concepts or principles explained in the text.

Still other texts explain formulas and the adjunct questions require the subjects to use these in calculations or to explain the formula conceptually. In view of the heterogeneity of higher order adjunct questions and the small number of outcomes available, the exploration of design variables' influence on effect size will not be as exhaustive as the exploration of factual adjunct-questions effects. Only the average effect sizes of the six types of effects will be discussed.

Remember that the effect sizes in the higher order studies represent comparisons between adjunct-questions groups, and not, as in the previous section, between an adjunct-questions group and a control group. It was explained in the Method section that in higher order questions research a factual adjunct-questions group is often employed as a control group. Therefore, the difference between performance in a higher order adjunct-questions group and a factual adjunct-questions group was used as the effect size measure. Positive effect sizes indicate that the higher order group did better than the group who answered factual adjunct questions; negative effect sizes mean that the factual group outperformed the higher order group.

Table XII shows that higher order groups do better on repeated higher order questions than factual adjunct-questions groups, $t(8) = 3.85$, $p < .01$. Answering a higher order question while studying a text facilitates performance when the same question is encountered again in the criterion test. Similarly, factual adjunct questions facilitate performance on repeated factual test questions, $t(16) = -2.29$, $p < .05$. Effects in the same direction, but smaller, are found in the related test questions. Higher order adjunct-questions groups do a little better on related higher order questions in the test, $t(30) = 1.66$, $p < .10$, one-tailed. Factual adjunct-questions groups do slightly better on related factual test questions than higher order groups, $t(9) = -1.41$, $p < .10$, one-tailed.

When we turn to unrelated test questions, we find small positive effect sizes irrespective of the cognitive level of the test questions. The facilitative effect of higher order adjunct questions on unrelated higher order test questions is significant, $t(14) = 3.21$, $p < .01$. Of the 15 outcomes making up this mean, 13 are from studies in which no time constraint was imposed. Therefore, this effect could be caused by a general facilitative influence of higher order adjunct questions, similar to the influence discussed in the context of factual adjunct-questions studies. The mean effect size of unrelated factual test questions is positive but not significantly so, $t(17) = 1.54$, $p > .10$.

TABLE XII

Mean Effect Sizes of Six Effect Types Comparing Higher Order and Lower Order Adjunct Questions

Adjunct-questions group	Criterion test	*M*	*SD*	*n*
Higher order	Repeated	.15	.12	9
	Related	.04	.14	31
	Unrelated	.05	.06	15
Factual	Repeated	−.09	.15	17
	Related	−.06	.13	10
	Unrelated	.03	.09	18

Two things are to be noted in the standard deviations in Table XII. First, the pattern is similar to that in Table VI. The effects on repeated questions have higher standard deviations than the effects on unrelated questions. Second, the standard deviations in Table XII are, on the average, larger than the standard deviations in Table VI. This reflects the heterogeneity of higher order adjunct-questions studies. These rather large standard deviations should make the reader wary of too strong conclusions. The regular pattern of mean effect sizes in Table XII should not make one forget that some of these means are derived from sets of studies that include studies with large effect sizes and studies in which the effects of higher order adjunct questions were totally absent.

Discussion

This paper has presented a comprehensive review of the published literature on adjunct-questions research between 1965 and 1983. The major findings are summarized here.

1. Factual prequestions facilitate the learning of material covered either directly (repeated test questions) or indirectly (related test questions) by these adjunct questions. Factual prequestions have a negative effect on the learning of text material unrelated to them.

2. Factual postquestions facilitate the learning of material covered either directly or indirectly by them; they have no general positive or negative effect on the learning of text material unrelated to them.

3. Factual adjunct questions in short-answer format have stronger facilitative effects than adjunct questions in multiple-choice format. This is due in part to the adjunct questions themselves and in part to the measurement of effects of short-answer adjunct questions by short-answer tests, and the effects of multiple-choice adjunct questions by multiple-choice tests. The effect of the format of adjunct questions could be explained by theories of recognition memory, suggesting that exposure to incorrect multiple-choice options would mitigate the adjunct-questions effect. The influence of test format may be a ceiling effect, because performance of control groups on multiple-choice tests is considerably higher than control performance on short-answer tests.

