

Making Legal Language Understandable: A Psycholinguistic Study of Jury Instructions

Robert P. Charrow*

Veda R. Charrow**

There has been a growing concern in recent years regarding the inability of lay persons to understand legal language.¹ This concern has generated a movement to rewrite legal documents in "plain English." At both the federal and state levels, laws have been enacted and regulations issued, requiring that automobile insurance policies, warranties, and other legal instruments be written in language that is clear and understandable to the average person.² Many of these efforts have been greeted with loud

© 1979 Robert & Veda Charrow

* Assistant Professor of Law, Howard University. B.S. 1966, Harvey Mudd College; J.D. 1969, Stanford University.

** Research Scientist, American Institutes for Research. B.A. 1967, McGill University; M.A., Ph.D. 1974, Stanford University.

The research reported in this Article was supported by National Science Foundation Grant SOC 76-01580, which provided funds for a series of six experiments; the results of two of those experiments are presented in this Article. This study was funded through the Center for Applied Linguistics. The authors gratefully acknowledge the logistical assistance of the judges, attorneys, and court personnel who facilitated this project. The authors are most grateful for the research assistance of Thelma Litsas, Joan Atherton, Judith Ottesen, Frank Humphrey, and Miles Pewitt, and for the special assistance of Herbert H. Clark, Michael Cramer, Annalee Elman, Charles Fillmore, Peg Griffin, Michael Karchner, Kari Sajavaara, and Roger Shuy. Thanks also to Ruby Berkemeyer who typed just about everything.

1. See, e.g., Goldfarb, *My Secretary, Hereinafter Referred to as Cuddles* . . . , BARRISTER, Summer, 1978, at 40; Winslow, *The Instruction Ritual*, 13 HASTINGS L.J. 456 (1962); DRAFTING DOCUMENTS IN PLAIN LANGUAGE (D. MacDonald ed. 1979).

2. At the federal level, President Carter's Executive Order No. 12044, "Improving Government Regulations," 43 Fed. Reg. 12661, 12662 (1978), requires that each federal regulation be "written in plain English and . . . understandable to those who must comply with it." The Magnuson-Moss Warranty Act, 15 U.S.C. §§ 2301-2312 (1976) provides in part that warranties accompanying consumer goods must be written in "simple and readily understandable language." 15 U.S.C. § 2302. Although the Act went into effect in 1975, the Federal Trade Commission has not yet promulgated regulations defining the phrase "simple and readily understandable language." S. 1312, 95th Cong., 1st Sess. (1977) would amend the Truth in Lending Act by requiring the Federal Reserve Board to issue model loan forms that utilize "readily understood language."

Many states have shown an interest in simplifying legal language. New York enacted the first "plain English" law in 1977. See Act of Aug. 5, 1977, 1977 N.Y. Laws, ch. 747 (codified at N.Y. GEN. OBLIG. LAW § 5-701(b), (c) (McKinney 1978)). Under the terms of the New York law a large class of consumer contracts must be "written in non-technical language and in a clear and coherent manner using words with common and every day meanings." Following the example of New York, two other states, Maine and Connecticut, have recently enacted similar laws. See Act of June 20, 1979, ch. 483, 1979 Me. Legis. Serv. 1206 (to be codified in ME. REV. STAT. ANN. tit. 10, §§ 1121-1126); Act of June 22, 1979, Pub. Act No. 79-532, 1979 Conn. Legis. Serv. 1517. Moreover, the New York law has stimulated many other state legislatures to consider adopting "plain English" laws. See, e.g., Alaska (House Bill No. 592, introduced Jan. 10, 1978); Arizona (Senate Bill 1096, introduced Jan. 16, 1978); California (Senate Bill 2067, introduced Apr. 3, 1978); (Assembly Bill 3607, introduced Apr. 3, 1978) (dealing with insurance policies); Colorado (Senate Bill No. 12, introduced Jan. 4, 1978); Illinois (House Bill 2509, introduced Nov. 4, 1977); Iowa (House File 2332, introduced Mar. 14, 1978); Maryland (House Bill 593, introduced Jan. 11, 1978); Massachusetts (House Bill 985, introduced 1977); Minnesota (Senate File No. 2291); (House File No. 2443, introduced Mar. 1, 1978); Nebraska (Legislative Bill 626, introduced Jan. 4, 1978); New Jersey (Assembly No. 536, adopted by Assembly June 22, 1978); Pennsylvania (Senate

praise and with confidence that rewriting will alleviate all comprehension problems.

But there are major issues that have not been addressed.³ For one thing, although it is assumed that all "legalese" is incomprehensible, there is no real data, aside from anecdotes, to support this assumption or to elucidate the exact nature of the problem. There is no empirical evidence of the extent to which legal language is not understood, nor is there any data regarding those segments of the population—aside from lawyers and bureaucrats—that may *not* have problems comprehending legalese.⁴

As for the rewriting of legalese into "plain English," there are no criteria for determining what constitutes "plain English," and no empirically determined rules for rewriting. Instead, there has been a reliance on such makeshift or invalid devices as readability formulas,⁵ frequency dictionaries and "commonsense"—although not necessarily accurate—grammatical rules of thumb. Moreover, while the questions of what is "plain English" and what is "clear and understandable" will undoubtedly stimulate a flurry of litigation, the courts, at this time, appear to be ill-prepared to grapple with such questions. They lack the linguistic expertise and the necessary empirical tools to make sound determinations concerning clarity or comprehensibility.⁶

This Article presents the results of the first empirical, objective linguistic study of the comprehensibility of one type of legal language—standard jury instructions. We have not merely attempted here to demonstrate that jury instructions are inadequately understood; we have also attempted to isolate those linguistic features typical of this brand of legalese—aspects of legal

Bill No. 1301, introduced Feb. 21, 1978); Rhode Island (78-H 7311, printed Feb. 3, 1978); West Virginia (House Bill 1339, introduced Jan. 27, 1978). In addition, Pennsylvania and Delaware require that automobile insurance policies be written in clear and understandable English. See 31 Pa. Code §§ 64.2-14 (1979); DEL. CODE ANN., tit. 18, § 2740 (Supp. 1978).

3. See V. Charrow, *Let the Rewriter Beware* (ms. 1977).

4. It is also assumed that everyone knows exactly what legalese is, but this jargon has never been systematically described, and there are no linguistic criteria for ascertaining what is and what is not legalese. Indeed, it is not entirely certain that legal language is a jargon; research could reveal it to be a dialect or sublanguage of English, but no such research has been done. See V. Charrow & R. Crandall, *Legal Language: What Is It and What Can We Do About It* (paper presented at the American Dialect Society Conference, Washington, D.C., Oct. 1978).

5. E.g., the Flesch Formula, the Gunning "Fog Index," and the Fry Scale. See note 10 *infra*.

6. Furthermore, the courts have been remarkably reluctant to consider seriously the comprehensibility of legal language. For instance, in *Garza v. Chicago Health Clubs*, 347 F. Supp. 955 (N.D. Ill. 1972), the court was called upon to determine whether a particular notice that had been included in the defendant's loan agreement in order to comply with the truth-in-lending laws could be understood by the average person. The section in question read as follows:

... and to consent to immediate execution upon any such judgment and that any execution that may be issued on any such judgment may be immediately levied upon and satisfied out of any personal property of the undersigned . . . and to waive all right of the undersigned . . . to have personal property last taken and levied upon to satisfy any such execution.

Id. at 960. The court, without revealing the basis for its decision, held that although the clause was written in somewhat legalistic terms, the language was understandable. See also *Rodgers v. United States Steel Corp.*, 70 F.R.D. 639, 645 (W.D. Pa. 1976).

grammar, semantics, vocabulary, and discourse structure—that cause the comprehension problems. We have then used this knowledge to rewrite jury instructions in a systematic fashion, and have empirically verified that such rewriting can yield positive results.

Most importantly, we have formulated an empirical psycholinguistic methodology that can be used by courts, legislatures, regulatory agencies, and the private sector to determine whether or not a document is comprehensible to the intended audience. The proposed methodology, when properly applied, can provide insight into the linguistic reasons underlying specific comprehension problems.⁷

I. AN OVERVIEW OF THE STUDY

Prior research in the area of jury instruction comprehension had modest goals: to measure the degree to which standard instructions were comprehended and, in a few cases, to illustrate that redrafting jury instructions could improve the comprehensibility of a packet of instructions as a whole.⁸ Little

7. We employed psycholinguistic methodology because psycholinguistics applies the techniques of experimental psychology to the problems of language processing and comprehension, and provides a number of useful tools for assessing the comprehensibility of legal language.

8. There have been relatively few studies concerning the comprehensibility of standard jury instructions. The studies that have been performed fall into two categories: first, studies that utilized subjective techniques—questionnaires, interviews, or observations—to determine jurors' comprehension; and second, studies that employed objective measures—e.g., multiple-choice tests—or a combination of objective and subjective measures.

The subjective studies provide some evidence—although equivocal at times—that jury instructions are less than models of clarity. See, e.g., Speech of Walter B. Wanamaker, The Cincinnati Conference (Feb. 20, 1937), printed in *Trail by Jury*, 11 U. CIN. L. REV. 119, 191 (1979), in which 36.4% of the 843 jurors who responded to the questionnaire believed that less than half of their peers had thoroughly understood their duties; Hervey, *The Jurors Look at Our Judges*, 18 ORLA. B.A.J. 1508 (1947), in which 40% of the 185 responding jurors believed that they did not clearly understand the instructions of the judge. See generally Broeder, *The University of Chicago Jury Project*, 38 NEB. L. REV. 744 (1959); Cook, *Instructions: Legalistic Lingo of Contrived Confusion*, 7 J. MO. B. 113 (1951); Hoffman & Brodley, *Jurors on Trial*, 17 MO. L. REV. 235 (1952); Hunter, *Law in the Jury Room*, 2 OHIO ST. L.J. 1 (1935); Kalven, *A Report on the Jury Project of the University of Chicago Law School*, 24 INS. COUNSEL J. 368 (1957); Maloney, *Should Jurors Have Written Instructions?*, TRIAL JUDGES J., Apr. 1967, at 18; Meyer & Rosenberg, *Questions Juries Ask: Untapped Springs of Insight*, 55 JUDICATURE 105 (1971).

The objective studies that have been performed either display methodological flaws or provide little or no information concerning those aspects of the jury instructions that actually cause comprehension difficulties. See Forston, *Judge's Instructions: A Quantitative Analysis of Jurors' Listening Comprehension*, TODAY'S SPEECH, Fall 1970, at 34, in which a multiple-choice test was used to measure the percentage of the judge's instructions that both individual jurors and deliberating juries retain and comprehend. Forston's methodology was flawed in at least two respects: the same subjects were used in both treatments, creating practice effects, and the author failed to perform the appropriate statistical tests. See also O'Reilly, *Why Some Juries Fail*, 41 I.B.A.D.C. 69 (1974), which analyzed the results of 173 quizzes that had been completed by jurors who had recently served. The study was seriously confounded because the quiz did not test jury instruction comprehension, but only the ability of people to retain the meanings of specific legal terms over a long period. Furthermore, given the variation in the amounts of time that elapsed between the jurors' jury duty and their completing the quiz, there was no proven correlation between incorrect answers on the quiz and a lack of understanding of jury instructions.

Three sets of more careful and sophisticated studies revealed that rewriting jury instructions could improve comprehension. None of these studies, however, empirically isolated the linguistic features that actually impeded comprehension, nor were they able to show consistent improvement on an instruction-by-instruction basis. All of these studies used multiple-choice

attention has been paid to measuring the changes in the comprehensibility of specific individual instructions as a function of changes in the language of those instructions, and no attempt has yet been made to correlate specific linguistic modifications with changes in subjects' performance. The experiments in this study are intended to fill these gaps by measuring the extent to which certain linguistic changes affect comprehension, on an instruction-by-instruction basis.

The study has two major goals. First, it seeks to ascertain the validity of several hypotheses about legal language in general, and about standard jury instructions in particular: (1) that standard jury instructions—when viewed as discourse—are not well understood by the average juror; (2) that certain linguistic constructions are largely responsible for this hypothesized incomprehensibility; and (3) that if the problematic linguistic constructions are appropriately altered, comprehension should dramatically improve, notwithstanding the "legal complexity" of any given instruction. A second and perhaps more important purpose is to develop a reliable and workable methodology, capable not only of assessing the relative comprehensibility of jury instructions, but also of isolating problematic linguistic constructions in legal language. Use of such a methodology can provide guidance in determining how the comprehension of legal language can be improved.

In order to accomplish these objectives, we designed and administered two experiments using a paraphrase task. A paraphrase task was chosen as the most powerful method for gathering the sorts of data necessary to test the three hypotheses set forth above.⁹ While there are several other possible methods of studying language comprehension and of evaluating the comprehensibility of discourse, all of them are indirect and incapable of isolating the sources of comprehension difficulty.¹⁰ Paraphrase testing, on the other

type tests. See H. Sigworth & F. Henze, *Jurors' Comprehension of Jury Instructions in Southern Arizona* (ms. 1973). One set of studies was conducted in Florida. See Strawn, Buchanan, Pryor, & Taylor, *Reaching a Verdict, Step by Step*, 60 JUDICATURE 383 (1977); Strawn & Buchanan, *Jury Confusion: A Threat to Justice*, 59 JUDICATURE 478 (1976); K. Taylor, B. Pryor, R. Buchanan, & D. Strawn, *An Investigation of Juror Comprehension and Attitudes Regarding Pattern Instructions* (paper presented at the Convention of the Speech Communication Association, San Francisco, 1976). The other set of studies was performed in Nebraska. See Sales, Elwork, & Alfini, *Improving Comprehension for Jury Instructions*, in 1 PERSPECTIVES IN LAW AND PSYCHOLOGY: THE CRIMINAL JUSTICE SYSTEM (H. Sales ed. 1977); Elwork, Sales, & Alfini, *Juridic Decisions: In Ignorance of the Law or in Light of It?*, 1 LAW & HUMAN BEHAVIOR 163 (1977).

9. See Bransford & Johnson, *Contextual Prerequisites for Understanding: Some Investigations of Comprehension and Recall*, 11 J. VERBAL LEARNING & VERBAL BEHAVIOR 717 (1972); Fillenbaum, *Memory for Gist: Some Relevant Variables*, 9 LANGUAGE & SPEECH 217 (1966); Fodor & Garrett, *Some Syntactic Determinants of Sentential Complexity*, 2 PERCEPTION & PSYCHOPHYSICS 289 (1967); Frederiksen, *Acquisition of Semantic Information from Discourse: Effects of Repeated Exposures*, 14 J. VERBAL LEARNING & VERBAL BEHAVIOR 158 (1975); Johnson, *Abstractive Processes in the Remembering of Prose*, 66 J. EDUC. PSYCHOLOGY 772 (1974).

10. Standard reading and oral comprehension tests (such as the SAT and Iowa Test) utilize open-ended questions or multiple-choice tasks. Such tests are not well suited for testing the hypotheses noted above. Standard tests are designed to test people, not materials. While they could be and indeed have been used to evaluate the comprehensibility of jury instructions, they provide little information concerning the reasons for poor performance. This is because in order to devise the questions and multiple-choice answers, the investigator

hand, is probably the closest thing to "getting inside the head" of the listener or reader. In a paraphrase task, a subject either listens to or reads some material and is then required to paraphrase it. The validity of the paraphrase task as a measure of comprehensibility rests on the premise that a subject will not be able to paraphrase accurately material that he or she has not understood. In addition, the subject will be more likely to focus upon concepts that are more comprehensible and those that are more important to the gist of the discourse, and to gloss over or omit less comprehensible or less important concepts.

Since jury instructions are usually given orally, and since jurors rarely have access to a printed copy of the instructions, we decided to employ a paraphrase task in which the instructions would be presented orally.¹¹ It was also decided that paraphrases should be oral, rather than written, to reduce the likelihood that subjects' writing skills would confound the results.¹²

In scoring paraphrases, one arguable limitation of the methodology is its apparent inability to distinguish whether an omission is due to lack of comprehension or such more benign factors as the triviality or obviousness of the item to be paraphrased. An examination of the omitted item, however, especially when taken in the context of the subjects' performance on the items surrounding the omitted item, can often resolve the problem. In any event, we designed certain statistical means for factoring out arguably

must understand what the problems are beforehand. If the investigator does not know what the problems are, and does not wish to influence the results with his or her own expectations, he or she cannot use a traditional test.

Tests of comprehensibility are even more of a problem. The most widely used tests for evaluating the comprehensibility of reading material are readability formulas, such as the Flesch test, the Gunning "Fog Index," and the Fry Readability scale. (For a review of most types of readability formulas, see G. KLARE, *THE MEASUREMENT OF READABILITY* (1963).) However, serious questions have been raised as to their validity as measures of comprehensibility or even as measures of "readability." See V. Charrow, *Let the Rewriter Beware* (ms. 1977); W. Kintsch & D. Vipond, *Reading Comprehension and Readability in Educational Practice* (paper presented at the Conference on Memory, Uppsala, Sweden, 1977). Even advocates of readability formulas recognize that they do not measure comprehensibility or comprehension, but rather "readability," which has to do with writing style. See G. KLARE, *supra*, at 16-17.

The "cloze" procedure was also originally devised as a readability measure. See Taylor, "Cloze Procedure": *A New Tool for Measuring Readability*, 30 *JOURNALISM Q.* 415 (1953). In a "cloze" procedure every fifth word or so is deleted from a passage and subjects are required to supply the missing words. (Sometimes multiple-choice answers are supplied.) Errors and omissions are then analyzed to ascertain a subject's reading problems. In addition, by observing error trends in the responses of a large number of subjects whose reading level is already known, the "cloze" test can serve as gauge of the "readability" of given passages. However, it is not necessarily a valid measure of comprehensibility. See W. Kintsch & D. Vipond, *supra*. Its major usefulness appears to be as a measure of children's reading comprehension when constructed using multiple-choice answers. See Dieterich, Freeman, & Griffin, *Comprehension Testing*, in *ABOUT A READING COMPREHENSION TEST: FOUR BOOKLETS* (Center for Applied Linguistics 1977).

