An Empirical Method of Assessing Topic Familiarity in Reading Comprehension Research

JAN H. SPYRIDAKIS, Department of Technical Communication, University of Washington
MICHAEL J. WENGER, Department of Psychology, State University of New York at Binghamton

ABSTRACT While numerous studies have shown that prior knowledge and topic familiarity affect comprehension and usability of a text, few researchers account for readers' topic familiarity in studies that investigate other reading comprehension variables. This research note briefly reviews the literature on the effect of prior knowledge and text familiarity on document comprehension and usability and discusses current methods for assessing subjects' topic familiarity. An empirically based method for effectively assessing subjects' topic familiarity is presented.

Over the past few decades, many researchers have demonstrated the effects of readers' topic familiarity and prior knowledge on the comprehension and usability of documents. Researchers interested in specifically studying the effects of prior knowledge or topic familiarity have acknowledged the effects in their experimental designs, by assessing the pre-existing knowledge of experimental subjects with respect to the experimental materials used in their studies. However, researchers interested in other variables affecting comprehension, recall, or usability rarely control for the effects of topic familiarity and prior knowledge when they assess the comprehensibility and usability of documents.

This article contends that topic familiarity and prior knowledge must be either assessed or controlled for in comprehension and usability studies; it presents a method that allows for assessing and controlling familiarity without introducing additional confounds. Whether researchers are specifically interested in the effect that topic familiarity or prior knowledge will have, they cannot ignore the issue and must define the requirements for subjects' background knowledge or experience in advance, just as they pre-select the levels of the independent variables under investigation. This article first briefly discusses the impact of subjects' prior knowl-
edge or experience on comprehension and recall, summarizing the effects that have been documented in the literature. Next, the current methods for assessing familiarity and prior knowledge are reviewed. Finally, a proposed method for assessing familiarity is presented.

The Impact of Familiarity on Comprehension.

Researchers have repeatedly pointed to the positive and negative impact that readers' prior knowledge or familiarity of the topic covered in experimental texts has on their comprehension (Anderson et al., 1978; Lantaff, 1978; Husain, 1983; Alvermann et al., 1985). Prior knowledge can provide ideational anchors for the acquisition of new information (Ausubel, 1968). Prior knowledge should facilitate readers in selecting text information to store in memory, and in linking new information with old information already present in memory. The findings of studies of readers' topic knowledge underscore the fact that comprehension depends on the nature of the reader or user as well as the nature of the document. This research has occurred under a number of theoretical rubrics, including the assessment of the effects of readers' schemata, prior knowledge, and cultural differences.

Researchers investigating the effects of schemata on reading comprehension continually identify the interaction of the reader's topic familiarity with the text. Anderson et al. (1978), investigating readers' interpretations of a passage in light of their existing cognitive schemata, found that readers' existing cognitive structures guide what they focus on in text, what they recall, and how they integrate and interpret the text. Pearson et al. (1979), investigating the effect of strong or weak schemata, concluded that subjects with strong schemata were able to infer implicit relationships. Lipson (1982), examining whether children use their familiarity with a topic in forming inferences from text, found that subjects who were familiar with the textual information were better at acquiring totally new information than correcting inaccurate old information.

Additionally, numerous studies of subjects with differing cultural backgrounds have shown that readers' cultural backgrounds and hence topic knowledge and world view affect their reading comprehension. Steffenson et al. (1979), comparing the performance of American and Indian adults who read texts about American and Indian weddings, identified interactions of nationality and text type on measures of reading speed, recall, and the ability to draw inferences or make elaborations. Defining cultural differences in terms of religious affiliation, Lipson (1983) found that subjects who read materials compatible with their religious backgrounds performed better on comprehension measures than subjects who read materials incompatible with their religious backgrounds. McCabe et al. (1983) probed the effect of socio-cultural differences in the comprehension of stories written with different structures and found that subjects with different socio-cultural backgrounds affected their expectations of story structure and comprehension.

