

Predicting performance on text: Delayed versus immediate predictions and tests

RUTH H. MAKI

North Dakota State University, Fargo, North Dakota

Delayed judgments of learning for word pairs are more accurate than immediate judgments of learning when the memory test is delayed. In the present experiment, I investigated a similar paradigm with text. Participants predicted performance on texts either immediately after reading the texts or after a delay following the reading of other texts, and tests were given either immediately or after a delay. Immediate ratings with an immediate test produced the most accurate predictions, and immediate ratings with a delayed test produced less accurate predictions. Delaying both the ratings and the test did not produce more accurate predictions than immediate predictions and a delayed test. The results for delayed judgments of learning with text were different from those with word pairs.

A major aspect of metacognition is the ability to predict future performance. Nelson and Dunlosky (1991) reported that participants are very accurate at predicting future performance on word pairs tested after a delay if predictions are also made following a short delay. Participants are much less accurate if predictions are made within a few seconds after study. Nelson and Dunlosky coined the term *delayed judgment of learning (JOL) effect* for the superiority of delayed over immediate judgments. Dunlosky and Nelson (1992) offered two possible explanations for the delayed-JOL effect. According to the two-store hypothesis, participants monitor both short-term (or a conscious working memory buffer) and long-term memory in the immediate prediction case, but they monitor long-term memory alone in the delayed case. Monitoring from long-term memory is more appropriate than monitoring from short-term memory because the memory tests are delayed. The second explanation involves transfer-appropriate monitoring (Dunlosky & Nelson, 1992). The conditions of retrieval must be similar to the conditions of the test for accurate monitoring to occur. A study by Begg, Duft, Lalonde, Melnick, and Sanvito (1989) supported the importance of transfer-appropriate processing for JOLs, but a recent study by Dunlosky and Nelson (1997) revealed that making JOLs to a stimulus-and-response pair did not produce more accuracy in recognition than making JOLs to the stimulus alone, although the pair was more similar to the recognition test. Dunlosky and Nelson (1997) argued that the transfer-appropriate monitoring hypothesis was not supported by their results, although they allowed that other retrieval monitoring hypotheses were viable.

Although the Nelson and Dunlosky studies (1991; Dunlosky & Nelson, 1992, 1997) have shown very high accuracy when predictions for responses from word lists were delayed, these results may not be relevant to a different metacognitive situation—namely, monitoring learning from text. Early research (Glenberg & Epstein, 1985; Glenberg, Sanocki, Epstein, & Morris, 1987) suggested that students cannot accurately predict future performance on text, but more recent studies have shown that they can predict with greater-than-chance accuracy (e.g., Maki, 1995; Maki & Serra, 1992; Weaver & Bryant, 1995) when more than one test item for each text is used (Weaver, 1990) to increase test reliability.

Many studies of predictions with text have been based on a paradigm that is similar to Nelson and Dunlosky's (1991) immediate-JOL paradigm. In such studies, prediction ratings (Maki & Serra, 1992) or comprehension ratings (Weaver & Bryant, 1995) are made immediately after reading, but the tests are delayed until all texts have been read. Although predictions are made immediately after a short section of text has been read, the time course differs from that of JOLs for word pairs. Predictions for individual text propositions are not made within a few seconds of study, as are immediate JOLs for word pairs. However, the procedures with text parallel those with immediate JOLs in that the delay between reading and predictions is shorter than the delay between reading and the test. In contrast, Glenberg and Epstein (1985; Glenberg et al., 1987) have often used a delayed procedure in which both the predictions and the tests were delayed (similar to Nelson & Dunlosky's, 1991, delayed-JOL procedure). Glenberg et al. compared this delayed procedure with an immediate procedure in their Experiments 1 and 2. In both experiments, half of the participants read texts, made immediate predictions, and took immediate tests. The other half of the participants read all texts first and then predicted performance from each text's title and immediately answered questions from each text. In their Experiment 1, Glenberg et al. used single-

Thanks to Dawn Jonas, Mary Kallod, Michelle Marek, and Michelle Schumacher for testing the participants in this experiment. R.H.M. is now at Texas Tech University. Correspondence should be addressed to R. H. Maki, Department of Psychology, Texas Tech University, Lubbock, TX 79409-2051 (e-mail: rumaki@ttu.edu).