4. The effect of factual prequestions and postquestions on repeated test questions is larger when study time is experimentally controlled than when time is free. When time is experimentally controlled, the effect of factual postquestions on the learning of unrelated text material is negative; when time is free, this effect is positive. This result establishes time control as a major design feature that may determine not only the size of adjunct-questions effects, but the way in which the pattern of learners' processing activities is changed by adjunct questions.

5. The positive and negative effects of prequestions are less pronounced in studies employing long texts than in studies using short texts. For postquestions the reverse trend was observed. Postquestion studies with long texts produce large positive effects on repeated test questions and small positive effects on unrelated test questions. When texts are short, postquestions generally produce slightly negative effects on unrelated test questions.

6. The pattern of results of factual adjunct questions is confounded by high intercorrelations between some design variables. Some notable clusters of interrelated variables are (a) adjunct-question format, test-question format, and control

performance, (b) text length, coverage, and distance, and (c) text length and time control. These interrelations obscure the mechanisms producing variations in the strength of the adjunct-questions effects.

7. A number of design variables could not be shown to be influential in determining the size of adjunct-questions effects. These include the interval between studying and the test, and whether or not the subjects were allowed to review the text while answering postquestions. The inconclusiveness of the review with respect to these variables may be due to the scarcity of studies with delayed testing and of studies that allowed lookbacks.

8. When higher order and factual adjunct questions are compared, higher order adjunct questions emerge as superior on repeated, related, and unrelated higher order test questions. There is a trend that performance on unrelated factual test questions is similarly facilitated by higher order adjunct questions. From the same comparison, factual adjunct questions emerge as superior on repeated and related factual test questions. These results indicate that higher order questions may have a somewhat broader general facilitative effect than factual adjunct questions.

Is it possible to develop a united theoretical framework in which these results can be fitted? The answer to this question necessarily reflects the author's conviction rather than objectively verifiable facts. Nonetheless, a few remarks about theory should be made. Theorizing about adjunct questions has two tasks. One task is to predict *under what conditions* adjunct questions are effective or ineffective; the other task is to explain *how* adjunct questions are effective when they are effective (cf. Wright, 1978).

The evidence in this review provides information relevant to the *when* question that may be helpful in pointing out directions for future research. With respect to the effect of postquestions on unrelated test questions, the review has pointed out some conditions that might be favorable to this effect. Specifically, the effect would be expected when long texts containing a relatively small number of adjunct questions are used, and when no limits are set on the amount of time to study the text. It is suggested that under these conditions learners are likely to adopt studying the text as their primary task, rather than answering the adjunct questions. Another way of saying this is that designs with long texts, few questions and no time limits are ecologically valid. These are conditions that obtain in normal study situations.

The relevance of this review for the *when* question can be illustrated in another way by an example. The example is provided by a recent study (Duchastel & Nungester, 1984), not included in the review data base because it was published after 1982. Duchastel and Nungester used a 1700-word history passage for which 12 adjunct questions were written. In one treatment these questions were inserted after each of the 12 paragraphs making up the text. In another treatment the adjunct questions were massed postquestions. A third group read a text version without adjunct questions. Review of the materials was permitted after an adjunct question was encountered. All students were given 20 min to study the text, and they were tested 2 weeks later. The criterion test contained 24 questions, 12 repeated questions and 12 unrelated new questions. On the repeated questions, the inserted questions group scored a little better ($M = 6.8$ out of a maximum of 12) than the massed questions group ($M = 6.1$), and both scored better than the control group ($M = 4.6$). On the unrelated questions, the control group ($M = 4.3$) did slightly better than the inserted and massed questions groups ($M = 4.0$ and $M = 4.1$,

respectively). Duchastel and Nungester ascribe the absence of a facilitative effect on unrelated questions to their allowing their students lookbacks while answering the adjunct questions.