11. In the case of oral discourse, there is an added problem for the listener—the need to remember what was said. See Miller, *Some Psychological Studies of Grammar*, 17 *AM. PSYCHOLOGIST* 748 (1962); Miller & Isard, *Some Perceptual Consequences of Linguistic Rules*, 2 *J. VERBAL LEARNING & VERBAL BEHAVIOR* 217 (1963); Miller & Selfridge, *Verbal Context and Recall of Meaningful Material*, 63 *AM. J. PSYCHOLOGY* 176 (1950). The added memory load involved in paraphrasing oral discourse would tend to enhance the negative effects of less comprehensible items or constructions in the discourse.

12. As it turned out, one of the subjects was functionally illiterate.

trivial and self-evident items from the instructions.¹³ Another potential limitation of the methodology is the large number of possibly interacting variables in the instructions. An item that is not well comprehended in an instruction might contain several problematic linguistic constructions—for example, a double negative¹⁴ embedded¹⁵ in a passive voice¹⁶ clause. It may be difficult, in some instances, to determine exactly which construction—or combination of constructions—is at the root of the comprehension difficulty. Fortunately, the instructions that were selected tended to have problematic constructions both in combination and in isolation, and in many cases it was possible to make a fair assessment of the amount that each construction contributed to the comprehension difficulty.

As mentioned above, the study consisted of two major experiments. The subjects of the experiments were people who had been called for jury duty in Prince Georges County, Maryland. In the first experiment, jurors were asked to paraphrase each of fourteen standard civil jury instructions from California; the paraphrases were then linguistically analyzed. The analysis revealed the existence of numerous grammatical constructions, phrases, and words that appear both to typify legal language and to affect jurors' comprehension adversely. In the second experiment, the jury instructions were rewritten to eliminate the apparently problematic items and constructions, and the paraphrase task was repeated with new subjects. From the results of the two experiments, we were able to show overall improvement for the modified instructions, and improvement on an instruction-by-instruction basis. Most significantly, we were able to isolate specific linguistic features of jury instructions—and of legalese in general—that interfere with the lay person's understanding of legal language.

II. THE FIRST EXPERIMENT

A. Description of the Experiment

For basic materials we chose fourteen California standard civil jury instructions, drawn from a corpus of fifty-two "BAJI"¹⁷ instructions. As a prelude to the present study, we had experienced California trial attorneys

13. See text preceding note 26 *infra* (discussing the "approximation measure").

14. See text accompanying notes 48 & 49 *infra*.

15. See text accompanying note 59 *infra*.

16. See text accompanying notes 50-55 *infra*.

17. BAJI stands for Book of Approved Jury Instructions, and is the common title for California's standard civil jury instructions. They are compiled in CALIFORNIA JURY INSTRUCTIONS—CIVIL—BOOK OF APPROVED JURY INSTRUCTIONS (5th ed. 1969). The texts of the instructions used in these experiments are set forth in section IV *infra*. It should be noted that a number of the instructions no longer reflect California law. Specifically, California no longer employs the doctrine of contributory negligence. However, since we were using jurors from a jurisdiction that still employed the doctrine, the court felt that exposing jurors to a different doctrine might be prejudicial. It should also be noted that we added an element to BAJI 3.01, to accommodate the issue of assumption of risk. BAJI 3.01 was often modified this way in actual California civil trials where assumption of risk was an issue. Unless the context indicates otherwise, subsequent references to BAJI 3.01 are to the instruction as so modified.

rate the fifty-two instructions for "conceptual complexity." The fourteen instructions were selected with a view to assuring a wide range of conceptual difficulty, as predetermined by the attorneys' ratings; a range of length—from 27 words to 154 words; and a variety of interesting linguistic constructions that are common to legal language. The final selection criterion was that the instructions, when read together, be a representative sample of a packet of instructions that would normally be presented in a highway accident case.

The instructions were recorded on audio cassettes by a male attorney acting the part of a judge.¹⁸ Each instruction was recorded twice, with a five-second pause between each recording.

As subjects, thirty-five persons were randomly selected from a panel of prospective jurors in Prince Georges County, Maryland. Twenty-two of the juror-subjects were male and thirteen were female. Twenty-seven jurors were white and eight were black. Educational levels ranged from twelfth grade through Ph.D.¹⁹ Although participation was voluntary, none of the jurors selected refused to participate in the study.

Jurors were selected at the beginning of their terms of duty, to ensure that they had not yet served on a jury. This was done in order to control for formal courtroom exposure to legal language, and particularly jury instructions. Although two of the jurors had at one time served on a jury, no subject had served within the past five years.

The thirty-five jurors were randomly divided into four almost equal groups. Because we wished to eliminate the possibility that the order in which the instructions were given would influence the results, four master tapes were recorded, each presenting the instructions in a different order, and a different master tape was presented to each of the four groups. (The first three instructions—BAJI 1.00, 1.01, and 1.02—were denoted "practice instructions," and were always presented first, and in that order.)²⁰ The instructions were presented in a one-on-one situation: one experimenter to one juror. No attempt was made to simulate an actual trial situation.²¹

The juror and the experimenter sat at a table in a room in the courthouse. There were two tape recorders on the table—one with the master

18. Tape recordings were used for consistency—to ensure that all subjects heard a given jury instruction in the same way. The male attorney was chosen because his delivery was better than that of many judges we listened to. The attorney was directed to read each instruction at approximately the same rate of speed; the mean rate of speed at which they were read was 131.6 words per minute, with a standard deviation of 17.37 words per minute. The variability in recording speed appears to be largely a function of natural pauses in the text.

19. See text accompanying note 39, and Tables 7 & 8 *infra*.

20. Thus, each of the four began with the three practice instructions, followed by the remaining eleven instructions in one of the three random orders, or in natural order (*i.e.*, the order in which they would normally be given in a trial).

21. Since our purpose was to isolate the linguistic factors that may impede comprehension of standard jury instructions (and similar legal language) and to examine the comprehensibility of jury instructions as *discourse*, the fact that some actual trial proceeding might elucidate one or more instructions was irrelevant.

tape, the other with a blank cassette. The experimenter read the following explanation of the task to the juror:

In this study I am going to play some jury instructions for you on this tape recorder. After I play each jury instruction twice, I would like you to paraphrase the instruction—that is to say, explain the instruction in your own words—as best you can. Your explanation will be recorded on this machine. We are evaluating the jury instructions, we are not testing you.

The experimenter gave the juror a drawing of an automobile accident that would give rise to a lawsuit in which the fourteen jury instructions would be given. The subject was then read a brief context paragraph describing the events in the picture, as follows:

The jury instructions that you will hear are like those that a judge might give at the end of a civil trial involving a highway accident. For example, as the drawing shows, a truck and an automobile collide. A passenger in the truck is injured. This passenger sues the driver of the automobile. The passenger, who is bringing suit, is known as the plaintiff in the case. The driver of the automobile, who is being sued, is known as the defendant. The truck driver's name is John Smith.

In addition, subjects were told that they could refer to the picture at any time.

The experimenter then played the first practice instruction twice²² on the master tape recorder. After the subject had heard the instruction for the second time, he or she orally paraphrased it into the second tape recorder, which was kept on "record" throughout the experiment in order to minimize distractions and reduce self-consciousness on the part of the jurors. After paraphrasing each of the three practice instructions, the subject was asked if he or she had any questions about the task, and then whether he or she was prepared to continue. Any questions concerning the procedure were answered, and all subjects chose to continue. The remaining instructions were then each played twice, in succession, and after the second playing the subject paraphrased the instruction. All subjects appeared to understand the task and paraphrased the instructions with varying degrees of completeness and accuracy.²³

The paraphrase task took approximately forty-five minutes per subject.

22. Since the purpose of this experiment was to obtain analyzable linguistic data, we found it necessary to play each instruction twice, as a pilot test with only one playing produced very few analyzable paraphrases.

23. During the course of the experiment, after completing a paraphrase, jurors were given encouragement, such as "Very good," or "You're doing fine." If a juror showed substantial hesitation in paraphrasing, he or she would be asked neutral questions, such as "What do you think the judge said?," "What does that mean?," and "Can you remember any more?" Any such prodding by experimenters was recorded, along with the subjects' responses. There were few such instances.

After completing the task, the subjects filled in a demographic questionnaire. Each subject's paraphrase cassette was transcribed, along with any comments or remarks made by either the subject or the experimenter. Each transcript was then checked against the cassette for transcription errors, pauses, and intonational clues to meaning.

B. *Analysis of the Data*

1. *Linguistic Breakdown.* Our strategy was first to ascertain whether a subject had paraphrased a given portion of an instruction accurately or inaccurately, or whether he or she had omitted it. In order to do this, each instruction was broken down into a number of constituent units. Originally, these constituent units corresponded to linguistic macro-units such as clauses or phrases, but where we felt, from prior psycholinguistic experiments or from an overview of the data, that a given grammatical or semantic construction or unit would affect comprehension, that construction or unit was isolated as a separate variable. Ultimately, the best breakdown was the one that not only maximized linguistic information but that made it easiest to judge objectively what a subject had understood. The breakdown for BAJI 1.00 was as follows:

Variable Number	Text
005	It is my duty
006	to instruct you in the law
007	that applies to this case
008	and you must follow the law
009	as I state it to you
012	As jurors
010	it is your exclusive duty
011	to decide all questions of fact
013	submitted to you
016	and for that purpose
014	(it is your duty)
015	to determine the effect and value of the evidence
017	You must not
018	be influenced by
019	sympathy
020	prejudice
021	or passion

We then compared a subject's paraphrase of an instruction to the breakdown of the instruction. For each constituent unit, the subject's response could receive one of the following scores: correct, correct by inference,²⁴ wrong, or omitted.²⁵

24. Where a subject had failed to paraphrase a given constituent unit explicitly, but where the existence of the unit could reasonably be inferred from the rest of the subject's paraphrase, the subject was given credit for that item. For purposes of determining a subject's performance, "correct by inference" was treated as the equivalent of "correct."

25. Each instruction was scored at first by two raters, acting independently. Agreement

2. *Measuring Performance.* There are a number of ways of computing the performance of each subject on a given instruction. The simplest and most straightforward approach—a full performance measure—is merely to sum the number of variables within an instruction that a subject paraphrased correctly and those that he paraphrased incorrectly or omitted. A problem with this approach is that it is very gross: it does not differentiate between variables within an instruction that are important to the legal concept and those that are trivial, redundant, or obvious. If, for example, half of the variables within an instruction were essential to a proper statement of the law and half were nonessential “padding,” the full performance measure would not be able to differentiate between a subject who correctly paraphrased all the essential features of the instruction and omitted all the trivial features and one who correctly paraphrased all the trivial features but omitted all the essential features. Both subjects would score 50% correct on that instruction.

For this reason, we constructed an approximation to each instruction consisting of only the most essential variables necessary for an accurate statement of the law. In order to choose the essential variables for each instruction objectively, we devised a set of rules that when applied to an instruction would generate the approximation. According to these rules the approximation would include any variable relating to jurors' duties, as well as the basic ideas—but only those ideas—necessary for a statement of the law. Where the instructions contained a list of items joined by “or,” the approximation would be satisfied by any *one* of these items. Similarly, the approximation would be satisfied by any one of equivalent phrases, sentences, or paragraphs appearing in an instruction. The approximation would *not* include statements of the judge's duties or definitions of terms where the term was then used in the same instruction.²⁶ The approximation also excluded exceptions to general rules that were stated in the same instruction.

Application of these rules to BAJI 1.00 resulted in the reduction of the seventeen variables of the full performance measure to an approximation with only eight variables:

Variable Number	Text
008	and you must follow the law
010	it is your exclusive duty
011	to decide all questions of fact
014	it is your duty

was very high (approximately 90%) and so, in order to save time, further instructions were scored by two raters working together. When problems or uncertainties arose, they were resolved through discussion with the other rater.

26. For example, in BAJI 3.50 the definition of “contributory negligence” would be excluded from the approximation because contributory negligence was then used in the statement of the law within the instruction. However, BAJI 3.75 consists entirely of the definition of “proximate cause,” and hence it would be included in the approximation.

Variable Number	<u>Text</u>
015	to determine the effect and value of the evidence
017	you must not
018	be influenced (by)
Any one of:	
019	sympathy
or	
020	prejudice
or	
021	passion

Variables 005 and 006 were excluded because they defined the judge's, not the jurors', duty. Variables 007, 012, and 013 were excluded because they stated obvious and trivial information. Variable 016, a connector, provided no substantive information.

In determining a subject's score for the approximation of BAJI 1.00, we added together the number of correct responses for the first seven derived variables, and if a subject had correctly paraphrased any one of variables 019, 020, or 021, we added an additional point. The sum was divided by eight. This approximation measure provides useful clues to the subjects' levels of comprehension, as well as one method for comparing an original text with a rewritten version. In essence, the approximation measure is a measure of "getting the gist."²⁷

The results for both the full performance and the approximation measures for each instruction are presented in Tables 1 and 2, respectively. As expected, the scores on the approximation measure were higher than the full performance scores, since some of the variables that subjects omitted were not essential to the important legal concepts embodied in the instructions. There was also a high correlation between the two measures, as reflected in Table 3. This correlation was expected, inasmuch as the approximation measure contains a subset of the data used to compute the full performance measure.

The performance scores suggest that the subjects did indeed have difficulty comprehending the instructions. The "mean of means" fraction of variables correct under the full performance measure was only 0.386, and even the approximation measure was only slightly better than half correct (0.540). Thus, the data provide some evidence that jury instructions are not adequately understood by the average juror.²⁸

27. On the other hand, the full performance measure is a diagnostic measure, which provides the information necessary to permit more comprehensible rewriting. The information about what people omitted or could not understand indicated that specific items should be modified. We needed all of the information available from the full performance measure in order to rewrite an entire instruction.

28. It should be recalled that BAJI 1.00, 1.01, and 1.02 were practice instructions always presented first. Thus, the relatively poor performance on these instructions may not be due to the nature of the instructions themselves, but rather may be attributable to a "newness" effect. However, in a trial, these instructions are also always presented first, and so the

However, the results should not be interpreted as definitive evidence that jurors or juries do not comprehend jury instructions. The ability of a juror to comprehend a given set of instructions depends on factors in addition to the linguistic construction and vocabulary of the instructions. The context provided by the trial itself may influence the comprehensibility of the charge. Specifically, factors such as the closing argument of counsel or a focus on specific issues might help to elucidate certain instructions. This experiment did not attempt to duplicate a trial situation, because the kind of information that we sought could be more readily obtained and replicated using the methodology we adopted, and in any case might be confounded in a simulated trial.²⁹

The results from Experiment I are largely descriptive, and cannot be fully interpreted without comparing them to results from a study of similar instructions that do not contain the problematic linguistic constructions. Consequently, based on the data obtained from this study, the instructions were rewritten, eliminating most of the apparently troublesome linguistic aspects, and the paraphrase task was repeated with new jurors in order to obtain the necessary comparison data. The results of this comparison are presented in the section on Experiment II.

C. *More Detailed Analyses of the Data*

In addition to the broad analyses of the data discussed above, several specialized analyses were performed to answer specific questions: Were there effects that were due to the order in which the instructions were given? How did the jurors' performance compare with attorneys' beliefs regarding the conceptual complexity of the instructions? Did sentence length affect comprehension? How was an individual's comprehension affected by such factors as education, occupation, age, sex, or other variables? This section presents the results of these analyses.

1. *Ordering Effects.* Since jury instructions tend to build on one another, the order in which the instructions are presented might affect a subject's performance. For example, a subject might conceivably be better able to paraphrase the definition of "proximate cause" if that instruction follows three or four other instructions that use the term. The presence of such ordering effects would complicate the analysis of the data.

When the performance for each instruction across subjects was analyzed for each order, however, no ordering effects were found beyond those expected by pure chance. Moreover, we discovered no apparent ordering

subjects' poor performance on them may in fact be an accurate reflection of how jurors actually process these instructions.

29. Even if this experiment had attempted to simulate a trial, the results from one simulated trial could not necessarily be generalized to other trials. It should also be noted that the apparent inability of individual jurors to adequately process jury instructions provides no reliable information as to how the deliberative process affects comprehension—i.e., how juror interaction in the jury room enhances or diminishes the comprehension of jury instructions.

effects when performance was computed across both subjects and instructions for each order: using the approximation measure, the differences in the mean fraction correct for each of the orders were not statistically significant (See Table 4). Thus, for the set of instructions used in this study, the methodology employed is relatively insensitive to order and therefore appears to provide a versatile means for measuring comprehension.

2. *Conceptual Complexity.* Many lawyers who acknowledge that jury instructions are difficult for the average juror to understand nonetheless do not believe that the instructions can be simplified to any great degree. This view reflects a widespread belief on the part of attorneys that it is the legal concept embodied in an instruction, and not the language, that causes comprehension difficulties; some legal concepts are so difficult for a lay person to understand that no amount of linguistic simplification will make an instruction comprehensible. We attempted to ascertain the validity of this belief.

As mentioned previously, experienced California trial attorneys had rated the conceptual complexity of fifty-two BAJI instructions on a scale of one (extremely easy to understand) to eleven (extremely difficult). The attorneys were asked to disregard any linguistic complexities (e.g., difficult vocabulary or grammatical and semantic constructions) and to rate how difficult the legal concept(s) contained in each instruction would be for the average juror to understand. Agreement among the attorneys for ratings of each instruction was very high.

We compared the attorneys' ratings of the conceptual difficulty of each instruction with subjects' performance on the instruction. When compared with the full performance measures, the ratings accounted for only 12% of the variability in subjects' scores ($r = -0.34$).³⁰ However, when the same comparison was made as a function of the subjects' education, we found that the attorneys' ratings correlated well with the performance of those subjects who had only a high school education, accounting for 39% of variability in scores ($r = -0.623$) (see Table 5).