—Accepted by previous editor, Geoffrey R. Loftus

Table 1
Sequence of Events in the Four Delay Conditions

P_iT_i	P_iT_d	P_dIT_d	P_dNIT_d
Read Text 1	Read Text 1	Read Text 1	Read Text 1
Predict Text 1	Predict Text 1	Read Text 2	Read Text 2
Test over Text 1	Read Text 2	Read Text 3	Read Text 3
Confidence, Text 1	Predict Text 2	[Read Texts 4–6]	[Read Texts 4–6]
Read Text 2	Read Text 3	Predict Text 1	Predict Text 1
Predict Text 2	Predict Text 3	Test over Text 1	Predict Text 2
Test over Text 2	[Read, Predict Texts 4–6]	Confidence, Text 1	Predict Text 3
Confidence, Text 2	Test over Text 1	Predict Text 2	[Predict Texts 4–6]
Read Text 3	Confidence, Text 1	Test over Text 2	Test over Text 1
Predict Text 3	Test over Text 2	Confidence, Text 2	Confidence, Text 1
Test over Text 3	Confidence, Text 2	Predict Text 3	Test over Text 2
Confidence, Text 3	Test over Text 3	Test over Text 3	Confidence, Text 2
	Confidence, Text 3	Confidence, Text 3	Test over Text 3
			Confidence, Text 3
[Read, Predict, Test, Confidence Texts 4–6]	[Test, Confidence Texts 4–6]	[Predict, Test, Confidence Texts 4–6]	[Test, Confidence Texts 4–6]

Note— P_iT_i , immediate predictions, immediate tests; P_iT_d , immediate predictions, delayed tests; P_dIT_d , delayed predictions, delayed intervening tests; P_dNIT_d , delayed predictions, delayed nonintervening tests.

item verbatim recognition and inference tests, and in Experiment 2, they used four-item idea recognition tests. In their Experiment 1, there was no relation between predictions and performance, and delay had no effect. However, in Experiment 2, the group that made immediate predictions followed by an immediate test showed a significant correlation between predictions and performance, but the group that made delayed predictions did not. This is one of the few significant relationships found by Glenberg and Epstein in their series of studies (Glenberg & Epstein, 1985; Glenberg et al., 1987). This unique finding may have occurred because the test for each text featured multiple questions, a condition necessary for showing accurate predictions (Weaver, 1990), and because the predictions and the test occurred immediately after reading.

The results of Glenberg et al.'s (1987) Experiment 2 suggest that the timing of the study, predictions, and tests over texts may be important, just as it is for word pairs. In order to examine this variable, I manipulated the delay between reading, predictions, and the test. In one condition, a prediction rating was made immediately after a short text had been studied and a multiple-choice test followed immediately (P_iT_i , designating immediate predictions and an immediate test). In a second condition, a prediction rating was made immediately after study, but the test was delayed (P_iT_d). The P_iT_d condition differed from the P_iT_i condition both in terms of the timing between predictions and the test (immediate or delayed) and in terms of the presence of intervening tests between predictions (present for P_iT_i but not for P_iT_d). In order to investigate the role of intervening tests while keeping timing constant, two delayed conditions were tested. In both of these, all texts were read and then predictions were made on the basis of text titles and short descriptions. The two delayed conditions differed in that multiple-choice tests followed each prediction in the delayed intervening tests condition (P_dIT_d), but the multiple-choice tests were not taken until all predictions had been made in the delayed non-

intervening tests condition (P_dNIT_d). Table 1 shows the procedures in these four conditions.