The present review leads to a different interpretation of Duchastel and Nungester's results. According to the review, crucial variables in their design are the length of the text and the experimenter-controlled time. Relatively short texts in combination with time control were shown to be associated with slightly negative effects of adjunct questions on unrelated test items. This leads to two conclusions. First, Duchastel and Nungester's findings could have been predicted from the review. The mean effect size of postquestions on unrelated questions in fixed time studies is −.03. The negative effect found by Duchastel and Nungester, averaged over the two adjunct-questions treatments, equals −.02. The mean effect of postquestions on repeated questions in fixed time studies is .19. Duchastel and Nungester found an effect size of .15. Second, the review suggests that Duchastel and Nungester may be seriously misguided in their explanation of the negative effect on unrelated questions. They ascribe this effect to their allowing their subjects to review the text while answering the adjunct questions. Their design does not enable them to compare lookback conditions directly with non-lookback conditions, so the evidence for their causal attribution is weak. The present review suggests that the crucial design feature causing the negative effect of postquestions on unrelated test questions is setting a fixed amount of time for the students to study the text.

The technique of the integrative review has become a standard methodological tool in recent years. The value of the technique is not disputed. In the author's view, the most important contributions of this review are the demonstration that the widely accepted general facilitative effect of adjunct questions is not general at all, and the provision of a basis for the interpretation of adjunct-questions studies and the refinement of theories about their effects.

References

Ackerman, A. S. (1981). Hocus-pocus, imagery in focus: Implications for individualized instruction. *Journal of College Science Teaching, 11*, 32–34.

Adejumo, D. (1980). Study strategy and type of test item in prose comprehension test. *Psychological Reports, 47*, 883–886.

Allen, D. I. (1970). Some effects of advance organizers and level of question on the learning and retention of written social studies material. *Journal of Educational Psychology, 61*, 333–339.

Anderson, R. C. (1967). Educational psychology. In P. R. Farnsworth, O. McNemar, & Q. McNemar (Eds.), *Annual review of psychology* (Vol. 18, pp. 129–164). Palo Alto, CA: Annual Reviews.

Anderson, R. C., & Biddle, W. B. (1975). On asking people questions about what they are reading. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9, pp. 89–132). New York: Academic Press.

Andre, T. (1979). Does answering higher-level questions while reading facilitate productive learning? *Review of Educational Research, 49*, 280–318.

Andre, T., Mueller, C., Womack, S., Smid, K., & Tuttle, M. (1980). Adjunct application questions facilitate later application, or do they? *Journal of Educational Psychology, 72*, 533–543.

Andre, T., & Womack, S. (1978). Verbatim and paraphrased adjunct questions and learning from prose. *Journal of Educational Psychology, 70*, 796–802.

Bing, S. (1982). Role of adjunct questions and reading ability levels on rote and conceptual learning from prose. *Instructional Science, 11*, 129–138.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16.

Boker, J. R. (1974). Immediate and delayed retention effects of interspersing questions in written instructional passages. *Journal of Educational Psychology, 66*, 96–98.

Boyd, W. M. (1973). Repeating questions in prose learning. *Journal of Educational Psychology, 64*, 31–38.

Bull, S. G., & Dizney, H. F. (1973). Epistemic-curiosity-arousing prequestions: Their effect on long-term retention. *Journal of Educational Psychology, 65*, 45–49.

Carrier, C. A., & Fautsch-Patridge, T. (1981). Levels of questions: A framework for the exploration of processing activities. *Contemporary Educational Psychology, 6*, 365–382.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Cooper, H. M. (1982). Scientific guidelines for conducting integrative research reviews. *Review of Educational Research, 52*, 291–302.

Duchastel, P. C. (1979). Retention of prose materials: The effect of testing. *Journal of Educational Research, 72*, 299–300.