Comparison of the attorneys' ratings with the approximation measures of the subjects' performance revealed that conceptual difficulty accounted for 21% of the variability in the subjects' scores ($r = -0.46$). This higher correlation for the approximation measure as opposed to the full performance

30. " $r = -0.34$ " represents the "correlation coefficient" for the comparison. For an explanation of the concept of correlation, and of the calculation of a correlation coefficient, see note d to the Tables, *infra*. The fact that the correlation coefficient is negative merely indicates that, in general, the higher the attorney rating (i.e., the greater the rated conceptual difficulty) the poorer the subjects' performance. It should be noted here that a correlation coefficient of -0.34 is relatively low (not statistically significant) and indicates only a very rough trend.

By squaring the correlation coefficient (i.e., $(-0.34)^2$ or 0.1156) we arrive at a figure indicating the percentage of variability in scores that can be accounted for by "conceptual complexity." In other words, only about 12% (11.56%) of the variation in the performance scores is attributable to the variation of the rated conceptual difficulty of the instructions; 88% of the variation is attributable to other factors.

measure is understandable, since the approximation measure was intended to measure performance only with respect to the important legal concepts in each instruction. Again, when the comparison was made as a function of education, the attorneys' ratings correlated well with the performance of those who had only a high school education, accounting for 35% of the variation in scores ($r = -0.588$). The correlation decreased, however, as the education level of the subjects increased (see Table 5). In short, if the raters viewed the average juror as one with a high school education, the rated conceptual difficulty of an instruction appears to be a rough predictor of performance. If, in fact, the conceptual complexity of an instruction limits its comprehensibility, we would expect to find that linguistic modification should not affect the comprehensibility of those instructions that were rated as conceptually complex. The results of Experiment II, however, throw the validity of that proposition into question.³¹

3. *Sentence Length.* There is another widespread belief, shared by certain educators,³² that comprehension is dramatically affected by sentence length; that, in general, the shorter the sentences, the more comprehensible the discourse. In the evaluation of written discourse, this belief has led to the adoption of various so-called readability formulas that incorporate as a primary parameter the length of the sentences in the discourse to be evaluated.³³ Readability formulas have gained such popularity that a number of jurisdictions have specified that insurance contracts must meet a minimum "readability" standard determined by a readability formula.³⁴

Although readability formulas are easy to use and certainly do indicate the presence of lengthy sentences, they cannot be considered measures of comprehensibility. Linguistic research has shown that sentences of the same length may vary greatly in actual comprehensibility. Consider the following pair of sentences:

- (1) "This morning I got up and brushed my teeth and got dressed and ate breakfast and went to work." (19 words)
- (2) "The boy whom the girl whom the gentleman in the white car hit kissed lives next door to me." (19 words)

Both these sentences would score the same on the most widely used readability formulas.³⁵ Yet the first is so simple that even a beginning reader

31. See text accompanying notes 78-80 *infra*.

32. See, e.g., R. FLESCH, *THE ART OF READABLE WRITING* (1949); R. FLESCH, *HOW TO WRITE PLAIN ENGLISH: A BOOK FOR LAWYERS AND CONSUMERS* (1979).

33. See note 10 *supra*. The other major parameter is usually word length, measured in syllables.

34. See DEL. CODE ANN. tit. 18, § 2741 (1976), which states in part that "all such automobile insurance policy forms shall have a total 'readability score' of 40 or more on the Flesch Scale"; 31 Pa. Code § 64.12 (1979), which provides in part that "[p]olicy forms should have a total readability score of 40 or more on the Flesch Scale."

35. See G. KLARE, *supra* note 10. Because the most widely used readability formulas measure only sentence length (number of words) and word length (number of syllables), and because both sentences have the same number of words and syllables, their scores would be identical.

could understand it, while the second is so difficult that it takes even a well-educated person some time to ascertain its meaning. It should suffice to say that it is the grammatical, semantic, and contextual complexity of discourse, not sentence length, that determines how difficult it will be for people to understand the discourse.

The results of our study clearly illustrate that sentence length has virtually no effect on subjects' performance. The instructions used in this study contained forty-four sentences, which ranged in length from seven³⁶ to seventy-two words.³⁷ Each sentence consisted of one or more variables, and so subjects' performance on a sentence was computed by averaging their performance on all the variables that made up a sentence. Sentence length accounted for only 1.7% of the variation in subjects' scores ($r = -0.13$). Similarly, we compared the mean sentence length per instruction with the full performance measure for that instruction (see Table 6). Here, too, the correlation between sentence length and comprehension was very low, with sentence length accounting for less than 3% of the variability in subjects' scores ($r = -0.162$). This means that sentence length had almost no effect on subjects' ability to paraphrase correctly any given piece of discourse. Actually applying the Flesch readability formula to the jury instructions and comparing the readability scores to subjects' performance, there was a nonsignificant *negative* correlation between readability score and performance, as measured by either full performance or approximation measure.³⁸ It is also interesting to note that the correlation between *instruction* length, measured in either words or constituent units (variables), and performance was also low. In other words, the length of the instruction as a whole similarly had little effect on its comprehensibility.

4. *Demographic Analysis.* As noted above, all subjects were asked to complete a demographic questionnaire after finishing the paraphrase task. The questionnaire sought information about the subjects' education, occupation, legal training (formal or informal, and amount), prior jury service, past military service, age, sex, native language, and a number of other variables. A multiple regression analysis³⁹ revealed that the only factor that consistently and significantly correlated with performance was the amount of education that a subject had had. Tables 7 and 8 set forth the

36. BAJI 2.40.

37. BAJI 3.01. Since we used traditional criteria for a written sentence, our results can be compared to results obtained using a readability formula. However, it should be noted that readability formulas are not often used to grade materials that are written to be read aloud.

38. The Flesch formula is constructed so that the higher the readability score, the easier ("more readable") the material. Thus, the results of our comparisons indicate that there is no real relationship between comprehensibility as measured by our very precise methodology, and scores on the Flesch formula. In fact, there is even a slight, though nonsignificant tendency in the opposite direction: the better the Flesch readability score, the poorer the subjects' performance.

39. A multiple regression analysis is a statistical technique for determining which factors (independent variables, such as, in our case, education, age, occupation, sex, etc.) account for how much of the variation in the performance scores (the dependent variable). For a fuller discussion, see M. EZEKIEL & K. FOX, *METHODS OF CORRELATION AND REGRESSION ANALYSIS* (3d ed. 1959).

full performance and approximation scores, respectively, for each instruction, as a function of the subjects' education. The Tables indicate, as might be expected, that comprehension rose as education level rose.

D. *Linguistic Analysis by Construction*

A primary purpose of this experiment was to determine how various linguistic constructions affect comprehension, and which constructions appear to impede it. In this section, we will describe a variety of specific linguistic constructions and indicate the degree to which each construction was comprehended. The constructions presented below were selected for analysis because either: (1) a review of the data on a variable-by-variable basis indicated that the construction was consistently associated with either a decrease or an increase in the subjects' performance; or (2) the construction had been the subject of prior psycholinguistic research indicating that under certain circumstances it might impede (or enhance) comprehension. Table 9 sets forth most of the constructions to be discussed and the degree to which each was comprehended.⁴⁰

1. *Nominalizations*. A nominalization is a noun that has been constructed from a verb. It may be created by simply adding the present participle ending "-ing" to the verb stem—as in "the doing (of)"—or it may involve the somewhat more complicated process of adding "-tion" or "-al" to the verb—as in "the investigation of" instead of "when we investigated," or "the removal of" instead of "when we removed." (There are several other ways of creating nominalizations, but these are the most common.) Linguistic theory indicates that nominalizations are more difficult to process than their equivalent verb forms, for a number of reasons. First, the work of a number of linguists suggests that verbs and verb phrases are more basic than certain classes of nouns, and that anything that makes a verb less verb-like and more noun-like creates abstraction.⁴¹ Second, nominalizations may be difficult because they generally replace entire subordinate clauses; for example, the subordinate clause "When you are incorporating the material into a chapter, make sure . . ." would be transformed into the nominalization "The incorporation of the material into a chapter necessitates . . ." By eliminating the true subject of the sentence—the "doer" of the action—the nominalized construction makes the sentence vague, impersonal, and hard to reconstruct. Reducing the grammatical complexity of the surface form of a sentence by shortening a whole subordinate clause into a single nominal usually increases the complexity of the deeper grammatical and semantic structure. The meaning of the sentence becomes less clear, and the mind must work harder to decode it.

40. Certain constructions—such as discourse structures and subordinate clause embeddings—could not be conveniently categorized, and hence are not included in the table even though they will also be discussed.

41. See McCawley, *Where Do Noun Phrases Come From?* in *READINGS IN ENGLISH TRANSFORMATIONAL GRAMMAR* 166 (R. Jacobs & P. Rosenbaum eds. 1970); R. Lees, *THE GRAMMAR OF ENGLISH NOMINALIZATIONS* (1968).

The fourteen BAJI instructions analyzed contained fourteen nominalizations of various sorts. For example, BAJI 1.02 instructs the jurors to regard a fact "as being conclusively proved as to the party or parties making the stipulation or admission."⁴² No subject correctly paraphrased this variable. BAJI 2.21 refers to "failure of recollection"—an exquisite double nominalization. Only thirty-four percent of the subjects paraphrased this correctly.

In order to obtain a measure of the subjects' performance on the nominalized forms, we computed each subject's correct score on the fourteen variables containing nominalizations. The mean of the scores was then computed and compared to the mean for all the other variables.⁴³ As indicated in Table 9, the mean percentage of correct paraphrasings across subjects for the fourteen variables containing nominalizations was 28.6%. This was significantly lower—at the .05 level of statistical significance⁴⁴—than the mean for all the other variables.

2. *Prepositional Phrases—"as to."* The BAJI instructions are replete with phrases introduced by "as to." For example, the following sentence appears in BAJI 1.02: "As to any question to which an objection was sustained, you must not speculate as to what the answer might have been or as to the reason for the objection." BAJI 1.01 contains the cryptic admonition, "The order in which the instructions are given has no significance as to their relative importance."

Unlike prepositional phrases beginning with "in," "on," "at," "for," "with," etc., phrases beginning with "as to" seem vague. The words "as to" do not refer to a time, location, or purpose, but rather serve as a somewhat ambiguous link between parts of speech. For example, it is clearer and more direct to say "you must not speculate about what the answer might have been," than to use "as to." Similarly, a word such as "concerning" or even a phrase such as "with regard to" has a more definite meaning than "as to."

Using the scoring method outlined above,⁴⁵ we found that the score for the eight "as to" prepositional phrases in the instructions was 28% correct. Although the sample size is small, the data nonetheless indicate a trend. It is possible that the vague preposition "as to" acts as a signal to the listener that what follows is unimportant, whether or not it is.

42. It should be noted that "stipulation," in addition to being a nominalization, is a somewhat technical legal term, or "lexical item."

43. In order to determine whether or not the mean score correct for a given construction differed significantly from the mean score correct for all the other variables, the "quasi-Fischer ratio" (F') was used. The quasi-Fischer ratio is an extremely conservative method of measuring statistical significance. Thus, a finding of statistical significance through use of the quasi-Fischer ratio is even more meaningful than a finding of statistical significance through use of another method.

44. For an explanation of the concept of statistical significance, see note c to the Tables *infra*.

45. See text accompanying notes 43 & 44 *supra*.

3. *Misplaced Phrases.* The instructions include several instances of phrases (mostly prepositional) inserted into the midst of otherwise normal clauses, or otherwise misplaced, so that they either break up the continuity of the clause or create ambiguity. Sometimes they are merely quaint, but sometimes they act in much the same way as that old chestnut from ancient grammar books: "The man ran down the street shouting 'Fire' in his pants."

For example, BAJI 1.01 begins with the words "*If in these instructions any rule, direction or idea is repeated . . .*" The placement of the italicized phrase directly after the word "if" confused the majority of the subjects. Since the word "if" is normally followed by the subject of the sentence, most listeners perceived the noun in the misplaced phrase—"instructions"—as the subject of the sentence, and paraphrased it as "If these instructions are repeated . . ." ⁴⁶

BAJI 3.75 instructs that "A proximate cause . . . is a cause which, in *natural and continuous sequence*, produces the injury . . ." Because the italicized phrase *precedes* the verb it is intended to modify, the construction leaves the listener with the impression that the cause itself is in a natural and continuous sequence. Inasmuch as a single "cause" cannot be in a continuous sequence, the listener is befuddled.

For the nine misplaced phrases in our instructions, the mean fraction correct was only 24%. Furthermore, for two of the items an inordinate number of jurors actually misunderstood—as opposed to omitting—the phrase. This apparent inability to adequately process such misplaced phrases has ramifications beyond the items themselves: because they disrupt the continuity of the sentence, misplaced phrases may contribute to poor performance on the entire instruction.

4. *"Whiz" and Complement Deletion.* The instructions contain several examples of subordinate clauses that are missing relative pronouns (that, which, who, etc.) and "copula" verbs ("Be" verbs, such as "was," "is," "am," "are," etc.). These include the phrases "questions of fact * submitted to you," from BAJI 1.00, and "any statement of counsel * made during the trial," from BAJI 1.02. This grammatical phenomenon is known to linguists as "whiz" (short for "which is") deletion, because what is missing is the implied phrase "which is." While whiz deletion is normal in English, in some instances it appears to add to the listener's processing load. Because some of the grammatical information is missing, the mind has to work harder to reconstruct it. The same is true for clauses from which the complementizers "that" or "which" have been omitted, such as "if you are convinced * it is erroneous" from BAJI 15.30. We therefore hypothesized that such constructions would not be well understood. For the twelve instances of whiz or complement deletion in the instructions, the mean score was 24.5% correct, thus tending to bear out this hypothesis.

46. Only five subjects correctly made reference to any of the three subject nouns; the rest omitted them.

5. *Lexical Items.* Probably the most obvious difference between legal language and ordinary discourse is the technical vocabulary of the law. As one would expect, those items containing technical (legal) terms, or infrequently used words or phrases, were omitted or incorrectly paraphrased a large percentage of the time. We looked in the instructions either for variables that consisted entirely of a legal term, an unfamiliar expression, or an uncommon word (as determined by frequency dictionaries), or for longer variables in which the predominant feature was the legal term, unfamiliar expression, or uncommon word. There were thirty-six such instances, including "credibility," "proximate cause," "deem," and "stipulate." The mean correct for these was 32%. There were a few surprises, however: "discrepancy(ies)" was *correctly* paraphrased 89% of the time and "imputed" was *correctly* paraphrased 54% of the time.

6. *Modals.* A linguistic construction that apparently enhances jurors' comprehension is the use of "modal" verbs,⁴⁷ specifically "must," "should," and "may," in phrases relating to the jurors' duties. Such modal constructions, even in the negative, produced a correct paraphrase rate of 57% in the sixteen such items in the instructions. This is significantly higher than the mean rate for the instructions as a whole, excluding modals (38.2%). Such a finding is understandable, since one would expect jurors to interpret as of primary importance those portions of the jury instructions that tell them what to do. Thus, they would probably be mentally tuned to such modal constructions as "you *must* not be influenced . . ." (BAJI 1.00) or "you *should* consider the qualifications . . ." (BAJI 2.40), which are quite strong in their assertion of the juror's duty.

7. *Negatives.* Much psycholinguistic research regarding negatives of various kinds has shown that negatives⁴⁸ apparently take longer to process and cause more comprehension errors than similar ideas stated in positive form.⁴⁹ Research on multiple negatives has indicated that as the number of negatives in a sentence increases, processing time and error rate similarly increase. Our own findings in this study show that for single negatives (excluding modals stating the jurors' duty) the correct paraphrase rate was about the same as the correct paraphrase rate for the instructions as a

47. Modals are a class of verbs, including "must," "may," "might," "should," "can," and "could," that are used as auxiliaries to other verbs and that carry meanings relating to ability, obligation, and permission.

48. Linguistic analysis and psycholinguistic research have shown that negatives are not only words such as "no," "not," "none," or prefixes such as "un-" or "in-." Conjunctions such as "unless" and "except" carry negative meaning, and function as negatives in a sentence. The mind also treats words like "absent," "excluding," "avoid," "less," "minority," and "few" as negatives, as it does such prefixes as "mis-," "il-," and "ir-." See authorities cited in note 49 *infra*.

49. See Clark & Chase, *On the Process of Comparing Sentences Against Pictures*, 3 *COGNITIVE PSYCHOLOGY* 472 (1972); Just & Carpenter, *Comprehension of Negation with Qualification*, 10 *J. VERBAL LEARNING & VERBAL BEHAVIOR* 244 (1971); Just & Clark, *Drawing Inferences from the Presuppositions and Implications of Affirmative and Negative Sentences*, 12 *J. VERBAL LEARNING & VERBAL BEHAVIOR* 21 (1973); P. Wason, *Psychological Aspects of Negation—An Experimental Enquiry and Some Practical Applications* (1962) (in-house publication of Communication Research Centre, University College, London).

whole (37% for single negatives; 39% for the instructions as a whole). Thus, single negatives did not appear to be an outstanding cause of confusion.

In contrast, for the six instances of double or triple negatives within a single sentence—such as “not to avoid it” (BAJI 3.11), and “without which the injury would not have occurred” (BAJI 3.75)—the average correct score was lower: 26%. However, because of the small sample size, the difference was not statistically significant. One interesting example of multiple negatives is the following from BAJI 2.21: “innocent *mist*collection is not *un*-common.” Only 26% of the subjects correctly paraphrased it.

8. *Passives*. One characteristic of legal language appears to be a high proportion of passive sentences. Although many people concerned with simplifying legal prose believe that the passive voice, when used indiscriminately, impedes comprehension,⁵⁰ psycholinguistic research regarding the comprehensibility of passive constructions has produced equivocal results.⁵¹

There were thirty-five instances of passive constructions in the jury instructions tested. These included such phrases as “no emphasis thereon is intended by me . . .” (BAJI 1.01), and “the conduct reasonably *could be avoided* . . .” (BAJI 3.11). As indicated in Table 9, the mean score for the passive constructions was almost identical to the mean score for the remaining 238 variables. This, of course, does not mean that the passive voice is necessarily easy to process. Many of these 238 other variables were themselves difficult to understand. However, it does indicate that passives, when viewed as a class, are not an outstanding source of confusion.