While researchers who have specifically investigated the issues of schemata, prior knowledge, topic familiarity, or cultural background have shown that subjects' topic knowledge interacts with their comprehension and recall, researchers investigating other comprehension and usability issues often neglect to assess their subjects' topic knowledge and in many instances may be ignoring the confounding effect of subjects' existing knowledge, familiarity and experience—interpreting the effect of
the variable of interest without noting the influence of topic familiarity. The next section focuses on current indicators of topic familiarity, noting the weaknesses that influenced the design of the proposed method.

Methods for Assessing Familiarity

Currently, there appears to be no standard, demonstrated, validated method for assessing or controlling for the influence of prior knowledge and familiarity in experimental settings. The more common methods include direct pre-tests of pertinent pre-existing knowledge, indirect pre-tests of pre-existing knowledge relevant (but perhaps peripheral) to the experimental text, pre-training in the concepts and content of the experimental stimuli, and a variety of other approaches. Each of these methods has its strengths, yet each also has confounding influences that suggest the need for a more robust approach.

Direct Pre-test

One of the most common and certainly one of the most logical methods of determining experimental subjects’ pre-existing knowledge of, or familiarity with, topics used in experimental contexts is a direct pre-test of subjects who will participate in the experimental condition (Langer & Nicolaich, 1981; Flentchev-Kleffer, 1988). While the direct pre-test method appears to answer the need of researchers seeking ways to assess or control for familiarity effects, two difficulties arise with the use of the direct pre-test: priming of the subject’s memory (Premer & Snyder, 1975; N_EDY & Duranego, 1985; Kemp-Wheeler & Hill, 1985) and, by relying on a specific behavioral response mode. The difficulty associated with the direct pre-test is the question of whether the pre-test affects the post-test results and thus threatens the internal validity of the study. Memory for content that is related to the experimental passages may be activated by the pre-test before the subjects encounter the experimental passages and thus may increase the likelihood of their performing well on outcome measures. In essence, the subjects’ memory for content-specific information may be primed by the pre-test, thus confounding the patterns observed on the experimental inventories. Witlheite (1989), summarizing some of his previous work (Witlheite, 1988), observed the possible confounding influences of a familiarity pre-test in his studies of the effects of headings in text and noted that “for the subjects of high levels of pre-existing knowledge, the pre-test questions contributed to the facilitative effect of the headings” (p. 115).

The second difficulty with direct pre-tests is the question of the subjects’ ability to produce responses (the behavioral response mode) that indicate their true level of familiarity, such as the free association and free recall methods employed by Langer & Nicolaich (1981). If, for example, a subject has a high degree of familiarity with the experimental topic but in some way possesses a deficit with regards to the response mode required by the assessment (e.g. writing block), that subject’s degraded performance may be quantitatively identical to another subject’s performance, a subject who may not have a response deficit but may have a lower level of familiarity. A number of authors have suggested that when free-recall protocols are used in comprehension research, some measure of a subject’s ability to produce written responses should be obtained (Estes & Wetmore, 1983; Taylor, 1984).
Indirect Pre-tests

To avoid the confounding influence of priming a subject's memory, a number of investigators have attempted to assess topic familiarity in more indirect manners. Johnston (1984) used a measure of subjects' knowledge of related vocabulary to investigate the effects of topic familiarity and found that prior knowledge accounted for 3.9% of the within subject variance on a comprehension test. Pearson, Halsey, and Gordon's investigation of inference operations based on topic familiarity also used an indirect pre-test (an eight question pre-test related to the topic at hand) to categorize subjects as possessing strong or weak schemata for the topics covered in the experimental tests (1979).

Although indirect pre-tests appear to avoid priming subjects' memory, they can still prime memory, depending on the depth of subjects' topic familiarity and general skill level. Johnston (1984) touched on this second issue by referring to the common confound of vocabulary knowledge with IQ; he also noted that vocabulary knowledge was highly related to measures of reading comprehension and document readability.