Duchastel, P. C. (1981). Retention of prose following testing with different types of tests. *Contemporary Educational Psychology, 6*, 217–226.

Duchastel, P. C. (1983). Interpreting adjunct question research: Processes and ecological validity. *Human Learning, 2*, 1–5.

Duchastel, P. C., & Nungester, R. J. (1984). Adjunct question effects with review. *Contemporary Educational Psychology, 9*, 97–103.

Eischens, R. R., Gaite, A. J. H., & Kumar, V. K. (1972). Prose learning: Effects of questions position and informational load interactions on retention of low signal value information. *Journal of Psychology, 81*, 7–12.

Ellis, J. A., Wulfeck, W. H., & Montague, W. E. (1980). The effect of adjunct and test question similarity on study behavior and learning in a training course. *American Educational Research Journal, 17*, 449–457.

Faw, H. W., & Waller, T. G. (1976). Mathemagenic behaviours and efficiency in learning from prose materials. Review, critique and recommendations. *Review of Educational Research, 46*, 691–720.

Felker, D. B., & Dapra, R. A. (1975). Effects of question type and question placement on problem solving ability from prose material. *Journal of Educational Psychology, 67*, 380–384.

Frase, L. T. (1967). Learning from prose material: Length of passage, knowledge of results and position of questions. *Journal of Educational Psychology, 58*, 266–272.

Frase, L. T. (1968a). Effect of question location, pacing, and mode upon retention of prose material. *Journal of Educational Psychology, 59*, 244–249.

Frase, L. T. (1968b). Questions as aids to reading: Some research and theory. *American Educational Research Journal, 5*, 319–332.

Frase, L. T. (1969). Structural analysis of the knowledge that results from thinking about text. *Journal of Educational Psychology Monograph, 60* (6, Pt. 2).

Frase, L. T. (1970). Influence of sentence order and amount of higher level text processing upon reproductive and productive memory. *American Educational Research Journal, 7*, 307–319.

Frase, L. T. (1971). Effect of incentive variables and type of adjunct questions upon text learning. *Journal of Educational Psychology, 62*, 371–375.

Frase, L. T., Patrick, E., & Schumer, H. (1970). Effect of question position and frequency upon learning from text under different levels of incentive. *Journal of Educational Psychology, 61*, 52–56.

Frase, L. T., & Schwartz, B. J. (1975). Effect of question production and answering on prose recall. *Journal of Educational Psychology, 67*, 628–635.

Friedman, F., & Rickards, J. P. (1981). Effect of level, review, and sequence of inserted questions on text processing. *Journal of Educational Psychology, 73*, 427–436.

Gagne, E. D., & Memory, D. (1978). Instructional events and comprehension: Generalization across passages. *Journal of Reading Behavior, 10*, 321–335.

Gagne, R. M., & Rohwer, W. D., Jr. (1969). Instructional psychology. In P. H. Mussen & M. R. Rosenzweig (Eds.), *Annual review of psychology* (Vol. 20, pp. 381–418). Palo Alto, CA: Annual Reviews.

Glaser, R., & Resnick, L. B. (1972). Instructional psychology. In P. H. Mussen & M. R. Rosenzweig (Eds.), *Annual review of psychology* (Vol. 23, pp. 207–276). Palo Alto, CA: Annual Reviews.

Glass, G. V, McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Grant, L., Keenan, J. B., & Hursh, D. (1980). The effects of study questions, the SQ3R system of studying, and reading and rereading on academic performance. *Journal of Personalized Instruction, 4*, 142–147.

Graves, M. F., & Clark, D. L. (1981). The effect of adjunct questions on high school low achievers' reading comprehension. *Reading Improvement, 18*, 8–13.

Green, B. F., & Hall, J. A. (1984). Quantitative methods for literature reviews. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual Review of Psychology* (Vol. 35, pp. 37–53). Palo Alto, CA: Annual Reviews.