The P_iT_d condition was most similar to Nelson and Dunlosky's (1991) immediate-JOL condition, and the P_dNIT_d condition was most similar to Nelson and Dunlosky's delayed-JOL condition. The P_iT_i condition was not possible in Nelson and Dunlosky's study because an immediate test of a single word pair would surely produce ceiling effects. However, Glenberg et al. (1987) found that predictions for text were more accurate than chance in this type of immediate condition, but predictions were not more accurate than chance in other conditions.

I also investigated posttest confidence judgments. Such posttest assessments of the accuracy of test performance on multiple-choice tests are typically higher than predictions (e.g., Maki & Serra, 1992), probably because participants know the incorrect alternatives after the test but not before. Posttest confidence judgments should not be affected by the delay between study and predictions, as long as the confidence judgments are made immediately after the test questions have been answered. Delaying tests, however, may produce lower judgment accuracy because participants may be less able to assess the match between the text representation and their selected answer.

METHOD

Participants

A total of 96 volunteers from North Dakota State University participated for extra credit in introductory psychology classes. Twenty-four participants were randomly assigned to each of the four groups. Eight of the original participants gave the same prediction rating to all texts, so they were replaced. Four were in the P_dIT_d condition; 2 were in the P_dNIT_d condition; 1 was in the P_iT_i condition; and 1 was in the P_iT_d condition.

Materials

Twelve 400-word texts of varying difficulty were taken from a timed reading series (Spargo, 1989). Reading difficulties as measured by the Flesch scale ranged from 42.2 to 77.0 with a mean of

Table 2
Median Ratings and Ranges for the Four Delay Conditions

Condition	Predictions		Confidence Judgments	
	Median	Range	Median	Range
Immediate predictions, immediate tests (P_iT_i)	4.00	2.0–5.5	4.00	2.0–6.0
Immediate predictions, delayed tests (P_iT_d)	4.00	2.5–6.0	4.00	3.0–6.0
Delayed predictions, delayed intervening tests (P_dIT_d)	3.50	2.5–5.0	3.50	2.0–5.5
Delayed predictions, delayed non intervening tests (P_dNIT_d)	3.75	2.5–5.0	4.00	1.0–5.0

62.12 (100 is the easiest). An example of medium-difficulty text is shown in the Appendix. The texts were divided into two sets, with each participant reading either one set or the other. The titles of the texts in each set are shown in the Appendix. Each text was divided in half with about 200 words in each half. Five multiple-choice questions that tapped the second half of each text were taken from the reading series, although the three-alternative questions from the series were modified so that each question had four, rather than three, alternatives. Half of the questions tested recall of presented facts and the other half were more integrative and designed to test comprehension.

Procedure

All passages were presented individually to participants on microcomputers that were programmed with Micro Experimental Laboratory (MEL; Schneider, 1988). Participants read the texts at their own pace by pressing the space bar to erase one sentence and present the next sentence in the middle of the screen. Because the texts in the first experiment were only about 200 words in length, the texts in this experiment were divided in half for predictions. It seemed possible that making immediate predictions over 400 words might reduce accuracy relative to making predictions over 200 words because of the longer delay between the beginning of the text and the prediction. Predictions over the first half of the text were made in response to the query, "How well will you do on multiple-choice questions on the section that you just read?" (1 = *very well*; 6 = *very poorly*). Then, all participants read the second half of the text.

Predictions concerning performance on the second halves were either made immediately after reading or they were delayed until all six texts had been read. In the P_iT_i condition, participants answered the prediction question immediately after reading the second half of each text. They then answered the five test questions for the text and made the confidence judgment by responding to the question, "How well do you think you did on the questions that you just answered?" on a scale ranging from 1 (*very well*) to 6 (*very poorly*). This procedure was repeated until participants had read, predicted, answered the test questions, and made confidence judgments for all six texts. In the P_iT_d condition, participants read each text and made predictions for the first and second halves. After they had read all six texts, they answered the five questions for each text and gave a confidence judgment following the questions for each text. In the delayed rating conditions, participants read all texts and made predictions on the first halves immediately after reading. However, they did not make predictions on the second halves until they had read all six texts. Then, the title of each text was given along with a short one-sentence description of the topic (shown in the Appendix). Participants were asked to predict their performance over the second half of the text using the same 6-point scale as in the immediate ratings conditions. In the P_dIT_d condition, participants answered the five multiple-choice questions for each text immediately following each prediction rating. In the P_dNIT_d condition, all six predictions were made first, followed by the multiple-choice questions for each text. In all conditions, the confidence judgments were made immediately following the five test questions for each text.