We also examined whether the type of passive or its location affected comprehension. We had originally believed that a “truncated passive”—one in which the agent (*i.e.*, the “doer” of the action) is omitted⁵²—would be more difficult to process than a full passive, because a truncated passive provides the listener with less information and forces the listener to “fill in the blanks” in order to decode the message. The following passage from BAJI 1.02 provides a good example of a truncated passive: “A question is not evidence and *may be considered* only as it supplies meaning to the answer.” The sentence gives no explicit indication of *who* may do the considering. As Table 9 indicates, however, truncated passives were actually better understood than full passives—although the difference was not statistically significant.

While the type of passive did not appear to affect comprehension sig-

50. See, e.g., Wydick, *Plain English for Lawyers*, 66 CALIF. L. REV. 727, 746-47 (1978).

51. Since passives have a different focus from their corresponding active sentences, in some situations the passive focus may be necessary, and hence easier to process. See Hornby, *Surface Structure and Presupposition*, 13 J. VERBAL LEARNING & VERBAL BEHAVIOR 510 (1974); Hupet & Le Bouedec, *Definiteness and Voice in the Interpretation of Active and Passive Sentences*, 27 Q.J. EXPERIMENTAL PSYCHOLOGY 323 (1975); Huttenlocher & Strauss, *Comprehension and a Statement's Relation to the Situation It Describes*, 7 J. VERBAL LEARNING & VERBAL BEHAVIOR 300 (1968); Huttenlocher & Weiner, *Comprehension of Instructions in Varying Contexts*, 2 COGNITIVE PSYCHOLOGY 369 (1971).

52. E.g., “John was bitten.” The missing information is “by whom?”

nificantly, the location of the passive did. Passive constructions in subordinate clauses⁵³ produced significantly lower scores than all the other variables—27% correct versus 40% correct—and also significantly lower scores than passive constructions in main clauses—27% correct versus 53.5% correct. On the other hand, the paraphrase scores associated with main clause passives were significantly higher than the scores for all other variables—53.5% correct versus 38.4% correct.⁵⁴ Thus, there is some evidence that passive constructions, when properly used and not obscured in subordinate clauses, do not impede comprehension.⁵⁵

9. *Word Lists.* In the search for precision, legal language often uses three or four words where one will do; an example is the ritual use in wills of the terms "give, bequeath, and devise." Our BAJI instructions also manifested this propensity: seven of the instructions contained a total of twelve lists of words—verbs, nouns, or attributes. Where the list contained more than two items, virtually no subject paraphrased all the items. Generally, subjects tended to pick one or at best two of the items in the list. In order to determine performance on the word lists, we assigned a score of correct if the subject correctly paraphrased any one item. For the twelve lists, the mean rate of correct paraphrasing was 43.1%.⁵⁶

There were a number of other interesting findings. If the structure of the sentence was such that there was little focus, or the wrong focus, on the list of items, subjects were unable to recall even one of the items. For example, in BAJI 1.01 the list "any rule, direction or idea" followed the misplaced phrase "in these instructions." Only four of the thirty-five subjects were able to paraphrase any of the items on the list correctly. Nor did the length of a list necessarily guarantee that even one item on it would be correctly paraphrased. For instance, for the list "knowledge, skill, experience, training or education," in BAJI 2.40, 46% of the subjects were unable to paraphrase even one of the items correctly.

10. *Discourse Structure.* Comprehension of discourse can depend on how the individual sentences are organized relative to each other and on the coherence among sentences; this overall organization is referred to as "discourse structure." A review of the data revealed that several instructions

53. A notable example appears in BAJI 1.02, which provides: "You must never speculate to be true any insinuation suggested by a question asked a witness." This sentence uses two passives in subordinate clauses—the first in an adjective clause modifying "insinuation"; the second in an adjective clause modifying "question."

54. These differences persisted even across types of passive. Truncated passives located in main clauses produced significantly higher scores than truncated passives located in subordinate clauses (55.3% correct versus 31.6% correct), and full passives located in main clauses produced better scores than full passives located in subordinate clauses (45.3% correct versus 17.3% correct). In the latter case, however, the differences were not statistically significant because of the small sample size.

55. For that reason, a number of passive constructions were retained, and some new passive constructions added, when the instructions were modified for the second experiment. See text accompanying note 89 *infra*.

56. However, because of the manner in which it was computed, this score cannot be compared with the mean for all items.

had problems at this level. Several instructions were poorly organized, as evidenced by an exceptionally poor full performance measure. BAJI 1.02, for example, had no overall "plan," scattering information concerning the treatment of "evidence" throughout the instruction; information about lawyers' "questions" appeared in three different places. Some subjects even commented on the disorganization of the instruction and had great trouble paraphrasing it. The mean full performance score across subjects was 25.7% correct.

One mechanism for organizing and highlighting the structure of discourse is numbering each of the ideas, and informing the listener beforehand of the number of ideas that will be covered.⁵⁷ BAJI 1.02 is a prime candidate for numbering, as is BAJI 1.01, which contains three separate ideas that are not clearly delineated. The mean full performance score across subjects for BAJI 1.01 was 23% correct.

However, numbering will not remedy all discourse-level problems. For example, BAJI 3.01 enumerates the four conditions under which the plaintiff is entitled to a verdict. These conditions are stated twice: once in the first paragraph, discursively, as attributes of the plaintiff, and once in the second paragraph in list form, as the hypothesis of a conditional. The subjects found the numbering useful, but many were totally confused by the unnecessary redundancy. Twenty-three percent thought the two paragraphs were saying different, even contradictory, things. The redundancy can be thought of as a violation of Grice's Maxim of Quantity: Say no more and no less than is necessary to convey the message.⁵⁸

11. *Embeddings.* The BAJI instructions, like much written legal language, contain a substantial number of "embeddings"—the use of numerous subordinate clauses within one sentence. In the fourteen instructions, forty-three sentences had one main clause and from zero to nine subordinate clauses. Linguistic theory would predict that the types of embeddings used should influence comprehension more than the absolute number of embeddings. However, if a writer indiscriminately embeds subordinate clauses, the likelihood that a difficult to comprehend embedding will be used in a sentence should increase as the number of embeddings increases. The results of the experiment revealed a high negative correlation between performance and the number of embeddings used: as the number of embeddings increased, comprehension decreased ($r = -0.98$).

A closer examination of the data showed that the types of embeddings used in the instructions probably affected performance more than the number of embeddings. Certain sentences contained types and combinations of embeddings that linguistic theory predicts would be difficult to compre-

57. Several of the modified instructions use this technique of numbering the ideas. See, e.g., Modified Instructions 1.01, 1.02, 3.11, and 3.01, in section IV *infra*.

58. H.P. Grice, a noted philosopher of language, has formulated a series of maxims concerning effective communication. See Grice, *Logic and Conversation*, in *SPEECH ACTS*, 3 SYNTAX AND SEMANTICS 41, 45 (P. Cole & J. Morgan eds. 1975).

hend,⁵⁹ and, indeed, the performance on these sentences was poor. BAJI 1.02, for example, reads: "You must never speculate to be true any insinuation suggested by a question asked a witness" (mean full performance score 22% correct), and BAJI 3.11 states:

One test that is helpful in determining whether or not a person was negligent is to ask and answer whether or not, if a person of ordinary prudence had been in the same situation and possessed of the same knowledge, he would have foreseen or anticipated that someone might have been injured by or as a result of his action or inaction.

(mean full performance score 32% correct). This latter sentence contains nine subordinate clauses; in this instance the sheer number of subordinate clauses could have affected subjects' memory and, hence, comprehension.

III. THE SECOND EXPERIMENT

The results of the first experiment suggest that standard jury instructions may not be adequately understood, and further indicate that specific linguistic constructions may be at the root of at least some of the comprehension problems. In the second experiment, the fourteen jury instructions used in Experiment I were rewritten by systematically eliminating, where appropriate, the various hypothesized problematic constructions. The modified instructions were then tested for comprehensibility, and the results show clearly that the modifications yielded improved performance.

A. Description of the Experiment

Based on the results of Experiment I, the jury instructions were modified in the following general ways:

—Where appropriate, we eliminated passives and used the active voice instead.

59. While multiple clause embedding may in general create comprehension problems, certain forms of embedding appear to be more difficult to understand and process than others. Among the most difficult are "self-embedding," or "center-embedding" clauses, illustrated by the sentence: "The boy (whom the girl [whom the man in the red car hit] kissed) lives next door to me." Here, there are two subordinate clauses; one clause is enclosed (centrally embedded) within the next one like a Chinese puzzle, as indicated by the parentheses and brackets in the sentence.

Another type of embedding that is difficult to process, if more than one such embedding is used at a time, is called "left-branching"; for example, "The fact that (the fact that [John loves Marcia] bothered Alex) surprised nobody." Again, there are three clauses in this sentence (corresponding to the three verbs); the two subordinate clauses are to the left of the main verb "surprised."

Instances of such relatively incomprehensible constructions are frequently found in standardized jury instructions. For example, the first sentence of California's "Contributory Negligence—Forgetfulness of Known Danger" instruction (BAJI 3.51) is: "Whether or not it is negligence for one to proceed into a dangerous situation of which he had previous knowledge is a question of fact." As can be noted, all of the clause embedding is to the left of the main verb phrase "is a question of fact." Another way of looking at this type of sentence is that its subject (the subject of the predicate "is a question of fact") is a sentence consisting of three subordinate clauses. By the time we get to the predicate, we have probably forgotten the subject. From the evidence of empirical studies, it can be hypothesized that this instruction will be difficult to understand and remember.

—In most instances of “whiz deletion,” we reinstated the relative pronoun and copula. Similarly, if a complement had been deleted, we reinstated it.

—We eliminated most instances of nominalization, substituting, where possible, the more verb-like form.

—We eliminated multiple negatives and vague prepositional phrases (such as those beginning with “as to”).

—We put misplaced phrases in their proper places.

—We changed several difficult lexical items to more commonly used words.

—We eliminated embeddings that appeared to be difficult to process.

—We changed the format of some instructions or parts of instructions in order to improve the organization, to provide the listener with more context, or to place the correct focus on an idea.

—We reduced strings of items or attributes to lists of one or two.

Table 10 summarizes the frequency of problematic constructions in the original instructions and in the modified instructions. The changes that we made were not substantial; wherever possible we retained the format or ordering of the original. Therefore, the rewritten instructions should not be viewed as model instructions, or as necessarily the best possible rewrites, but as instructions that have been systematically altered for experimental purposes.

We created two treatment groups for Experiment II. Each group received seven original and seven modified instructions; the seven original instructions received by one treatment group corresponded to the seven modified instructions received by the other treatment group, and vice versa.⁶⁰ However, we did not randomly select the original and modified

60. This design provided an enormous amount of flexibility, combining the advantages of both a “repeated measures” and an “independent groups” design. In a repeated measures design, each subject would be presented with both the original and the modified versions of all of the instructions. Such a design would eliminate the possibility that differences in results were due to differences in subject characteristics (e.g., educational level, ability to perform a paraphrase task, exposure to legal language), but it would be impractical because of “practice effects”: information gleaned from paraphrasing the original instruction would affect the paraphrase scores for the modified instructions, or vice versa. In an independent groups design, two groups would be used; one would paraphrase only modified instructions, and the other would paraphrase only original instructions. However, the independent groups design requires a large sample size in order to maximize the probability that the two groups are comparable; if employed on a relatively small sample, it would be difficult to assure that differences in performance were indeed the result of the modifications in the instructions and not the result of differences in group characteristics. Unfortunately, the number of subjects available to us was limited. Furthermore, it is not practical to use a paraphrase task on large subject samples because only one subject can be tested at a time.

The methodology chosen, like an independent groups design, minimizes the effect of differences in group characteristics. Thus, if performance on all modified instructions is significantly better than performance on the corresponding original instructions, differences between each group's characteristics cannot be responsible for the overall improvement.

This design also has features of a repeated measures design. By exposing each subject to seven original instructions, it is possible to predict, from the data obtained from Experiment I,

instructions to be presented to each group. Because lexical changes made in one modified instruction necessarily appeared in several other modified instructions (for example, "proximate cause" was changed to "legal cause"), random groupings of original and modified instructions would have resulted in two different terms for the same idea within a single set of instructions. This could have confounded the results, since the subjects would have had the added burden of dealing with two terms without any indication that they meant the same thing. In addition, we wanted to ensure that each treatment group received original instructions that were, on the average, equally difficult to paraphrase (as measured by the results from Experiment I). Finally, we wanted to ensure that the fourteen instructions received by each treatment group had the same average length, so that differences in results could not be attributed to differences in the number of words heard by each treatment group.

For these reasons, we grouped the twenty-eight instructions (fourteen original and fourteen modified) as indicated in Table 11. Two master tapes were created, one for each treatment group.⁶¹ The order in which the instructions were taped was the same as in Table 3. As in Experiment I, each instruction was recorded twice in succession, with a five-second pause between each recording.

The subjects were forty-eight persons called for jury duty in Prince Georges County, Maryland, during the winter of 1977. As in the first experiment, participants were randomly selected from a panel of prospective jurors. Participation was voluntary, but again none of the jurors selected refused to participate. As in the first experiment, jurors were selected before they had served on a jury, in order to ensure that none of them had had recent experience with jury instructions.⁶²

The subjects were randomly divided into the two equal-sized treatment groups, and each group received one of the two versions of the master tape. In general, the procedure was identical to that used in the first experiment: the subjects were presented with the same directions, the same picture, and the same fact situation as in Experiment I.⁶³

how that subject would have performed on the other seven original instructions had he received them. In this way it is possible to compare each subject's performance on his or her seven modified instructions with his or her predicted performance on the corresponding original instructions. It is also possible to compare predicted scores on the original instructions for one set of subjects with the actual scores for the other set of subjects on the same instructions, and to pair subjects by their ability to perform a paraphrase task. Thus, the scores for original and modified instructions for each matched pair of subjects can be compared, and any changes should be due to differences in the instructions.

61. The 14 modified instructions were recorded by the same male attorney who had recorded the original 14; his speed of delivery was identical for both the original and modified instructions (131.6 words per minute).

62. The education levels of the 48 subjects who participated in Experiment II were significantly lower than those of the 35 subjects who had participated in the first experiment. The mean level of education per subject in Experiment I was 15.06 years, while the mean level of education in Experiment II was 14.00. The mean educational level of the subjects in each of the two groups in Experiment II was the same (14.00 years).

63. For a description of the procedure, see text accompanying notes 20-23 *supra*.

B. Analysis of the Data

We used the same procedures we had used for Experiment I to analyze the data from Experiment II.⁶⁴ The original and modified instructions were broken down into constituent units (variables); the breakdowns for the original instructions were the same as those used in the first experiment. Jurors' paraphrases were then compared to the breakdowns in order to ascertain whether they had paraphrased the variables correctly (explicitly or implicitly), omitted them, or paraphrased them incorrectly.⁶⁵

The data from Experiment II were analyzed from a number of perspectives. The broadest view involved comparing jurors' performance on the modified instructions with performances on the original instructions.⁶⁶ In addition, we reexamined the question of whether the attorneys' ratings of the conceptual difficulty of the instructions were a reliable means of predicting the subjects' performance.

1. *Overall Performance.* We measured the subjects' overall performance in three ways, using (1) a full performance measure, (2) an approximation (gist) measure, and (3) a concordance measure. Even though each measure provides different information, the results were quite consistent across all three: the subjects' overall performance was significantly and substantially better on the modified instructions than on the original instructions. The mean of means percentage improvement across all instructions ranged from 41%, as determined by the full performance measure, to 35% as determined by the approximation measure.

a. *Full Performance Measure.* The full performance measure, as described above,⁶⁷ is the least informative of the three measures, because it simply measures the subjects' performance (*i.e.*, the mean fraction of variables correctly paraphrased) across *all* variables, without taking into account the relative importance of a variable to the message of the instruction. In rewriting three of the instructions (BAJI 3.50, 3.71, and 4.31), we included a certain amount of introductory filler material to provide a context for the listener, and because this material was not essential to the meaning of the instruction, we hypothesized that the jurors would be unlikely to paraphrase it. Thus, the full performance measures for those modified instructions might well be lower than for the original instructions. However, we did hypothesize that there would be improvement in the full

64. See text accompanying notes 24-27 *supra*.

65. This analysis was again performed by two raters acting in concert.

66. This included a comparison of full performance measures and approximation measures, and a "concordance" measure. See text accompanying notes 67-77 *infra*. Although these overall comparisons showed that linguistic modification of jury instructions produces substantially better scores, they did not show which kinds of changes produced the improvement. That information was obtained from a second set of analyses, which examined the effect on subjects' performance of specific linguistic changes. See text accompanying notes 81-96 *infra*. A third method of analysis involved examining each pair of instructions (original and modified) in order to ascertain which modifications accounted for the improvement in paraphrase scores. See section IV *infra*.

67. See text following note 25 *supra*.

performance measure for the other eleven modified instructions. In fact, we anticipated that the improvement in modified BAJI 3.01 would be amplified for the very reasons that BAJI 3.50, 3.71, and 4.31 would probably not show an increase in full performance measure. This is because in re-writing BAJI 3.01, we removed redundancy that we felt was potentially confusing in the original.

The results of the full performance measure are set forth in Table 12. There was improvement in twelve of the fourteen instructions; in nine of these, the mean improvement was statistically significant, and in one case the improvement was marginally significant. The mean of the means for correct paraphrasing of the instructions increased from 0.319 to 0.427.⁶⁸ For the reasons stated earlier, we were not surprised to find that the full performance scores for BAJI 3.50 and 3.71 declined, and that the improvement in BAJI 3.01 was abnormally high.⁶⁹

In addition to measuring change in performance by instruction across subjects, we also measured the change in performance by subjects across instructions. Using the data from Experiment I, it was possible to predict, from each subject's mean full performance score for the seven original versions heard, what that subject's mean full performance score would have been on the other seven original instructions not heard.⁷⁰ The change in performance (from original to modified instruction) by subject was then determined by comparing the predicted mean score for those seven original instructions with the subject's actual mean score for the corresponding seven modified instructions. For forty-one of the forty-eight subjects

68. This increase was statistically significant at the .001 level.

69. The reader may also note that the scores for the original instructions used in Experiment II are substantially lower than the scores in Experiment I. It would appear that this difference is primarily a result of the significantly lower educational level of the jurors in Experiment II. See note 62 *supra*. Indeed, using the mean scores for each instruction from Experiment I, broken down by education level (in a manner similar to Table 7), we can predict what the scores for each original instruction would have been if the subjects in Experiment I had had the same level of education as the subjects in Experiment II. Table 13 compares those predicted scores with the actual scores in Experiment II.