Gustafson, H. W., & Toole, D. L. (1970) Effects of adjunct questions, pretesting, and degree of student supervision on learning from an instructional text. *Journal of Experimental Education, 39*, 53–58.

Hiller, J. H. (1974). Learning from prose text: Effects of readability level, inserted question difficulty, and individual differences. *Journal of Educational Psychology, 66*, 202–211.

Hunkins, F. P. (1969). Effects of analysis and evaluation questions on various levels of achievement. *Journal of Experimental Education, 38*, 45–58.

Koran, M. L., & Koran, J. J. (1975). Interaction of learner aptitude with question pacing in learning from prose. *Journal of Educational Psychology, 67*, 76–82.

LaPorte, R. E., & Voss, J. F. (1975). Retention of prose materials as a function of postacquisition testing. *Journal of Educational Psychology, 67*, 259–266.

Mayer, R. E. (1975). Forward transfer of different reading strategies evoked by testlike events in mathematics text. *Journal of Educational Psychology, 67*, 165–169.

McConkie, G. W., Rayner, K., & Wilsons, S. J. (1973). Experimental manipulation of reading strategies. *Journal of Educational Psychology, 65*, 1–8.

McGaw, B., & Grotelueschen, A. (1972). Direction of the effect of questions in prose material. *Journal of Educational Psychology, 63*, 580–588.

McKeachie, J. (1974). Instructional psychology. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. 25, pp. 161–193). Palo Alto, CA: Annual Reviews.

McKenzie, G. R. (1972). Some effects of frequent quizzes on inferential thinking. *American Educational Research Journal, 9*, 231–240.

Memory, D. M. (1981). An extended comparison of adjunct aids with low, average and good readers in the sixth grade. *Journal of Educational Research, 74*, 405–410.

Natkin, G., & Stahler, E. (1969). The effects of adjunct questions on short and long-term recall of prose materials. *American Educational Research Journal, 6*, 425–432.

Peeck, J. (1970). Effect of prequestions on delayed retention of prose material. *Journal of Educational Psychology, 61*, 241–246.

Prosser, G. V. (1974). The role of active questions in learning and retention of prose material. *Instructional Science, 2*, 451–476.

Resnick, L. B. (1981). Instructional psychology. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. 32, pp. 659–704). Palo Alto, CA: Annual Reviews.

Reynolds, R. E., & Anderson, R. C. (1982). Influence of questions on the allocation of attention during reading. *Journal of Educational Psychology, 74*, 623–632.

Reynolds, R. E., Standiford, S. N., & Anderson, R. C. (1979). Distribution of reading time when questions are asked about a restricted category of text information. *Journal of Educational Psychology, 71*, 183–190.

Rickards, J.P. (1976a). Interaction of position and conceptual level of adjunct questions on immediate and delayed retention of text. *Journal of Educational Psychology, 68*, 210–217.

Rickards, J.P. (1976b). Stimulating high-level comprehension by interspersing questions in text passages. *Educational Technology, 9*, 13–17.

Rickards, J. P. (1979). Adjunct postquestions in text: A critical review of methods and processes. *Review of Educational Research, 49*, 181–196.

Rickards, J. P., Anderson, M. C., & McCormick, C. B. (1976). Processing effects of common-word and number questions inserted in reading materials. *Journal of Educational Research, 69*, 274–277.

Rickards, J. P., & Denner, P. R. (1979). Depressive effects of underlining and adjunct questions on children's recall of text. *Instructional Science, 8*, 81–90.

Rickards, J. P., & Di Vesta, F. J. (1974). Type and frequency of questions in processing textual material. *Journal of Educational Psychology, 66*, 354–362.

Rickards, J. P., & Hatcher, C. W. (1977–1978). Interspersed meaningful learning questions as semantic cues for poor comprehenders. *Reading Research Quarterly, 13*, 538–553.

Rothkopf, E. Z. (1965). Some theoretical and experimental approaches to problems in written instruction. In J. D. Krumboltz (Ed.), *Learning and the educational process* (pp. 193–221). Chicago: Rand McNally.