RESULTS

Proportion Correct

Mean proportions correct on the multiple-choice tests were .76 ($SE = .02$) for the P_iT_i group, .71 ($SE = .03$) for the P_iT_d group, .67 ($SE = .04$) for the P_dNIT_d group, and .69 for the P_dIT_d group ($SE = .03$). These differences were not significant [$F(3,92) = 1.54$, $MS_e = .02$, $p > .05$ (the significance level used for all analyses)]. Thus, the short delays experienced by participants in this experiment did not influence memory for the text as measured by the multiple-choice questions.

Median Ratings

Ratings were converted in such a way that higher numbers indicate more confidence. Median prediction ratings and posttest confidence judgments for the four delay groups can be seen in Table 2. A Kruskal–Wallis test showed that the groups' median prediction ratings did not differ [$\chi^2(3) = 3.30$]. A Kruskal–Wallis test for the median posttest confidence judgments did not reach a conventional level of significance, although it did indicate a trend [$\chi^2(3) = 6.71$, $p < .10$]. The group that made delayed predictions with intervening test experience tended to give lower ratings than the other groups.

Rating and Performance Relationships

Each participant's prediction ratings and posttest confidence judgments for each text were correlated separately with performance on the questions from each text. Contingency tables defined by predictions from 1 to 6 and numbers correct from 0 to 5 were set up for prediction and confidence judgments for each participant. The number of texts falling into each of these cells was counted. Correlations were calculated from these contingency tables using the Goodman–Kruskal (Goodman & Kruskal, 1954) gamma, which is a nonparametric correlation coefficient that can vary from -1.0 for a perfect negative relationship to $+1.0$ for a perfect positive relationship. Nelson (1984) has argued that gamma is the appropriate correlation to use for discrete ordinal data such as the ratings in this experiment. Participants who predict performance accurately or who are accurate in making posttest confidence judgments should have high, positive gammas.

Mean gamma correlations between prediction ratings and performance and between confidence judgments and performance are presented in Figure 1. Single-sample t

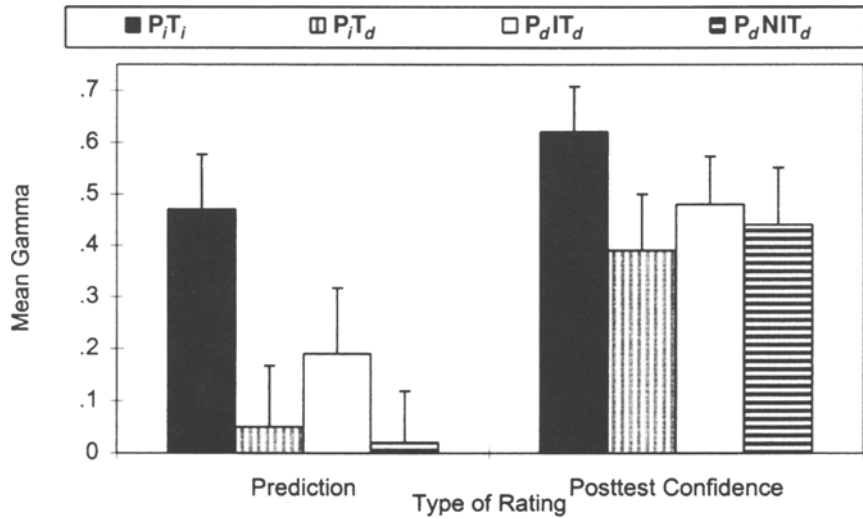


Figure 1. Mean prediction rating and posttest confidence gammas with SEMs for the four delay conditions.

tests showed that prediction accuracy was significantly greater than zero only in the P_iT_i condition. The mean gammas in the P_iT_d condition and the two delay conditions did not differ reliably from zero. In contrast, posttest confidence judgment accuracy was significantly greater than zero in all conditions.