The predicted mean score if the subjects in Experiment I had had the same educational distribution as those in Experiment II is 0.328; the actual mean score of subjects in Experiment II is 0.318. These scores are substantially the same. This means that the actual differences in performance between the subjects in Experiment I and the subjects in Experiment II are principally due to the difference in the level of education between the two groups.

70. The data from Experiment I revealed that a subject's mean full performance score on the original versions of BAJI 1.00, 1.01, 2.21, 2.40, 3.11, 15.22, and 15.30 correlated highly with a subject's mean full performance score on the remaining seven instructions ($r = 0.94$). Hence, by using the data from Experiment I, we were able to generate a highly predictive linear equation relating a subject's mean score for one group of seven instructions with that subject's mean score for the other group of seven. The equation is as follows:

Let $\bar{Q} 1$ = a subject's mean full performance score on BAJI 1.00, 1.01, 2.21, 2.40, 3.11, 15.22, and 15.30;

and $\bar{Q} 2$ = a subject's mean full performance score on BAJI 1.02, 3.01(A), 3.50, 3.71, 3.75, 4.30, and 4.31;

then $\bar{Q} 2' = 1.32 \bar{Q} 1 - 12.76 (\pm 6.38)$,

and $\bar{Q} 1' = 0.76 \bar{Q} 2 + 9.67 (\pm 4.83)$.

(The above equations are in percentage units, generated using the statistical procedure known as the least squares solution. See L. HOBOWITZ, *ELEMENTS OF STATISTICS FOR PSYCHOLOGY AND EDUCATION* 319-40 (1974).)

tested (85%), the actual full performance mean on the modified instructions was greater than the predicted mean for the corresponding seven original instructions.⁷¹ For one subject there was no difference, and for six subjects the full performance mean for the modified instructions was less than the predicted mean for the corresponding original instructions.

b. *Approximation Measure.* As explained above,⁷² for each original instruction we devised an approximation, which consisted of only the most essential variables necessary for an accurate statement of the law. Using the same rules that we had used to generate the original approximations, we generated approximations for each of the modified instructions. The approximation measure is far more useful than the full performance measure for comparing performance on original and modified instructions, because it does not have the same artificial skewing effects as the full performance measure. We therefore hypothesized that there would be improvement for each of the fourteen modified instructions.

Table 14 sets forth the approximation scores for each of the original and modified instructions in Experiment II. As anticipated, there was improvement in the approximation scores for each of the fourteen modified instructions; nine of these improvements were statistically significant. The mean of the mean scores for the instructions increased from 0.447 to 0.592.⁷³

c. *Concordance Measure.* Another useful comparison between the scores on the original and modified instructions is between variables that have the same meaning in the two sets. For each pair of instructions, we drew up a "concordance," that is to say, a list of variables in the modified instruction that have semantic counterparts in the original instruction.⁷⁴ The original instructions had been broken down into 273 variables (meaning units); the modified instructions were broken down into 295 variables. Of these variables, 179 were common to both; that is to say, 179 items had the same meaning in both the original and modified instructions.⁷⁵

The results of the concordance measure are presented in Table 15. There was a statistically significant improvement for the common items in nine of the instructions, and one marginally significant improvement.⁷⁶ There was also consistent improvement, on a subject-by-subject basis, across instructions. As noted above, based on each subject's mean score for the seven original instructions received in Experiment II, we could predict what

71. $t(47df) = 6.99, p < .001$.

72. See text preceding note 26 *supra*.

73. This increase was statistically significant at the .001 level.

74. Some trivial or redundant variables were retained when the instructions were rewritten; however, since the same essential ideas necessarily appear in both, a comparison of concordance scores is a valid way of measuring change in scores from original to modified instructions.

75. For example, in Modified BAJI 1.00, the phrase "I am now going to tell you the laws" corresponds to the original, "[I] instruct you in the law"; the modified phrase "It is your job and no one else's . . ." corresponds to the original "it is your exclusive duty . . ."

76. This measure provides the first real empirical indication that the improvement in performance on the modified instructions was due to the linguistic alterations.

his or her mean score would have been for the seven original instructions that he or she did not receive. In this way, subjects in Treatment Group 1 could be paired by performance with subjects in Treatment Group 2.

Of the twenty-four subjects in each treatment group, twenty pairs could be created.⁷⁷ For each pair of subjects matched in this way, we compared the total number of concordance variables that one subject in the pair paraphrased correctly for the modified instructions with the other subject's score for the original instructions, and vice versa. This resulted in forty comparisons. The results, which are set forth in Table 16, reveal that in only two of the forty cases did a subject have a lower concordance score for the modified instructions. Thirty-six of the forty cases (90%) showed improvement for the modified instructions.

2. *Conceptual Complexity Revisited.* The results of Experiment I showed a statistically significant correlation between attorneys' ratings of the conceptual difficulty of the original jury instructions and the paraphrase scores (both full performance and approximation) of those subjects with a high school education. If the attorneys were in fact rating only conceptual complexity, and not linguistic complexity or some combination of the two, these correlations should have remained even after the language of the instructions was modified. In other words, there should have been, at the very least, a significant correlation between the attorneys' ratings and subjects' scores on the modified instructions.

There was, in fact, a significant correlation between the attorneys' ratings and the full performance measure for the modified instructions ($r = -0.45$ over all subjects, and -0.40 for those subjects with a high school education or less).⁷⁸ However, as explained above, the full performance measure may be a comparatively poor way to assess a subject's overall grasp of the fundamental concepts of an instruction. The approximation measure is much better suited for that task. When the approximation measure for the modified instructions was compared with the attorneys' ratings, the comparison actually yielded a correlation in the opposite direction.⁷⁹

In other words, the more conceptually complex an instruction was rated, the better the subjects' approximation scores were. This could mean either that the attorney raters were not in fact able to disassociate conceptual complexity from linguistic complexity, or that conceptual complexity plays a secondary role in the comprehensibility of jury instructions. In either event, these results cast doubt on attorneys' assertions that it is the conceptual complexity of a jury instruction that creates comprehension problems and that therefore rewriting instructions will not help.

In order to investigate the problem further, we compared the attorneys'

77. Subjects were paired only if the actual score for one subject differed from the other subject's predicted score by no more than 0.03. Eight subjects—four in each treatment group—could not be paired, because they failed to meet this standard.

78. There were too few subjects with higher education to perform a meaningful analysis.

79. The correlation coefficient for all subjects was 0.19; for subjects with 12 years of education or less, it was 0.37.

ratings of conceptual complexity with the improvement in performance on the modified versions of the instructions. If the attorneys had rated the conceptual complexity of an instruction without reference to its linguistic complexity, and if conceptual complexity was a major cause of comprehension problems, one would expect that those instructions rated as most conceptually difficult would show the least change in paraphrase scores when they were linguistically modified. In fact, for the approximation measure the opposite is the case. The greater the rated conceptual difficulty, the greater the improvement when the instruction was rewritten.⁸⁰ Even though the results of correlating attorneys' ratings with performance and changes in performance are equivocal, it nonetheless appears that even jury instructions containing very complex concepts can be made more understandable when their language is simplified.

C. *The Effect of Specific Linguistic Changes*

Because we had deliberately set out to reduce the number of problematic constructions in the modified instructions, we compared first the number of various types of "problematic" constructions in the original and the modified instructions. Table 10⁸¹ shows that the modified instructions contain far fewer of the types of constructions that appeared to be difficult. Obviously, this table is intended only to show some of the ways in which we attempted to simplify and clarify the modified instructions. As will be noted, we did not mechanically eliminate all passives, or replace all complements or "whiz"es. In certain cases, the passive construction was essential for proper focus, and where we used a truncated passive in a modified instruction, the agent was obvious from the context.

We also examined, where possible, how changing a problematic construction affected paraphrase scores. We did this by looking at all the variables that contained a particular type of problematic construction in the original instructions, and comparing them to their counterpart variables without the problematic construction (where they existed)⁸² in the modified instructions.

Table 17 sets forth, for each item or grammatical construction examined, the number of pairs of variables compared across original and modified instructions, and the mean change in paraphrase scores for those constructions. In addition, the table sets forth the percent improvement by

80. For the approximation measure, there was a statistically significant correlation between an instruction's rated conceptual difficulty and the improvement on the modified instructions ($r = 0.68$, $p < .005$). The correlation between rated conceptual difficulty and improvement on the full performance measure was not statistically significant ($r = -0.1$).

81. Table 10 sets forth the number of occurrences of each type of construction—passives, whiz/complement deletions, nominalizations, misplaced phrases—in the original instructions and in their modified counterparts. Other problematic constructions, such as multiple embeddings and illogical discourse structure, could not conveniently be set forth in a table.

82. In many cases there were no counterpart variables in the modified instructions, because in rewriting we either deleted the variable or made multiple changes that precluded comparison with the original. We will discuss in the third set of analyses why we deleted or radically altered certain variables.

construction, and the level of statistical significance of this improvement. In all cases the deletion of the problematic construction resulted in improvement; in four cases the improvement was statistically significant.

1. *Nominalizations.* Nine instances of nominalizations⁸³ in the original instructions had "de-nominalized" counterparts in the modified instructions. "De-nominalizing" increased paraphrase scores for those items by an average of 45%. Thus, for example, changing the nominalized phrase "any offer of evidence" in BAJI 1.02 to "items . . . were offered as evidence" in the modified version resulted in a raw score improvement of 0.20 for that variable.

2. *"As to" Phrases.* Six phrases beginning with "as to" in the original instructions had rewritten counterparts in the modified instructions. It appears that changing the "as to," and in some cases rewriting its phrase or clause in the modified instructions, increased paraphrase performance by an average of 19% for those items. In only two cases, however, did the modification involve only changing or eliminating the "as to" element. For the other four, the accompanying phrase or clause was also changed grammatically, so the improvement may not be due simply to eliminating "as to." This is understandable, since we believe that the presence of "as to" is often simply a symptom of other grammatical problems and semantic fuzziness. The two pairs for which only the "as to" was changed improved by an average raw score of 0.07.

3. *Misplaced Phrases.* Six sentences in the original instructions containing misplaced phrases had counterparts in the modified instructions; the rewritten versions remedied the misplacements. The mean overall improvement was 0.06, or 24% over the originals. Moreover, while the misplaced phrases in the original instructions had a tendency to produce *incorrect* responses (as opposed to omissions), there were few erroneous paraphrases for the rewritten versions.

4. *Lexical Items.* Of the thirty-six apparently difficult lexical items in the original instructions, seventeen were replaced with simpler synonyms or phrases in the modified instructions. The mean score for these seventeen went from 0.34 correct to 0.50 correct, a 47% increase. For example, while in BAJI 3.71 the subjects were told that his agent's negligence "must be imputed" to the plaintiff, in the modified version they were told that the agent's negligence "would transfer" to the plaintiff. This change accounted for a raw score improvement of 0.46 (0.25 to 0.71). When the words "foreseen or anticipated" in BAJI 3.11 were changed to the phrase "would . . . have realized in advance," the performance increased from a raw score of 0.18 correct to 0.46 correct.⁸⁴

83. See text accompanying note 41 *supra*.

84. One lexical change produced unexpected results: changing "proximate cause" to "legal cause" in BAJI 3.75 did not result in the improvement expected. See text following note 108 *infra*.

5. *Multiple Negatives.* The original instructions contained six instances of multiple negatives, and the results of Experiment I indicated that the subjects were having considerable difficulty processing them. However, in order to eliminate these multiple negatives in our rewritten instructions, it was necessary to make major structural changes in the affected sentences. As a result, no modified variables could be paired with multiple negative constructions. However, the trend in the results from the original instructions (only 0.26 correct for Experiment I; 0.27 correct for Experiment II) and the large body of psycholinguistic research in this area⁸⁵ leave little doubt that multiple negation impedes comprehension and that eliminating multiple negatives will improve comprehension.

6. *Passives.* Modifications of passives, full and truncated, in the original instructions created the greatest number of paired variables.⁸⁶ When the passive construction was deleted, there was an overall improvement of 48.5%.⁸⁷ For example, in BAJI 1.01, the change of the phrase "no emphasis thereon is intended by me" to "that does not mean that I am trying to emphasize those . . ." resulted in a raw score improvement of 0.16, from 0.17 correct to 0.33 correct. Of even greater significance, however, is the fact that seven subjects who heard the original version in Experiment II actually misunderstood the phrase; with the rewritten version, only one subject did. In BAJI 3.11, the phrase "the conduct reasonably could be avoided" was changed to "[could the reasonably careful person] have avoided behaving . . ." with a resulting 0.25 improvement in raw score, from 0.33 correct to 0.58. Overall, in seventeen of the twenty-two pairs examined, subjects performed much better in paraphrasing active-voice phrases than their passive counterparts.

The most interesting finding, however, had to do with the location of the original passive in a main or subordinate clause. As indicated earlier,⁸⁸ it appeared that the passive construction created serious comprehension problems only when located in a subordinate clause. The results of Experiment II confirmed this. When passive constructions in subordinate clauses were changed to the active voice in the modified instructions, paraphrase scores improved from 0.22 correct to 0.39 correct, an improvement that was statistically significant at the .01 level. On the other hand, when main clause passives were changed or eliminated, the improvement was not statistically significant.

In rewriting the instructions, we did not change all passives to the active voice. In some cases, we even introduced a passive construction where we considered that sort of focus necessary for the meaning of the sentence; examples included "you *are not required* to accept those opinions" (Modified

85. See note 49 and accompanying text *supra*.

86. Eighteen original truncated passives and four original full passives were paired with modified instructions that did *not* contain passive constructions.

87. This was statistically significant at the .01 level.

88. See text accompanying notes 53 & 54 *supra*.

BAJI 2.40), and "performing the duties he *was hired to do*" (Modified BAJI 3.71). The average paraphrase score for our thirteen uses of passives in the modified instructions was 0.49 correct, as compared to 0.29 correct for passives in the original instructions.⁸⁹

7. *Whiz/Complement Deletions.* Six of the seven whiz/complement deletions in Table 17 were actually part of passive constructions. Although this does not affect the validity of our conclusions with respect to the passives examined, it does prevent us from making claims with regard to restoring deleted whizes or complements.⁹⁰

As we have said, the original jury instructions contained only four instances of whiz/complement deletion that were independent of a passive construction. As the passive constructions in subordinate clauses appeared to be causing more egregious problems, we believe that the improvement for those variables in which passive and whiz deletion interacted was due primarily to the change from passive to active. This does not mean that reinstating the deleted whiz or complement does not also improve comprehension; we simply cannot make any definitive claims for it. However, for the four original whiz deletions that were not part of a passive construction, the mean score was a very low 0.23 correct.

8. *Word Lists.* The original instructions contained twelve instances of word lists,⁹¹ eleven of which had counterparts in the modified instructions. Since subjects in Experiment I had paraphrased one, or at best two, of the items in any given list, in rewriting the lists we reduced them to one or two items. Thus, in each case, the modified list contained fewer items than the corresponding original list.

One might expect that if a list contains more items, there is a greater likelihood that a subject will remember *one* of those items. In order to determine whether or not this is the case, a subject was given full credit if he or she correctly paraphrased any one item in a list, for both the original and modified instructions. In eight of the eleven cases, subjects actually performed better on the shorter lists in the modified instructions than on the corresponding longer lists in the original instructions. The mean improvement—0.07, a 17% increase—was not statistically significant, but it suggests that it is not necessary to bludgeon jurors with a long list of items or attributes in the hope that they will remember one of them. To the contrary, doing so appears to be counterproductive.

9. *Discourse Structure.* Earlier⁹² we pointed out four instructions whose overall structures appeared to be at least partially responsible for subjects' poor performance. BAJI 1.01 and 1.02 were badly organized;

89. This increase is significant at the .001 level.

90. For the one complement deletion that was not part of a passive construction (in BAJI 4.30)—"resulting therefrom" rewritten as "as a result of taking the risk"—the improvement was 0.09, a 38% improvement.

91. See text accompanying note 56 *supra*.

92. See text accompanying notes 57 & 58 *supra*.

ideas were not grouped logically, and there was no indication of the number of different ideas that each instruction covered. BAJI 3.01 did list and number four conditions, but with confusing redundancy. BAJI 3.71 lacked introductory material putting the listener on notice that the rule might be counterintuitive.

In rewriting BAJI 1.01 and 1.02, we reordered and grouped the ideas logically, and numbered them to provide mental "signposts" to the listener. For both instructions, performance on the modified version was significantly better than on the original versions, for full performance, approximation, and concordance measures.⁹³

Obviously, the improvement in paraphrase scores for BAJI 1.01 and 1.02—particularly BAJI 1.02—is also due to specific syntactic and semantic changes. However, an examination of the way in which subjects attempted to paraphrase the original and modified versions shows that they had far less difficulty organizing their ideas—and their paraphrases—for the modified versions. Even when they could not remember a given idea in the modified instructions, they at least were aware that their paraphrases were incomplete. This was not true for the original instructions. For the modified versions, typical comments were "That's about it, it was just three things" (BAJI 1.01); and "and there was a fifth. I can't remember" (BAJI 1.02).

10. *Embeddings.* It was not possible to make meaningful pairings of embeddings (*i.e.*, subordinate clauses) in the original instructions and the language that replaced them in the modified versions. To achieve a better understanding of the effect on comprehensibility of embeddings as a class, however, we were able to analyze the embeddings used in the modified instructions, and to compare their effect on performance with the effect on performance of embeddings in the original instructions.