Rothkopf, E. Z. (1966). Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal, 3*, 241–249.

Rothkopf, E. Z. (1972). Variable adjunct question schedules, interpersonal interaction, and incidental learning from written material. *Journal of Educational Psychology, 63*, 87–92.

Rothkopf, E. Z., & Bisbicos, E. E. (1967). Selective facilitative effects of interspersed questions on learning from written materials. *Journal of Educational Psychology, 58*, 56–61.

Rothkopf, E. Z., & Bloom, R. D. (1970). Effects of interpersonal interaction on the instructional value of adjunct questions in learning from written material. *Journal of Educational Psychology, 61*, 417–422.

Sagaria, S. D., & Di Vesta, F. J. (1978). Learner expectations induced by adjunct questions and the retrieval of intentional and incidental information. *Journal of Educational Psychology, 70*, 280–288.

Samuels, S. J. (1969). The effect of post-test relevant pre-tests and discussion-type feedback on learning and retention. *Psychonomic Science, 16*, 67–68.

Sanders, J. R. (1973). Retention effects of adjunct questions in written and aural discourse. *Journal of Educational Psychology, 65*, 181–186.

Santiesteban, A. J., & Koran, J. J. (1977). Instructional adjuncts and learning science from written materials. *Journal of Research in Science Teaching, 14*, 51–55.

Snowman, J., & Cunningham, D. J. (1975). A comparison of pictorial and written adjunct aids in learning from text. *Journal of Educational Psychology, 67*, 307–311.

Swenson, I., & Kulhavy, R. W. (1974). Adjunct questions and the comprehension of prose by children. *Journal of Educational Psychology, 66*, 212–215.

Tergan, S. O. (1979). Der Einfluss von Textverstaendlichkeit und Orientierungshinweisen auf den Lernerfölg von Funkkollegiaten [The influence of text comprehensibility and orienting instructions on learning performance of radio college students]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 11*, 67–76.

Walker, B. S. (1974). Effects of inserted questions on retroactive inhibition in meaningful verbal learning. *Journal of Educational Psychology, 66*, 486–490.

Watts, G. H. (1973). The "arousal" effect of adjunct questions on recall from prose materials. *Australian Journal of Psychology, 25*, 81–87.

Watts, G. H. (1974). Effects of prequestions on control of attention in written instruction. *Australian Journal of Education, 18*, 79–85.

Watts, G. H. (1975). The effect of adjunct questions on learning from written instruction by students from different achievement levels. *Australian Journal of Education, 19*, 78–87.

Watts, G. H., & Anderson, R. C. (1971). Effects of three types of inserted questions on learning from prose. *Journal of Educational Psychology, 62*, 387–394.

Wiesendanger, K. D., & Wollenberg, J. P. (1978). Prequestioning inhibits third graders' reading comprehension. *Reading Teacher, 31*, 892–895.

Wilson, M. M. (1979). The processing strategies of average and below average readers answering factual and inferential questions on three equivalent passages. *Journal of Reading Behavior, 11*, 235–245.

Winne, P. M. (1979). Experiments relating teachers' use of higher cognitive questions to student achievement. *Review of Educational Research, 49*, 13–50.

Wittrock, M. C., & Lumsdaine, A. A. (1977). Instructional psychology. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. 28, pp. 417–459). Palo Alto, CA: Annual Reviews.

Wright, P. (1978). Feeding the information eaters: Suggestions for integrating pure and applied research on language comprehension. *Instructional Science, 7*, 249–312.

Yost, M., Avila, L., & Vexler, E. B. (1977). Effect on learning of postinstructional responses to questions of differing degrees of complexity. *Journal of Educational Psychology, 69*, 399–408.

AUTHOR

CHRISTIAAN HAMAKER, Associate Professor, Developmental Psychology, University of Amsterdam, Weesperplein 8, 1018 XA, Amsterdam, The Netherlands. *Specializations:* Cognitive development, mathematics learning.