An analysis of variance (ANOVA) showed that the prediction gammas for the four groups differed significantly [$F(3,92) = 3.34, MS_e = .307$]. Post hoc pairwise comparisons (both with and without a Tukey correction) showed that the P_iT_i group made more accurate predictions than the P_iT_d group and that the P_iT_i group predicted more accurately than the P_dNIT_d group. The P_dIT_d group fell between these two sets of means, but its mean gamma was not different from that of the P_iT_i group or from those of the two groups that had no intervening tests (P_iT_d and P_dNIT_d). An ANOVA of the posttest confidence gammas showed no differences among the groups ($F < 1$).

DISCUSSION

The P_iT_i condition, in which participants made predictions after reading a short section of text and immediately taking the multiple-choice test, yielded more accurate predictions than the delayed test conditions. When the test was delayed, delayed predictions were not more accurate than immediate predictions, as would be expected if a delayed-JOL effect (Nelson & Dunlosky, 1991) occurred with text. However, the definition of *immediate* was defined differently here from that in the delayed-JOL paradigm. With word pairs, immediate judgments are made within a few seconds of original study. With text, only the last proposition can fall within the few-second time period before the prediction is made. In the present experiment, the time period between most of the predic-

tions and the prediction was greater than 30 sec. Thus, the timing in the text paradigm is such that a traditional short-term memory buffer would not be involved in either immediate or delayed predictions. In addition, retrieval of short texts is much more complex than retrieval of a single word. It would be unlikely that participants could retrieve all of the propositions in a text when they make a prediction rating.

Immediate predictions matched performance on immediate tests better than they matched performance on delayed tests. In both the P_iT_i condition and the P_iT_d condition, predictions were made immediately after reading, so the information accessed in making predictions should have been similar. However, students in the pure immediate condition had test experience between predictions, whereas students in the P_iT_d condition did not. It is not possible to eliminate test experience from a pure immediate condition because delaying the test until all ratings are made would transform the P_iT_i condition into the P_iT_d condition. Thus, the presence of intervening test experience was manipulated with delayed ratings that followed the reading of all texts. The P_dNIT_d condition produced prediction accuracy at the same level as the P_iT_d condition, and this delayed condition produced prediction accuracy that was significantly lower than that of the P_iT_i condition. Accuracy in the delayed ratings condition with test experience, however, was intermediate and not significantly different from accuracy in either the P_iT_i condition or the P_iT_d condition. Thus, test experience may have played some role in the higher accuracy of the P_iT_i condition relative to the P_iT_d condition.

In addition to test experience, the P_iT_i and the P_iT_d conditions differed in terms of the timing of the test—either immediately after the ratings or after all texts had been read and rated. There are several possible explanations for why immediate predictions with immediate tests

were the most accurate. One possibility, as pointed out by Nelson (1996), is that the processes involved in acquisition are maximally available immediately after learning. Thus, students may have accessed information about the difficulty of processing the text. However, students in both the P_iT_i and the P_iT_d conditions would have been equally able to access processing information. For this explanation to be viable, assessments of processing difficulty would need to correlate better with immediate than with delayed tests.

Another possibility is that text propositions are retrieved, an explanation that might take the form of the accessibility model proposed by Koriat (1993) for feeling-of-knowing (FOK) judgments made when participants predict recognition after a recall attempt. According to Koriat's model, participants base their FOK judgments on how much total information about an item can be retrieved, with more confident judgments being given with more retrieved information. Such retrieval-based accounts are relevant to text: Morris (1990) found that the amount of text retrieved on a speeded immediate test correlated with predictions of later memory performance. The total amount retrieved may correlate better with an immediate test than with a delayed test.