Table 18 sets out the mean full performance measures for sentences containing various numbers of embeddings, for both the original and the modified instructions. In contrast to the results in Experiment I, there were no significant changes between the scores for sentences in the modified instructions as the number of embeddings increased.⁹⁴ Our analysis showed that the decrease in comprehension for sentences with greater numbers of embeddings in Experiment I actually related more to the types, rather than the number, of embeddings used.⁹⁵ Moreover, performance on the sentences in the modified instructions was substantially better than performance on original instructions containing the same number of embeddings.⁹⁶ This is

93. See Tables 12, 14 & 15 *infra*.

94. In Experiment II subjects' performance on the original instructions decreased as the number of embeddings increased ($r = -0.86$). However, subjects' performance on the modified instructions was relatively unaffected by the number of embeddings ($r = -0.05$).

95. See text accompanying note 59 *supra*.

96. It might be argued that this difference is due to a difference in average sentence length between sentences with the same number of clauses in the original and modified instructions. However, even when sentences are matched by both number of clauses and number of words, the observed differences persist. The comparison is nevertheless somewhat limited, because the sentences being compared relate to unrelated subject matter and are drawn from unrelated parts of the instructions.

because in rewriting the instructions we made certain to use types of embeddings that linguistic theory and previous research have shown to be relatively easy to process.

An examination of the kinds of sentences containing multiple embeddings that were used in both the original and modified instructions suggests why embeddings created less problems in the modified versions. For example, the following sentences both contain one main clause and three embeddings:

You must never speculate *to be true* any insinuation *suggested by a question asked a witness*. (BAJI 1.02)

In order *to decide whether or not the defendant was negligent*, there is a test *you can use*. (Modified 3.11)

In BAJI 1.02, the subordinate clauses are complicated by whiz deletions; in addition, the last two clauses are in the passive voice, which forces one to be embedded within the other. The modified instruction, on the other hand, is a typical subordinate clause/main clause construction, with the added factor that each has another subordinate clause. The logical structure, however, is still very straightforward.

The following sentences each contain one main clause and four or more embeddings:

[H]owever, *if counsel for the parties have stipulated to any fact, or any fact has been admitted by counsel*, you will regard that fact *as being conclusively proved as to the party or parties making the stipulation or admission*. (BAJI 1.02)

If a lawyer objected to a question, and I did not allow the witness to answer the question, you must not try to guess *what the answer might have been*. (Modified 1.02)

On the surface, these two sentences are somewhat similar: both are complex conditionals. However, in the original instruction the antecedent "if" clauses are not parallel—one is active and one is passive. In the modified instruction, both are active and thus probably require less mental gymnastics. The consequent "then" clauses are quite different. The original uses two impersonal constructions: "as being conclusively proved" and "as to the party . . . making the stipulation." The modified instruction has only one consequent, and it is a personal construction in the normal order of subject-verb-object.

D. Readability

One would have expected that using conventional readability formulas, improved comprehensibility would have been reflected in improved readability scores. This was the case for half the rewritten instructions. However, for two of the rewritten instructions that showed large improvements over the originals in comprehensibility (BAJI 2.40 and 3.01),

the Flesch readability scores were worse than the originals. For one rewritten instruction that showed significant improvement in comprehensibility (BAJI 4.31), the readability score was almost exactly the same as the original. Conversely, for the four rewritten instructions that showed no significant increase in comprehensibility over the originals (BAJI 1.00, 2.21, 3.50, and 3.75) the readability scores were better than the readability scores for the originals. Thus, the scores using a readability formula gave erroneous information regarding the actual comprehensibility of half of the instructions. Moreover, the correlations between readability scores and the performance on the modified instructions did not significantly differ from zero.⁹⁷ When the change in full performance scores between original and modified instructions was compared with the change in readability scores for each, there was a statistically significant negative correlation.⁹⁸ Similar comparisons using the approximation and concordance measure yielded correlations that did not significantly differ from zero. These trends tend to cast doubt on the usefulness of readability formulas for gauging the real improvement in a document.

IV. A COMPARISON OF ORIGINAL AND MODIFIED INSTRUCTIONS

The third set of analyses is an instruction-by-instruction comparison of the original and modified instructions, showing the salient differences between them. We will focus on the reasons for rewriting the instructions as we did.

BAJI 1.00

Ladies and Gentlemen of the Jury:

It is my duty to instruct you in the law that applies to this case and you must follow the law as I state it to you.

As jurors it is your exclusive duty to decide all questions of fact submitted to you and for that purpose to determine the effect and value of the evidence.

You must not be influenced by sympathy, prejudice or passion.

1.00 Modified

Members of the Jury:

I am now going to tell you the laws that apply to this case. As jurors you have two major duties:

First, you must look at the evidence, and decide from the evidence what the facts of this case are. It is your job and no one else's to decide what the facts are.

97. For the full performance measure the correlation coefficient was $r = 0.056$ and for the approximation measure the correlation coefficient was $r = -0.327$.

98. The correlation coefficient was $r = -0.468$; this was statistically significant at the .05 level.

Second, you must listen to the laws that I am now telling you, and follow them, in order to reach your verdict.

In fulfilling these duties, you must not be influenced by your feelings of sympathy or prejudice.

On the basis of the results from Experiment I, we made several changes in this instruction, most of them minor. The subjects in Experiment I, and those in Experiment II who heard the original version, largely ignored the first sentence, which expresses the judge's duty—"It is my duty to instruct you" We therefore changed the sentence to a statement of intention, "I am now going to tell you" This increased the paraphrase score for this phrase in Experiment II from 0 to 0.42.⁹⁹ In the original instruction the judge's duty and the jurors' duty appear together in the same sentence, with the clause outlining the jurors' duty—"and you must follow the law . . ."—trailing behind. This clause lacks focus. Thus, when we rewrote the instruction, we rewrote the clauses as separate sentences, and focused on the jurors' duties separately. This resulted in some improvement: from 0.29 to 0.33 correct.

The phrase "it is your *exclusive* duty" had been misinterpreted by several subjects in Experiments I and II to mean "it is your *only* duty." We therefore changed it to "it is your job and no one else's." This modification raised paraphrase scores from 0.08 to 0.17, and, more important, totally eliminated misunderstanding. Reducing the number of items in the list "sympathy, prejudice or passion" from three to two appeared to have no effect.

For this instruction, improvement from the original to the modified version was not large.

BAJI 1.01

If in these instructions any rule, direction or idea is repeated or stated in varying ways, no emphasis thereon is intended by me and none must be inferred by you. For that reason you are not to single out any certain sentence or any individual point or instruction and ignore the others, but you are to consider all the instructions as a whole and are to regard each in the light of all the others.

The order in which the instructions are given has no significance as to their relative importance.

1.01 Modified

As you listen to these instructions of law, there are three things you must keep in mind:

First, throughout these instructions, you may find that some ideas or

⁹⁹ The results for the original instructions, or parts of them, are taken from the results of Experiment II unless we have indicated to the contrary.

rules of law are repeated. That does not mean that I am trying to emphasize those rules or ideas.

Second, you must consider all the instructions together, as a package, and you must not ignore any instruction or any part of an instruction. And

Third, the fact that I am giving the instructions in a particular order does not mean that the first are more important than the last, or vice-versa. In other words, the order has no significance.

The original instruction appeared to have a number of serious problems.¹⁰⁰ As discussed earlier, the phrase "If in these instructions any rule, direction or idea is repeated" generated a great deal of confusion. Jurors thought they were being told that the *instructions* themselves might be repeated rather than the rules, directions, or ideas *within* the instructions.¹⁰¹ The subjects also had trouble with the segment "no emphasis thereon is intended by me/and none must be inferred by you," which consists of two negative phrases in the passive voice. The subjects did considerably worse, however, in paraphrasing the first phrase;¹⁰² the difference in performance might be explained by the presence in the first phrase of the archaic legal pro-form "thereon," and by the fact that the subjects were generally better able to paraphrase their own duties than those of the judge.

The subjects in Experiment I omitted most of the elements contained in the segment "For that reason/you are not to single out/any certain sentence/or any individual point/or instruction/ and ignore the others" The poor performance is understandable when one considers that comprehension of this segment depends on an understanding of what has gone before; since most subjects thought that it was instructions—and not parts of them—that were being referred to in the first sentence, they naturally omitted reference to anything that was not a full instruction. The principal thrust of this instruction is that the jurors are to "consider all the instructions as a whole." Twenty-six percent of the subjects, however, misinterpreted "as a whole" to mean "as of equal weight or importance." "As a whole" is a cliché, and it appears from our data that many people simply do not understand what it means. The following clause, "and are to regard each in the light of all the others," is also hackneyed and probably no longer understood; it provides little elucidation for the cliché that precedes it.

From the point of view of eighteenth-century English and Biblical writing, this jury instruction is esthetically pleasing: each idea has its parallel; the rhythm is enchanting. But as a means of communicating the law to the layman, it is clearly inadequate. The entire instruction is poorly organized and lacks any introductory material to create a transition. In modifying the instruction, we wrote an introductory sentence, to function as an advance

100. Overall performance was only 23% correct.

101. See note 46 and accompanying text *supra*.

102. Seventy-seven percent omitted the first phrase, and 3% paraphrased it incorrectly; 46% omitted the second phrase, and 9% paraphrased it incorrectly.

organizer, and divided up and grouped the mass of thoughts in the original into three important points, which we numbered.

Our rewriting of the first clause, which contained the misplaced phrase "in these instructions," increased the correct paraphrase score from 0.24 to 0.38. More importantly, the misunderstandings caused by the misplacement (0.28 incorrect) were eliminated. The archaic and ambiguous phrases were replaced: instead of "no emphasis *thereon* is intended by me," we wrote "That does not mean that I am trying to emphasize those rules or ideas"; "together" replaced the phrase "as a whole." The overall modifications resulted in a significant improvement in paraphrase scores for each of the three measures: full performance (41% increase), approximation (33% increase), and concordance (52% increase).

BALI 1.02

You must not consider as evidence any statement of counsel made during the trial; however, if counsel for the parties have stipulated to any fact, or any fact has been admitted by counsel, you will regard that fact as being conclusively proved as to the party or parties making the stipulation or admission.

As to any question to which an objection was sustained, you must not speculate as to what the answer might have been or as to the reason for the objection.

You must not consider for any purpose any offer of evidence that was rejected, or any evidence that was stricken out by the court; such matter is to be treated as though you had never known of it.

You must never speculate to be true any insinuation suggested by a question asked a witness. A question is not evidence and may be considered only as it supplies meaning to the answer.

1.02 Modified

As I mentioned earlier, it is your job to decide from the evidence what the facts are. Here are five rules that will help you decide what is, and what is not, evidence.

1. *Lawyers' Statements. Ordinarily, any statement made by the lawyers in this case is not evidence. However, if all the lawyers agree that some particular thing is true, you must accept it as the truth.*

2. *Rejected Evidence. At times during this trial, items or testimony were offered as evidence, but I did not allow them to become evidence. Since they never became evidence, you must not consider them.*

3. *Stricken Evidence. At times, I ordered some piece of evidence to be stricken, or thrown out. Since that is no longer evidence, you must ignore it, also.*

4. *Questions to a Witness. By itself, a question is not evidence. A question can only be used to give meaning to a witness's answer. Further-*

more, if a lawyer's question to a witness contained any insinuations, you must ignore those insinuations. And

5. Objections to Questions. *If a lawyer objected to a question, and I did not allow the witness to answer the question, you must not try to guess what the answer might have been. You must also not try to guess the reason why the lawyer objected to the question.*

Like BAJI 1.01, BAJI 1.02 is very poorly organized. In addition, it contains numerous instances of inappropriate passives (e.g., "A question . . . may be considered"), whiz and complement deletions (e.g., "statement of counsel * made during the trial"; "insinuation * suggested by a question"), nominalizations (e.g., "stipulation"), misplaced phrases (e.g., "speculate to be true any insinuation"), "as to" phrases and odd embeddings (e.g., "you will regard that fact as being conclusively proved as to the party or parties making the stipulation or admission").

The original instruction contained nine passives, most of which were in subordinate clauses. The mean correct paraphrase score for these was 0.18. Of the nine passives, seven had analogues in the modified instruction. For example, "evidence that was rejected" became "but I did not allow them to become evidence"; "and may be considered only" became "a question can only be used"; "question to which an objection was sustained" became "a lawyer objected . . . and I did not allow the witness to answer the question."¹⁰³ For these seven, the increase in scores averaged 0.16.

By using passives only where appropriate, eliminating nominalizations, changing "as to" phrases, correcting misplaced phrases, rethinking sentence structures to eliminate bizarre multiple embeddings, and giving the instruction a more logical organization, correct scores were significantly increased. For the full performance measure, there was a 49% improvement; for the approximation measure, 52%; and for the concordance, 64%.

BAJI 2.21

Discrepancies in a witness's testimony or between his testimony and that of others, if there were any, do not necessarily mean that the witness should be discredited. Failure of recollection is a common experience, and innocent misrecollection is not uncommon. It is a fact, also, that two persons witnessing an incident or a transaction often will see or hear it differently. Whether a discrepancy pertains to a fact of importance or only to a trivial detail should be considered in weighing its significance.

2.21 Modified

As jurors, you have to decide which testimony to believe and which testimony not to believe.

¹⁰³ Not all of the rewriting changed the passive to the active form. See text preceding note 89 *supra*.

You may be tempted to totally disbelieve a witness because he contradicted himself while testifying. Keep in mind, however, that people sometimes forget things, and end up contradicting themselves.

You might also be tempted to totally disbelieve a witness because another witness testified differently. But keep in mind also that when two people witness an incident they often remember it differently.

When you are deciding whether or not to believe a witness, you should consider whether contradictions or differences in testimony have to do with an important fact or only a small detail.

This instruction, in its original form, contains several impersonal constructions, including truncated passives (e.g., "the witness should be discredited"; "should be considered") and nominalizations (e.g., "failure of recollection"). The last sentence has an entire subordinate clause as the subject of a passive verb. In addition, the instruction is intentionally vague; it does not tell jurors exactly what to do with discrepancies in a witness's testimony. Rather, it attempts to convey even-handedly the idea that witnesses are not perfect and can make innocent mistakes, and that therefore jurors must use their own judgment.

In rewriting this instruction, we found it necessary to preserve this vagueness, as it is central to the meaning of the instruction. We did, however, personalize the instruction by eliminating truncated passives, nominalizations, and odd embeddings. We also provided an introductory sentence for context and split up the key ideas that were originally packed into one sentence. These alterations improved the scores for several individual items.

However, in light of the results, the linguistic changes were insufficient. Although paraphrase scores improved for the modified instruction, the improvements were not statistically significant. It is possible that the inherent vagueness of the message of this instruction overwhelmed the effect of the linguistic changes.

BAII 2.40

A witness who has special knowledge, skill, experience, training or education in a particular science, profession or occupation may give his opinion as an expert as to any matter in which he is skilled. In determining the weight to be given such opinion you should consider the qualifications and credibility of the expert and the reasons given for his opinion. You are not bound by such opinion. Give it the weight, if any, to which you deem it entitled.

2.40 Modified

Ordinarily, a witness may not give an opinion. However, an expert witness may give his opinion. An expert witness is a person with special skill or education in a particular field.

Even though I allow expert witnesses to give their opinions, you are not required to accept those opinions. In order to determine the value of the expert's opinion, you should take into account the expert's qualifications and believability. You should also take into account how he reached his conclusions.

The structure of the first sentence of the original instruction appeared to cause some serious comprehension problems. Because an enormously long relative clause, listing attributes, immediately follows the subject "a witness," and because the specific type of witness to which the instruction refers, "an expert," is lost in a participial phrase after the verb, nine of the twenty-four subjects misinterpreted the meaning of the sentence. Those nine subjects thought either that an expert witness could not give an opinion or that an ordinary witness could.

The second sentence contains the puzzling construction "the reasons given for his opinion."¹⁰⁴ Only 17% of the subjects who heard the original version could decipher and paraphrase this ambiguous item. The phrase "[i]n determining the weight to be given such opinion" was very poorly understood (12% correct): it is a truncated passive infinitive, with no real indication of who is to give the weight to the opinion. This, in turn, appeared to affect performance on the third sentence, "You are not bound by such opinion." Subjects may have been confused as to whose opinion was referred to (the witness's, their own opinion of the witness's opinion, or someone else's entirely), as 12% of them incorrectly thought that they *were* bound by some opinion or other.

In rewriting BAJI 2.40, we made explicit the implied rule that ordinarily a witness may not give an opinion. We highlighted the fact that an *expert* witness may give an opinion by focusing on it syntactically. Rather than burying a prepositional phrase toward the end of the sentence, as in the original version, we made "expert witness" the subject of a simple sentence. These modifications resulted in only one misunderstanding of who may give an opinion—as opposed to nine in the original version.

We cut down the original list of five attributes to two and the original list of three synonyms for profession to one—"field." This also enhanced the subjects' ability to paraphrase. We made it clear whose opinion was being referred to and clarified the jurors' latitude in using the expert's opinion. This was done by means of a complex sentence—"Even though I allow expert witnesses to give their opinions, you are not required to accept those opinions"—setting out the ideas in a logical fashion. This reduced misunderstanding from 12% to 4%, and increased correct paraphrasing by 0.20. We also changed the ambiguous phrase "the reasons given for his opinion" to "how he reached his conclusions"; this produced a 0.50 increase in correct paraphrase scores.

104. This construction is in the passive voice and contains a whiz deletion.

Overall, the modified version produced dramatically better scores than the original. The increase was 63% for the full performance measure, 30% for the approximation measure, and 58% for the concordance measure.

BAJI 3.01

A plaintiff who was not contributorily negligent and who did not assume the risk of harm and who was injured as a proximate result of some negligent conduct on the part of a defendant is entitled to recover compensation for such injury from that defendant.