Pre-training

Another obvious way to control the effects of familiarity is to directly manipulate subjects' familiarity through pre-training on the concepts and content of the experimental materials (Linde & Bergström, 1988; Barnes et al., 1989; Stahl et al., 1989; Wilhite, 1989).

While pre-training appears to be the newest approach to assessing or controlling subjects' familiarity with the experimental topic, it too has problems. Investigators who rely on pre-training have no information on the presence of serendipitous and compatible topic familiarity that may contribute to subjects' performance on outcome measures. For example, Wilhite (1989) assumed that subjects who had not taken a course on the organization of human memory would have much less familiarity with the topic than those subjects enrolled in a class dedicated to the topic. However, the other subjects may have taken courses (e.g. an introductory survey of human psychology course) that may have given them more pre-existing knowledge than the investigator assumed. Additionally, the belief that subjects have little prior knowledge because the topic concerns imaginary entities (Stahl et al., 1999) or because the subjects do not own or have any direct experience with the experimental task (Linde & Bergström, 1988) may ignore the fact that certain subjects may be familiar with topics that are compatible with the experimental topic familiarity that could aid their performance on experimental outcome measures. These issues seem to point to the need to assess the familiarity of a representative sample drawn from the experimental sample frame in order to assess actual topic familiarity.

Other Methods for Assessing Topic Familiarity

Researchers have used a variety of other methods to determine topic familiarity. Some researchers choose subjects from contrasting social or economic backgrounds and then use experimental materials that have some relationship to those backgrounds (Steffenson et al., 1979; Johnson, 1981; Lipson, 1983; Phillips, 1988).
Johnston (1984) suggested two other alternative means of controlling for the influence of topic familiarity. Investigators could select experimental passages covering a broad range of topics, passages that could be assumed to cover the range of topic familiarity possessed by all subjects in the sample frame. This 'shotgun' approach, however, seems expensive and could well be critically flawed if the topics selected did not show enough 'scatter' to cover the actual (versus the assumed) range of topic familiarity. A second alternative would be to assess the passage dependency of the outcome measure (as suggested by Trueman, 1973-74). However, passage dependency measures probe the construction of the outcome measure relative to the experimental text rather than relative to the pre-existing knowledge of subjects. While subjects who possess prior knowledge of a topic might perform well on a passage dependency test, their scores would be an inseparable composite of the passage dependency of the text and their familiarity with the topic. The actual effect of topic familiarity could only be observed after subjects had read the experimental text and then could well be confounded with having participated in the passage dependency test.

Design of an Empirical Method for Assessing Topic Familiarity

The preceding discussion has suggested a number of considerations for the design of an empirical method for assessing subjects' topic familiarity. These would include: (1) avoiding the potential priming effect of a pre-test; (2) controlling for variations in subjects' ability to produce responses indicative of topic familiarity; (3) controlling for variations in subjects' skills or intelligence and variations in serendipitous familiarity with compatible topics; and (4) maintaining an empirical link to the familiarity of representative members of the subject population.

With these considerations in mind, we devised a method for assessing topic familiarity that we believe to be effective and efficient. This method was originally devised and successfully used by the first author and was subsequently employed by the second author in order to assess its robustness (Spyridakis, 1987; 1989a; b; Wenger, 1989; Wenger & Spyridakis, submitted for publication). We have found that this method allows a researcher to assess the topic familiarity of members of the sample population without priming subjects' long-term memory, that it does not rely on a well developed repertoire of subject response skills (e.g., producing written or verbal responses), that it allows for scores that can be used as covariates for analysis of comprehension or recall data, and that it allows investigators to select experimental texts based on actual responses of members of the experimental population.

In addition, the proposed method is grounded on the premise that subject responses are better indicators of subject familiarity than an investigator's analytical assessment. Thus, subject ratings are used at all stages of the method. To control for the disjunction between studies of topic familiarity and the operation of that familiarity in actual reading behavior, our method depends entirely on the behavioral ratings of members of the experimental population used in the study. In essence, this method uses the topic familiarity responses of one sample to predict the topic familiarity responses of the experimental sample.