A variation on the retrieval account is similar to Dunlosky and Nelson's (1992) two-store hypothesis of the delayed-JOL effect. Propositions from the recently created text representation are probably activated in short-term working memory, and they may also be activated in long-term working memory (Ericsson & Kintsch, 1995). The context for the immediate test was very similar to the context available for immediate predictions. After a delay, the context had changed and the immediate cues used to predict performance may have been different from those available on the delayed multiple-choice test. This transfer-appropriate monitoring interpretation predicts high accuracy not only for a P_iT_i condition, but also for conditions in which both predictions and tests are delayed. Prediction accuracy in the delayed prediction conditions (P_dIT_d and P_dNIT_d) was generally low. However, delayed predictions with text add difficulty in that the cues must be sufficient for the participant to access the same knowledge about the text that will be used on the multiple-choice tests. The present results suggest that the match between the text information retrieved by the cues and the information used on the test may not have been high in the P_dIT_d and P_dNIT_d conditions, making the correlations between predictions and test performance low. Whether more complete cues would increase the prediction accuracy in delayed rating conditions remains to be seen. The conclusion from this experiment, however, is that under the present conditions, making immediate predictions followed by an immediate test pro-

duces higher prediction accuracy than does delaying either the predictions or the test.

REFERENCES

- BEGG, I., DUFT, S., LALONDE, P., MELNICK, R., & SANVITO, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory & Language*, **28**, 610-632.
- DUNLOSKY, J., & NELSON, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, **20**, 374-380.
- DUNLOSKY, J., & NELSON, T. O. (1997). Similarity between the cue for judgments of learning (JOL) and the cue for test is not the primary determinant of JOL accuracy. *Journal of Memory & Language*, **36**, 34-49.
- ERICSSON, K. A., & KINTSCH, W. (1995). Long-term working memory. *Psychological Review*, **102**, 211-245.
- GLENBERG, A. M., & EPSTEIN, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 702-718.
- GLENBERG, A. M., SANOCKI, T., EPSTEIN, W., & MORRIS, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, **116**, 119-136.
- GOODMAN, L. A., & KRUSKAL, W. H. (1954). Measures of association for cross classification. *Journal of the American Statistical Association*, **49**, 732-764.
- KORIAT, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, **100**, 609-639.
- MAKI, R. H. (1995). Accuracy of metacomprehension judgments for questions of varying importance levels. *American Journal of Psychology*, **108**, 327-344.
- MAKI, R. H., & SERRA, M. (1992). The basis of test predictions for text material. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 116-126.
- MORRIS, C. C. (1990). Retrieval processes underlying confidence in comprehension judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 223-232.
- NELSON, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, **95**, 109-133.
- NELSON, T. O. (1996). Consciousness and metacognition. *American Psychologist*, **51**, 102-116.
- NELSON, T. O., & DUNLOSKY, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science*, **2**, 267-270.
- SCHNEIDER, W. (1988). *Micro Experimental Laboratory (MEL)* [Computer program]. Pittsburgh: Psychology Software Tools.
- SPARGO, E. (1989). *Timed readings* (3rd ed.). Providence, RI: Jamestown Publishers.
- WEAVER, C. A., III (1990). Constraining factors in calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 214-222.
- WEAVER, C. A., III, & BRYANT, D. S. (1995). Monitoring of comprehension: The role of text difficulty in metamemory for narrative and expository text. *Memory & Cognition*, **23**, 12-22.

NOTE

1. The significant difference between the P_iT_i and the P_iT_d conditions was replicated in two other experiments. In one, mean gammas for the two groups, respectively, were .45 and .16 [$F(1,48) = 4.75$, $MS_e = 0.14$]; in the second experiment, the mean gammas were .48 and .24 [$F(1,135) = 6.68$, $MS_e = 0.30$].