Thus, the plaintiff is entitled to a verdict in this case if you find, in accordance with my instructions:

- 1. That defendant was negligent;*
- 2. That such negligence was a proximate cause of injury to the plaintiff;*
- 3. That the plaintiff was not negligent; or, if negligent, that his negligence did not contribute as a proximate cause to his injury; and*
- 4. That the plaintiff did not assume the risk of harm.*

3.01 Modified

The plaintiff is entitled to recover money if you find that all four of the following conditions are true:

- 1. The defendant was negligent;*
- 2. The defendant's negligence was a legal cause of the plaintiff's injury;*
- 3. The plaintiff was not contributorily negligent; and*
- 4. The plaintiff did not voluntarily take the risk of being injured.*

As indicated earlier,¹⁰⁵ the major problem with this instruction was the redundant first paragraph. It appeared to confuse some subjects by leading them to believe that the second paragraph said something different. In addition to eliminating the first paragraph, we changed the words "entitled to a verdict" (0.50 correct) and "entitled to recover compensation" (0.46 correct) to "entitled to recover money" (0.92 correct). The original version contained a conditional within a conditional: "if you find . . . 3. That the plaintiff was not negligent; or, if negligent, that his negligence did not contribute as a proximate cause . . ." We changed this to a clause with more legalistic vocabulary "if . . . the plaintiff was not contributorily negligent," but whose logical and linguistic structure is much simpler.

Because we removed an entire redundant paragraph, the full performance measure for the modified instruction was far higher than for the original instruction—a 100% increase—though, as we have pointed out, this cannot be considered a valid measure of improvement. The approximation measure, however, which counts any one correct paraphrase of a redundant

105. See text preceding note 58 *supra*.

item, does provide a valid, in fact a conservative, gauge of improvement. For this instruction, the modified version produced a 28% improvement in the approximation measure. Similarly, the concordance measure, which also eliminates the effects of redundancy, showed a 34% improvement.

BALI 3.11

One test that is helpful in determining whether or not a person was negligent is to ask and answer whether or not, if a person of ordinary prudence had been in the same situation and possessed of the same knowledge, he would have foreseen or anticipated that someone might have been injured by or as a result of his action or inaction. If such a result from certain conduct would be foreseeable by a person of ordinary prudence with like knowledge and in like situation, and if the conduct reasonably could be avoided, then not to avoid it would be negligence.

3.11 Modified

In order to decide whether or not the defendant was negligent, there is a test you can use. Consider how a reasonably careful person would have acted in the same situation. Specifically, in order to find the defendant negligent, you would have to answer "yes" to the following two questions:

1. Would a reasonably careful person have realized in advance that someone might be injured as a result of the defendant's conduct?

And,

2. Could the reasonably careful person have avoided behaving as the defendant did?

If your answer to both of these questions is "yes," then the defendant is negligent. You can use the same test in deciding whether the plaintiff was negligent.

The original instruction consists of two very long, complex sentences. The first contains a conditional within a conditional within an embedding ("One test . . . is to ask and answer whether or not, if a person . . . had been . . . , he would have . . ."). It also contains the term of art "person of ordinary prudence"; various doublets such as "foreseen or anticipated," "by or as a result of," and "action or inaction"; and the archaism "possessed of the same knowledge." The overall mean paraphrase score for the first sentence was 0.27 correct. Subjects had particular difficulty with the doublets and some of the embeddings. Eleven subjects misinterpreted the term "person of ordinary prudence" to mean the average person or the average juror.

The second sentence is a conditional with two antecedent clauses, both of which have passive or passive-like ("would be foreseeable") verb forms. The consequent (main clause) contains an infinitive phrase with a double negative ("not to avoid it"). In addition, the antecedent contains the ambiguous phrase "from certain conduct," the problematic "person of ordinary

prudence," and an archaic use of "like." The overall mean correct score for this sentence was 0.24.

In rewriting this instruction, we broke up both sentences into shorter ones, in order to clarify the logic of the instruction. We changed the focus of the first sentence from "[o]ne test that is helpful . . .," to "in order to decide . . . there is a test you can use." The original phrase, by its placement, presupposed that jurors know about legal tests for determining negligence. We changed "person of ordinary prudence" to "reasonably careful person," shifting the focus from "ordinary" to "careful," which is essential to the meaning. This latter change reduced the number of subjects who misinterpreted the phrase from eleven to one.

We changed the nebulous "[o]ne test . . . is to ask and answer . . .," to "in order to find . . . you would have to answer 'Yes' to the following two questions . . ." This alerted the listener to the fact that there were *two* conditions that had to be met. We then presented the two conditions in the active voice, and reiterated the fact that both conditions had to be met. We also eliminated the archaic terminology and the doublets.

Overall improvement for the modified instruction was statistically significant and dramatic for all three measures. Improvement by the full performance measure was 83%; by the approximation measure, 43%; and by the concordance, 79%.

BAII 3.50

Contributory negligence is negligence on the part of a plaintiff which, combining with the negligence of a defendant, contributes as a proximate cause in bringing about the injury.

A plaintiff who is contributorily negligent cannot recover for such injury.

3.50 Modified

There is a type of negligence that involves the conduct of the plaintiff, rather than the defendant. It is called "contributory negligence."

If a plaintiff is negligent, and his negligence helps cause his own injury, we say that the plaintiff is contributorily negligent.

A plaintiff who is contributorily negligent cannot recover money for his injury.

The original version of this instruction was well understood (see Tables 1 and 2), and, indeed, there is not much wrong with it linguistically. The principal change that we made was to add an introductory sentence to highlight the difference between negligence and contributory negligence. Inasmuch as this information is not essential to the legal concept, we anticipated that few subjects would bother to paraphrase it, and that therefore the full performance measure for the modified version might be less than for the original. As expected, there was a substantial, although not statistically significant, decrease in the full performance measure for the modified version—22%. However, for the approximation measure, which disregards non-

essential concepts, there was a 14% increase for the modified version. This increase was not statistically significant, which did not surprise us, as the original scores had been very good.¹⁰⁶

One small change that apparently made a large difference was rewriting "recover for such injury" as "recover money for his injury." The score for the original version was 0.50 correct; for the modified it was 0.96, a 0.46 increase.

BALI 3.71

If you should find that John Smith, who, at the time of the accident in question, was driving the vehicle in which plaintiff was riding, was negligent and that his negligence contributed as a proximate cause of plaintiff's injury, then you must determine whether said driver was then the agent of the plaintiff and acting within the scope of his employment.

If the driver was plaintiff's agent and acting within the scope of his employment, his negligence, if any, must be imputed to the plaintiff, with the same effect as if the plaintiff himself were contributorily negligent.

But if said driver was not then the agent of plaintiff or was not acting within the scope of his employment, his negligence, if any, may not be imputed to the plaintiff.

3.71 Modified

As you recall, John Smith was driving the truck at the time of the accident, and the plaintiff was a passenger in that truck. Ordinarily, in deciding whether the plaintiff was contributorily negligent, you would only look at the plaintiff's conduct. However, there is one situation where John Smith's conduct affects the plaintiff's ability to recover money. That situation is where, at the time of the accident, John Smith was the plaintiff's agent, and was performing duties he was hired by the plaintiff to do.

If you find that at the time of the accident, John Smith was the plaintiff's agent, and was performing duties that he was hired to do, then any negligence on John Smith's part would transfer to the plaintiff. It would be as though the plaintiff himself were negligent.

On the other hand, if you find that John Smith was not the plaintiff's agent, or that he was not performing duties that he was hired by the plaintiff to do, then any negligence on John Smith's part would not transfer to the plaintiff.

The first sentence of this instruction is a long conditional sentence with two antecedents and a conclusion that contains another conditional ("then you must determine whether . . ."). Both the logical structure and the syntactic structure appeared to cause enormous problems. Moreover, the terms of art "agent" and "acting within the scope of his employment" were poorly

106. The concordance measure showed a very slight increase for the modified version—2%.

understood. The second sentence is also a complex conditional, but here the conditional format was not the major problem. Rather, subjects appeared to be confused by two additional "if's that were not conditionals (*i.e.*, "if any"; "as if the plaintiff himself were contributorily negligent"). In addition, the terms "agent" and "acting within the scope of his employment" appeared again, as well as another unfamiliar lexical item—"impute"—which appeared as a truncated passive.¹⁰⁷

In rewriting, we began with the information that had been buried in two embedded relative clauses in the original instruction: "John Smith was driving . . . and the plaintiff was a passenger . . ." In this way, the contextual information was highlighted. We then added two sentences providing explicit information that could only have been inferred—at best—from the original version: "However there is one situation . . . that situation is where . . ." These two sentences alert the listener to the fact that the rule of law that follows states something unusual, or not in accord with common sense.

We changed the improper focus of the conditional in the original first sentence from John Smith's negligence to John Smith's status as an agent. In addition, we changed the phrase "acting within the scope of his employment" to "performing duties that he was hired to do," which increased correct paraphrases from a mean of 0.22 to a mean of 0.44 for the three instances of each. The passive phrases "must be imputed" and "may not be imputed" were changed to "would [not] transfer," which increased paraphrase scores from 0.25 to 0.71 for the positive form, and from 0.29 to 0.58 for the negative form. We continued to use the conditional formats—both in the positive and in the negative—but we directed them toward the jurors rather than toward the parties in the case by eliminating truncated passives.

The addition of the introductory material to the modified version artificially lowered its full performance measure, so that it showed a 9% decrease from the original version. However, the approximation measure showed a 93% improvement. The concordance measure showed a 36% improvement.

BAJI 3.75

A proximate cause of an injury is a cause which, in natural and continuous sequence, produces the injury, and without which the injury would not have occurred.

3.75 Modified

A legal cause of an injury is something that triggers a natural chain of events that ultimately produces the injury.

Without the legal cause, the injury would not occur.

107. The third sentence is similar to the second but is in the negative—"If . . . not . . . [then] . . . not . . ."

In Experiment I, the full performance measure for this instruction was almost 40%, but it produced proportionally the most misunderstanding—*i.e.*, incorrect responses per variable—of the instructions tested. Only ten of the thirty-five subjects were able to paraphrase some or all of the instruction without any errors. This is interesting to note, particularly because this is the shortest instruction tested, at twenty-seven words. One would think that it would be possible, if nothing else, to memorize such a short instruction, especially after two playings. However, not only did no one memorize it, four subjects were unable to paraphrase any of it.¹⁰⁸

First of all, the term "proximate cause" was misunderstood by 23% of the subjects in Experiment I.¹⁰⁹ They interpreted it as "approximate cause," "estimated cause," or some fabrication. As noted above, the phrase "in natural and continuous sequence" produced a whole variety of misunderstandings. These misunderstandings not only affected the comprehension of that phrase, but also caused comprehension problems on subsequent variables.

We rewrote this misplaced phrase as "triggers a natural chain of events that ultimately produces the injury." This change eliminated misunderstandings (*i.e.*, there were no *erroneous* paraphrases for this variable), and increased the paraphrase score for this variable from 0.25 correct in the original to 0.46 in the modified version.

We also rewrote "proximate cause" as "legal cause." This change, however, created some unexpected results. Although the subjects' performance improved, from 0.46 to 0.71 correct, 25% of the subjects who heard "legal cause" misinterpreted it as the opposite of an "illegal cause." We would therefore recommend that the term "legal cause" not be used in jury instructions; instead, the simple term "cause" should be used, with the explanation that the law defines "cause" in its own particular way.

For the full performance measure there was a 22% improvement for the modified instruction, which was not statistically significant. For both the approximation and concordance measures, the improvement was 7%.

BAJI 4.30

If plaintiff assumed the risk of harm he may not recover damages for an injury resulting therefrom.

In order for plaintiff to have assumed such risk, he must have had actual knowledge of the particular danger and an appreciation of the risk involved and the magnitude thereof, and must thereafter have voluntarily assumed such risk.

For a person to act voluntarily he must have freedom of choice. This freedom of choice must come from circumstances that provide him a rea-

108. This is further confirmation of the psycholinguistic finding that one cannot remember what one has not understood. Only one subject—who had a Ph.D.—paraphrased the entire instruction correctly, by means of an example of cause and effect.

109. In Experiment II, 38% of the subjects omitted any reference to the term, while 17% paraphrased it erroneously.

sonable opportunity, without violating any legal or moral duty, to safely refuse to expose himself to the danger in question.

In determining whether the plaintiff assumed such risk, you may consider his maturity, intelligence, experience and capacity, along with all the other surrounding circumstances as shown by the evidence.

4.30 Modified

If the plaintiff voluntarily took a risk, and was injured as a result of taking the risk, then he cannot recover money for his injury.

In order for someone to voluntarily take a risk, he must know and understand the dangers involved, and decide to take the chance anyway.

In deciding whether the plaintiff voluntarily took a risk, you may consider his maturity and intelligence, and all the other circumstances of the case.

This instruction had several problems. The sentence structure, in general, is not particularly complex, but unfamiliar terms of art such as "assumed the risk of harm," the unusual use of the common word "appreciation" (which is also a nominalization), and the legalistic pro-forms "therefrom," "thereof," and "thereafter," whose antecedents are unclear, appeared to cause difficulty for the subjects. In addition, the lengthy, convoluted definition for the relatively simple word "voluntarily" confused virtually all the subjects.

In rewriting, we translated "assumed the risk of harm" into the common expression "took a risk." We changed "damages" to "money" and the nominalized "appreciation" to the verb form "understand," and we eliminated the archaic pro-forms. Since almost none of the subjects who heard the original instruction paraphrased the convoluted explanation of "voluntarily," and since the explanation did not really explain anything, we eliminated it. We also cut the list of the four attributes of the plaintiff to two—"maturity and intelligence."

These modest changes produced significant improvement for all three measures. Full performance improved 37%; the approximation measure increased 57%; and the concordance measure increased 39%.

BAJI 4.31

While the same conduct on the part of the plaintiff may amount to both assumption of risk and contributory negligence, there is a distinction between the two defenses. An essential factor in contributory negligence is that it contribute as a proximate cause of the injury. Assumption of risk, however, if it meets the requirements stated to you, will bar recovery of damages although it plays no part in causing the accident except merely to expose the person to danger.

4.31 Modified

As I have previously said, a plaintiff cannot recover money either if

he was contributorily negligent, or if he voluntarily took the risk of being injured.

Even though both contributory negligence and voluntarily taking a risk keep the plaintiff from winning, there is a difference between the two. If the plaintiff was contributorily negligent, he actually helped cause his own injury, through his own negligence. In contrast, if the plaintiff voluntarily took a risk, he merely exposed himself to danger. He may not have been negligent, and may not have caused his own injury.

This instruction was very poorly understood; the full performance score was 0.21. The first sentence is unfocused and poorly organized. The grammatical form of the opening clause—"While the same conduct on the part of the plaintiff . . ."—presupposes that the juror knows that there are circumstances under which the same conduct may "amount to both assumption of risk and contributory negligence." A juror, even in the context of a trial, may have no such knowledge, and hence the significance of the entire instruction may be lost. This is best illustrated by the fact that only 0.38 of the subjects recognized that the instruction was contrasting two legal doctrines.

In addition, the term "assumption of risk"¹¹⁰ had not appeared before. The prior instruction referred to the plaintiff having "assumed the risk," which was itself poorly understood. The instruction further presupposes that the listener is aware of the fact that contributory negligence and assumption of the risk are legal "defenses." The term "defense," with its legal meaning, had never been used before.

The focus of the second sentence is also strange: the subject of the sentence is "an essential factor in contributory negligence," rather than contributory negligence itself, which had been referred to before. This phrase is then contrasted with the subject of the third sentence, "assumption of risk," which further underlines the oddness of the subject of sentence two. In short, sentences two and three, which should have parallel construction, do not, and the listener is in the position of comparing oranges with apple cores.

Sentence three has a complicated structure. There is an embedded conditional, "if it meets the requirements . . .," and, at the end, a series of embeddings in an "although" clause which also contains the double negative "no . . . except." Furthermore, the word "however" makes it appear that assumption of the risk does not have the same legal effect as contributory negligence, and indeed, in Experiment I, 20% of the subjects incorrectly interpreted the instruction to mean that contributory negligence barred recovery of damages while assumption of risk did not, or vice versa.

In modifying this instruction, we added an initial sentence to clarify the notion that assumption of risk and contributory negligence have the

¹¹⁰. This term is both a technical legal phrase and a nominalization. The nominalized form adds complexity to an already difficult phrase.

same effect: both bar recovery. The second modified sentence shifts the focus from the plaintiff's conduct to the existence of a distinction between the two doctrines. It is this distinction that is important in this instruction, and the rewritten version produced a 0.16 improvement in subjects' paraphrases of this idea (from 0.38 to 0.54). We rewrote the two nonparallel explanations of assumption of risk and contributory negligence using parallel sentence structures ("If . . . then . . ."; "In contrast, if . . . then . . ."). Finally, we changed the structure of the last part of the instruction to simple negative sentences, eliminating the complex embeddings and the double negative.

There was a dramatic improvement for the modified version of this instruction, across all three measures. The full performance measure showed a 93% increase; the approximation measure, a 78% increase; and the concordance, a 100% improvement.

BAJI 15.22

The court has given you instructions embodying various rules of law to help guide you to a just and lawful verdict. Whether some of these instructions will apply will depend upon what you find to be the facts. The fact that I have instructed you on various subjects in this case including that of damages must not be taken as indicating an opinion of the court as to what you should find to be the facts or as to which party is entitled to your verdict.

15.22 Modified

I have given you various rules of law to help you reach a just verdict. Some of these rules of law may or may not apply to this case, depending upon what you find to be the facts.

Simply because I have instructed you on subjects such as damages does not mean that I am giving my opinion. I am not giving any opinion regarding what the facts are, or regarding who should win.

BAJI 15.22 is not a difficult instruction conceptually, but it contains a number of linguistic problems. First of all, it uses the impersonal term "the court" instead of "I." The participial phrase "embodying various rules of law" is redundant; "just and lawful" is also redundant. Furthermore, the phrase "[t]he court has given you instructions embodying . . ." was misunderstood: two subjects thought it meant that the judge was telling them the facts of the case.

The subject of sentence two is a subordinate clause ("[w]hether some of these instructions will apply") as is the object ("what you find to be the facts"). This makes the sentence somewhat difficult to understand (0.22 correct).

The last sentence contains a truncated passive main verb and numerous embeddings and prepositional phrases, including two embeddings beginning with the famous "as to." Once again, the judge is referred to as "the court."

The phrase "must not be taken as indicating an opinion of the court" carries an inappropriate thrust and lacks an agent, and the subject of "indicating" is unclear. One subject interpreted this to mean that the judge's instructions were to be disregarded.