The selection of experimental texts begins with the investigator specifying the degree of familiarity required for the study. For example, an investigator may wish to compare the effects of a highly familiar text with those of an unfamiliar text, or, an investigator may specify that the text to be read by subjects should be unfam-
358 J. H. Spyrakakis & M. J. Wenger

ilius. In any case, the investigator should pre-specify the level of topic familiarity desired.

The next step is to select at least ten naturally occurring texts on different topics that the investigator believes would be acceptable to the study needs and would also cover the range of desired familiarity levels. Once these texts are selected, the investigator should read each thoroughly to gain a working knowledge and then write a two- to three-sentence description of each text.

With the goal of having subjects rank order the ten topics in terms of familiarity, the investigator presents the topic descriptions, accompanied by the titles of the articles, to a representative sample of subjects, drawn from the sample frame for the experimental treatments. (These subjects should not participate in the ensuing experimental treatments.) The subjects in this preliminary sample are asked to read the titles and descriptions and then complete two tasks. First, they divide the text summaries with their titles into two groups, one representing topics that the subjects perceive as more familiar and the other representing topics that subjects perceive as less familiar. The rationale for dividing the ten topics into two groups of five can be found in studies of memory limitations that have pointed out that humans are best able to discriminate among like objects when those objects are arranged in small groups of five to seven items (Miller, 1956; Mandler, 1967, 1970). Subjects then rank the five topics in each category in order of increasing familiarity (with '1' in each group being the least familiar and '5' in each group being the most familiar). Thus, the researcher ends up with two groups of topics from each subject, each group which is ordered according to perceived familiarity.

While these initial rankings represent two ranked sets on a five-point scale, the two sets are converted to a single scale, ranging from a low of 1 (the least familiar of the "unfamiliar" group of topics) to a high of 10 (the most familiar of the "familiar" group of topics). This conversion is done for each subject's response. The data from all subjects are then aggregated, and the investigator calculates the minimum, maximum, and median rank given each article. Since the data are based on an ordinal scale, medians are calculated instead of means. The medians can then be used to select articles with the desired familiarity level. Finally, in the experimental conditions after new subjects have read the complete passages, they can rate their familiarity with the text on a five-point familiarity scale; this allows the investigator to check the validity of the original familiarity rankings.

Conclusions

As noted earlier in this article, many researchers have noted the impact of topic familiarity on comprehension and recall of textual information. However, the impact of familiarity will not be observable without some method for empirically identifying the familiarity of representative members of the experimental population with the topics selected for experimental materials. The method we have proposed was designed given the difficulties documented for the more commonly used methods of assessing and controlling for prior knowledge. First, by using two subject groups (each drawn from the same sample frame)—one group for the familiarity assessment of ten topics and one group for the experimental conditions—our method avoids priming the memory of experimental subjects. Second, since our method relies on behavioral responses that can be assumed to be part of the repertoire of all subjects (reading and making discriminatory judgments among a
Assessing Topic Familiarity

small set of items), the method avoids relying on skills that may or may not be equally distributed in the sample. Third, because the method uses actual representative subject responses to guide the selection of experimental materials, the presence of serendipitous forms of topic familiarity is controlled far more effectively than when experimenter assumptions guide the selection of materials. Fourth, the method maintains an empirical and verifiable link between the judgements of topic familiarity and the responses of experimental subjects. Fifth, the method is easily administered and exacts only nominal costs in terms of the preparation of materials, administration of the pre-test, and analysis of the data. Finally, the responses of representative readers can be seen as better predictors of topic familiarity than experimenter analyses.

Correspondence: Jan H. Spirydaklis, Department of Technical Communication, University of Washington, Seattle, WA 98195, USA.

REFERENCES


Miller, G. A. (1957) The magical number seven, plus or minus two: some limits on our capacity for processing information, Psychological Review, 63(2), pp. 81-96.