APPENDIX

**FIRST HALF OF THE
EXAMPLE TEXT "FOOD ADDITIVES"
USED IN EXPERIMENTS 2 AND 3**

Food Additives

We can get along without food additives, but not very well. Were it not for food additives, we would have to go back to the old idea of bakery freshness—good today, stale tomorrow. Many adults remember when the cottage cheese separated, cookies dried up in two days, any food with fat or oil in it became rancid, canned vegetables and fruits were soft and mushy, and marshmallows got too hard to toast. Without additives, the variety and quality of foods would return to those familiar to grandmother. The quantities available would definitely be less, and quick foods would not exist. Some people have called these additives poisonous chemicals. If all chemicals are poisonous, then people should stop eating because all foods are chemicals. Some common additives are pure chemicals, such as the potassium iodide in table salt and many familiar vitamins that are all essential to man's health. Expressing foods in chemical terms can be a lengthy job. For example, milk is made of water, fats, proteins, lactose, salts, acids, pigments, enzymes, vitamins, and gases. Milk is a wonderful chemical, but hardly poisonous.

SECOND HALF OF EXAMPLE TEXT

Food Additives

Many allowable food additives are taken directly from food itself. For example, lecithin is found in all living things. It is obtained mostly from soybeans and is used as an emulsifier to keep ingredients in processed foods from separating. A large number of laboratory-created additives are also found naturally in foods. Calcium and sodium propionate, for example, are made during fermentation in the production of Swiss cheese. The propionates help to prevent mold. They are used primarily in baked goods. The vitamins used to improve the nutritive value of many foods are identical to the vitamins found naturally in food. Actually, all foods and additives are chemicals, and the body cannot tell the difference between them. The cost of an additive does not increase the cost of a food. For example, enough calcium propionate to help protect 1,000 loaves of bread from becoming moldy costs less than ten cents. In addition to these additives, there are also incidental additives that have no planned function in food but become a part of it during some phase of processing, packaging, or storing. Good examples are substances that might get into the food from a packaging material. The safety of incidental additives is scientifically controlled as are the other additives. A food additive must be of some benefit to a food or its production.

QUESTIONS FOR AN EXAMPLE TARGET TEXT

Fact Recall

According to the author, propionates prevent

- A. mold formation +
- B. thinning
- C. mushiness
- D. sourness

The author says that foods must be classified as

- A. chemicals +
- B. fuels
- C. calories
- D. organic

Comprehension Questions

For the most part, food additives are

- A. inexpensive +
- B. costly
- C. unnecessary
- D. essential

Incidental additives are

- A. of no benefit to food +
- B. put into food before processing
- C. high in vitamin content
- D. functional

The reader can conclude that

- A. most additives are safe to eat +
- B. most additives are organic by nature
- C. several additives are toxic in large quantities
- D. additives are unnatural

+ Correct alternative. Question alternatives were randomized for each participant.

**TITLES, BRIEF DESCRIPTIONS
OF THE TEXTS, AND FLESCH
READING SCORES (100 = EASIEST)**

Set 1

Ice Cream Plants \ patterns of consumption on the range resulting from overpopulation (69.5)

The Kitchen: An Expensive Room \ the location and layout of the essential parts of a kitchen (77.0)

Household Pests: Cracks, Crevices, and Crannies \ methods of preventing the breeding of household pests (70.2)

Food Additives \ types and functions of food additives (64.3)

The Andersonville Horror \ the lives of federal soldiers jailed in Andersonville, Southern prison during the Civil War (46.1)

Evaluation without Intimidation \ how employers can best evaluate employees to help them improve their performance (55.4)

Set 2

The Virgin Islands \ the discovery and early settling of various Caribbean Islands (67.1)

Tablets, Capsules, and Liquids \ the history of drugs from centuries past to modern day uses (68.5)

Home Wine Making: Rose Hips and Dandelions \ the history of and methods used in making homemade wine (69.9)

The Mystery of Migration \ theories behind why birds migrate (68.5)

Chocolate: The Magical Food \ the history and lesser-known nutritional benefits of chocolate (42.2)

Problems of Recycling \ history of and current methods and obstacles to recycling (46.8)

Note—The material in this Appendix is from *Timed Readings* (3rd ed., Book 7, p. 19-20) by E. Spargo, 1989, Providence, RI: Jamestown Publishers. Copyright 1989 by Jamestown Publishers. Reprinted with permission.