In the modified instruction we used "I" in place of "the court," and removed the redundancies in the original first sentence. This increased the score for the first sentence from 0.22 to 0.54; there were no misinterpretations. We gave the second sentence a more normal structure, changing the subordinate clause subject into a main clause ("Some of these rules of law may or may not apply to this case") and subordinating the second part of the sentence to it ("depending upon . . ."). The final clause ("what you find . . .") remained unchanged. Changing the structure of the sentence made it more logical and improved performance from 0.22 to 0.39.

In the final sentences, we used the active voice and put more emphasis on the idea that the judge is not giving his opinion, by using the words "simply because" to begin the initial subordinate clause rather than the very neutral "the fact that." In order to give this additional emphasis, we added the clause "I am not giving any opinion . . ." Even though we changed the awkward and formal "as to what you should find to be the facts or as to which party is entitled to your verdict" to the simpler construction "regarding what the facts are or regarding who should win," the subjects' performance was equally poor for both versions. This is understandable, as the important concept is that the judge is not giving an opinion, and the last two clauses only modify "opinion."

Overall, for the full performance measure, there was a 44% improvement for the modified instruction; this is statistically significant. The improvement for the approximation measure (13%) was not statistically significant, while the improvement for the concordance was marginally significant (26%).

BAJI 15.30

When you go to the jury room it is your duty to discuss the case for the purpose of reaching an agreement if you can do so.

Each of you must decide the case for yourself, but should do so only after a consideration of the case with the other jurors.

You should not hesitate to change an opinion if you are convinced it is erroneous. However, you should not be influenced to decide any question in a particular way simply because a majority of the jurors, or any of them, favor such a decision.

15.30 Modified

When you go to the jury room, it is your duty to discuss the case and reach a verdict.

You should discuss the case with the other jurors, but each of you must decide the case for yourself.

Do not be afraid to change your opinion if you think you are wrong. However, you should not come to a particular decision simply because other jurors think it is the right decision.

The original version is a very easy instruction, and the full performance measure (0.508) reflects this. Nonetheless, we felt that some of its grammatical structures were needlessly complex and that it could be improved.

The first sentence contains an overly wordy embedding, "for the purpose of reaching an agreement." In the second sentence, the order of the two main clauses does not parallel their temporal sequence (*i.e.*, "Each of you must decide . . . but . . . only after . . ."). In addition, the second main clause contains the nominalization "a consideration of," which makes this part of the sentence abstract and impersonal. We felt that the words "hesitate," "convinced" (a passive-like form), and "erroneous" in the third sentence were too complex for the average juror. The fourth sentence uses the passive "you should not be influenced to decide . . ." when it is not really necessary for focus; it is redundant and only adds to the syntactic complexity of the sentence. The verb "favor" is not in ordinary usage and the phrase "a majority of the jurors, or any of them" is stilted.

We changed the awkward purpose clause in the first sentence to "[to] reach a verdict" in the modified instruction. In the second sentence, we reordered the two main clauses to reflect the appropriate temporal (and logical) sequence and changed the nominalization to a verb form, using the verb "discuss" rather than "consider." In the third sentence, we made several lexical changes, changing "[y]ou should not hesitate" to "do not be afraid," "are convinced" to "think," and "erroneous" to "wrong." In the final sentence, we eliminated the redundant passive verb phrase "you should not be influenced" and changed "a majority of the jurors, or any of them" to simply "other jurors."

These simple modifications significantly improved the subjects' performance for all three measures. The full performance measure increased 39%; the approximation measure, 20%; and the concordance measure, 34%.

CONCLUSION

This study set out to test three hypotheses: (1) that the standard jury instructions used in this study—when viewed as discourse—are not well understood by jurors; (2) that certain linguistic constructions are largely responsible for the incomprehensibility; and (3) that if the problematic linguistic constructions are appropriately altered, comprehension will dramatically improve, notwithstanding the "legal complexity" of any given instruction. The results of the two experiments that we have described in this Article support these hypotheses. In addition, we have demonstrated that a paraphrase methodology and linguistic analysis are powerful tools not only for testing relative comprehensibility, but also for discovering which

aspects of legal discourse are potentially difficult for non-lawyers to understand.

The study provides evidence that there is more to legal language than merely "jargon"—an esoteric vocabulary. Certain grammatical constructions and discourse structures found in the jury instructions appear to be recurring elements in legal language.¹¹¹ Although these constructions are found in ordinary usage, they appear with much greater frequency in legalese and tend to characterize it as a distinct sublanguage. The results of the study also indicate that these constructions—rather than the legal complexity of the jury instructions—were responsible for comprehension problems.

The results of this study—in conjunction with the results of other studies of jury instruction comprehension—underscore the fact that jury instructions are not written for their major intended audience. The inability of jurors to comprehend the charge adequately has obvious implications concerning the soundness of the jury system: if many jurors do not properly understand the laws that they are required to use in reaching their verdicts, it is possible that many verdicts are reached either without regard to the law or by using improper law.

Lack of comprehension of jury instructions has implications beyond the trial level: the vitality of the appellate process depends, to a large extent, on the assumption that the jury instructions have been understood and properly applied.¹¹² If the propriety of a jury instruction is questioned on appeal, the appellate court must first decide whether the challenged instruction is legally correct, and, if it is not, decide whether the incorrect charge amounted to prejudicial error. The appellate court makes no attempt to ascertain the comprehensibility of the controversial instruction. In some instances, however, an instruction that appears correct to an appellate judge may have in fact been so incomprehensible to the average juror that its use, in effect, resulted in prejudicial error. In other cases, it may be that an important, yet substantively erroneous, jury instruction was not understood by the jurors, and that the correct instruction, had it been given, would also have defied comprehension. Thus it is possible that appellate decisionmaking with respect to jury instruction is little more than an academic exercise. If appellate courts were forced to consider whether jury instructions were properly understood, the drafters of jury instructions would probably take more care to draft instructions with a lay audience in mind.

The results of this study, and particularly the linguistic findings, can

111. See Charrow, Crandall, & Charrow, *Characteristics and Functions of Legal Language*, in *SUBLANGUAGE: STUDIES OF LANGUAGE IN RESTRICTED SEMANTIC DOMAINS* (R. Kittredge & J. Lehrberger eds., to be published 1980). Examples of such constructions include subordinate clause passives (often with *whiz* deletion); nominalizations; negatives; misplaced phrases; vague prepositional phrases (often beginning with "as to"); unusual subordinate clause embeddings; unusual discourse conventions (such as restating entire passages in different words); and strings of synonyms. These characteristics were isolated because they were consistently not well understood and poorly paraphrased by our subjects.

112. As Roger Traynor stated, "In the absence of definitive studies to the contrary, we must assume that juries for the most part understand and faithfully follow instructions." R. TRAYNOR, *THE RIDDLE OF HARMLESS ERROR* 73-74 (1970).

be used by appellate courts as criteria for judging the comprehensibility of jury instructions, and drafters of jury instructions can use the linguistic findings as guidelines for drafting more comprehensible instructions. Our study also has significance beyond jury instructions. All of the grammatical constructions, discourse conventions, and vocabulary items that we isolated in the jury instructions commonly appear in other branches of legal language. It has become clear that many lay persons who are affected by legal language cannot effectively understand it. Studies such as the present one will help to define those aspects of legal language that should be changed to make them accessible to the non-lawyer.

TABLE 1
 FULL PERFORMANCE MEASURE ACROSS SUBJECTS
 (N = 35 Subjects)
 Mean Fraction Correct Per Instruction
 (Explicit Correct + Inferred Correct) ÷ Number
 of Variables

Instruction (BAJI)	Mean Fraction Correct Per Instruction (Explicit Correct + Inferred Correct) ÷ Number of Variables	Standard Deviation σ
1.00	0.395	0.185
1.01	0.230	0.127
1.02	0.257	0.134
2.21	0.511	0.249
2.40	0.396	0.152
3.01	0.497	0.243
3.11	0.337	0.183
3.50	0.619	0.297
3.17	0.325	0.182
3.75	0.404	0.291
4.30	0.270	0.165
4.31	0.324	0.274
15.22	0.394	0.139
15.30	0.549	0.174
Mean \bar{b}	0.386	0.118
Mean of Means \bar{c}	0.393	0.149

TABLE 2
 APPROXIMATION MEASURE ACROSS SUBJECTS
 (N = 35 Subjects)

Instruction (BAJI)	Mean Fraction Correct Per Instruction	Standard Deviation σ
1.00	0.554	0.227
1.01	0.420	0.200
1.02	0.428	0.207
2.21	0.682	0.302
2.40	0.718	0.264
3.01	0.564	0.222
3.11	0.475	0.243
3.50	0.810	0.316
3.17	0.438	0.214
3.75	0.390	0.326
4.30	0.502	0.295
4.31	0.435	0.336
15.22	0.500	0.206
15.30	0.643	0.193
Mean \bar{b}	0.540	0.169
Mean of Means \bar{c}	0.540	0.128

TABLE 3
 PEARSON CORRELATION COEFFICIENTS—FULL PERFORMANCE MEASURE
 ACROSS SUBJECTS BY APPROXIMATION MEASURE ACROSS SUBJECTS

Instruction (BAJI)	Correlation Coefficient d (r)	Level of Significance e
1.00	0.89	$p < .001$
1.01	0.94	"
1.02	0.94	"
2.21	0.88	"
2.40	0.74	"
3.01	0.84	"
3.11	0.75	"
3.50	0.88	"
3.71	0.82	"
3.75	1.00	"
4.30	0.76	"
4.31	0.97	"
15.22	0.76	"
15.30	0.95	"
Across All Instructions	0.98	"
Mean Coefficient for Instructions	0.87	

TABLE 4
 ORDER OF PRESENTATION V. APPROXIMATION MEASURE ACROSS SUBJECTS AND INSTRUCTIONS

Order No.	Mean Fraction Correct	Std. Dev. (σ) a	No. Subjects
1	0.555	0.179	9
2	0.542	0.168	9
3	0.519	0.073	9
4	0.543	0.251	8
$F = 0.0643$ f			

TABLE 5

A. ATTORNEYS' RATINGS OF CONCEPTUAL COMPLEXITY V. FULL PERFORMANCE
AND APPROXIMATION MEASURES ACROSS SUBJECTS

Instruction (BAJI)	Attorney Rating (1 = conceptually easy; 11 = conceptually difficult)	Full Performance Measure	Approximation Measure
1.00	1.72	0.395	0.554
1.01	2.32	0.230	0.420
1.02	3.97	0.257	0.428
2.21	3.25	0.511	0.682
2.40	2.84	0.396	0.718
3.01	*	0.497	0.564
3.11	5.06	0.337	0.475
3.50	3.47	0.619	0.810
3.71	8.00	0.325	0.438
3.75	5.62	0.404	0.390
4.30	4.97	0.270	0.502
4.31	7.69	0.324	0.435
15.22	2.16	0.394	0.500
15.30	1.53	0.549	0.643

Rating v. Full Performance $r = -0.34$
 $r^2 = 0.12$

Rating v. Approximation $r = -0.46$
 $r^2 = 0.21$

* 3.01 was not rated

B. CORRELATION BETWEEN RATINGS OF CONCEPTUAL
COMPLEXITY V. PERFORMANCE BROKEN DOWN BY
EDUCATION

	Correlation Coefficient Full Performance Measure	Correlation Coefficient Approximation Measure
High School Education ≤ 12 yrs. (N = 8)	-0.623	-0.588
College > 12 yrs. ≤ 16 yrs. (N = 16)	-0.208	-0.391
Graduate > 16 yrs. (N = 11)	-0.349	-0.388

TABLE 6
 MEAN SENTENCE LENGTH PER INSTRUCTION V. FULL PERFORMANCE MEASURE PER INSTRUCTION

Instruction (BAJI)	Mean Sentence Length (Words)	Full Performance Measure (Mean Fraction Correct Per Instruction)
1.00	18.0	0.395
1.01	30.3	0.230
1.02	30.8	0.257
2.21	20.75	0.511
2.40	19.75	0.395
3.01	58.5	0.497
3.11	50.5	0.337
3.50	19.5	0.619
3.71	43.0	0.325
3.75	27.0	0.404
4.30	25.6	0.270
4.31	26.33	0.324
15.22	28.67	0.394
15.30	23.75	0.549
Mean of Means	30.175	0.386
	$r = -0.162$	
	$r^2 = 0.026$	

TABLE 7
FULL PERFORMANCE MEASURE BY INSTRUCTION BY EDUCATION

Instruction (BAJI)	High School ≤ 12 years (N = 8) (Full Performance)	College > 12 years ≤ 16 years (N = 16) (Full Performance)	Graduate School > 16 years (N = 11) (Full Performance)	F(2, 32) ^a	Statistical Significance
1.00	0.250	0.368	0.540	8.73	p < .01
1.01	0.136	0.241	0.281	3.63	p < .05
1.02	0.188	0.273	0.285	1.47	N.S.
2.21	0.257	0.576	0.601	7.54	p < .01
2.40	0.234	0.411	0.492	10.78	p < .01
3.01	0.267	0.557	0.579	6.09	p < .05
3.11	0.151	0.367	0.428	8.15	p < .01
3.50	0.361	0.694	0.697	4.79	p < .05
3.71	0.161	0.349	0.409	5.80	p < .05
3.75	0.156	0.461	0.500	4.62	p < .05
4.30	0.112	0.304	0.333	6.27	p < .05
4.31	0.133	0.392	0.364	2.80	N.S.
15.22	0.275	0.433	0.424	4.62	p < .01
15.30	0.400	0.558	0.642	5.81	p < .01

TABLE 8
APPROXIMATION MEASURE BY INSTRUCTION BY EDUCATION

Instruction (BAJI)	High School ≤ 12 years (N = 8)	College > 12 years ≤ 16 years (N = 16)	Graduate School > 16 years (N = 11)	F(2, 32) ^a	Statistical Significance
1.00	0.391	0.531	0.705	5.85	p < .01
1.01	0.275	0.463	0.464	3.07	N.S.
1.02	0.359	0.447	0.450	0.56	N.S.
2.21	0.411	0.768	0.753	5.21	p < .05
2.40	0.423	0.829	0.771	10.31	p < .01
3.01	0.411	0.556	0.688	4.35	p < .05
3.11	0.193	0.589	0.515	11.95	p < .01
3.50	0.583	0.917	0.818	3.39	p < .05
3.71	0.271	0.424	0.576	5.88	p < .01
3.75	0.104	0.469	0.485	4.94	p < .05
4.30	0.250	0.536	0.636	5.20	p < .05
4.31	0.208	0.493	0.515	2.60	N.S.
15.22	0.391	0.555	0.500	1.78	N.S.
15.30	0.475	0.638	0.773	7.70	p < .01

TABLE 9

A. PERFORMANCE ACROSS INSTRUCTIONS BY LINGUISTIC CONSTRUCTION

Linguistic Construction	Frequency of Occurrence	Mean Fraction Correct/Construction	Mean Fraction Correct for All Variables Except Those In Col. 1	Mean Col. 4 Less Col. 3	Min. F' s (i, j)
Nominalizations	14	0.286	0.399	0.113	(j = 308) 3.87*
"As to" Phrases	8	0.282	0.396	0.114	N.S.
Misplaced Phrases	9	0.240	0.398	0.158	(j = 305) 7.04**
Whiz/ Complement Deletion	12	0.245	0.400	0.155	(j = 294) 4.52*
Problematic Lexical Items	36	0.321	0.404	0.083	(j = 301) 4.86*
Modals	16	0.570	0.382	-0.188	(j = 299) 9.39**
Single Negatives (non-modals)	13	0.369	0.394	0.025	N.S.
Multiple Negatives	6	0.262	0.396	0.134	N.S.
Passives—All	35	0.392	0.393	0.001	N.S.
Full Passives	9	0.268	0.397	0.129	N.S.
Truncated Passives	26	0.434	0.389	-0.045	N.S.
Subordinate Clause Passives	19	0.271	0.402	0.131	(j = 303) 7.45**
Main Clause Passives	16	0.535	0.384	-0.151	(j = 299) 4.50*
Word Lists	12	0.431			

* $p < .05$ ** $p < .01$

N.S. = Not Significant

B. PASSIVE CONSTRUCTIONS—TYPE BY LOCATION (MEAN FRACTION CORRECT)

Location	Type		Total
	Truncated	Full	
Subordinate Clause	(N = 13) 0.316	(N = 6) 0.173	(N = 19) 0.271
Main Clause	(N = 13) 0.553	(N = 3) 0.457	(N = 16) 0.535
Total	(N = 26) 0.434	(N = 9) 0.268	

Main Clause v. Subordinate Clause Passive for:

Truncated Passives, Min. F' (1, 28) = 6.05, $p < .05$

Full Passives, N.S.

Truncated and

Full Passives, Min. F' (1, 38) = 10.14, $p < .01$

Truncated v. Full Passives for:

Subordinate Clause Passives, N.S.

Main Clause Passives, N.S.

Subordinate and Main

Clause Passives, N.S.

TABLE 10
FREQUENCY OF PROBLEMATIC CONSTRUCTIONS FOR ORIGINAL AND MODIFIED BY INSTRUCTION

BAJI		Problematic Constructions (Number of Occurrences)							
Instruction	Version	A	B	C	D	E	F	G	H
1.00	Original	0	0	1	2	0	1	1	1
	Modified	0	0	0	0	0	1	0	0
1.01	Original	0	1	1	1	0	2	3	1
	Modified	0	0	0	0	0	1	1	1
1.02	Original	3	4	3	3	0	3	6	4
	Modified	0	0	0	1	0	0	1	0
2.21	Original	2	0	0	1	2	0	2	2
	Modified	0	0	0	0	0	0	2	0
2.40	Original	0	1	0	2	0	0	3	1
	Modified	0	0	0	0	0	0	1	0
3.01	Original	0	0	0	5	0	0	3	0
	Modified	0	0	0	0	0	0	1	0
3.11	Original	2	0	0	5	1	2	1	0
	Modified	0	0	0	0	0	1	0	0
3.50	Original	0	0	0	2	0	0	0	0
	Modified	0	0	0	0	0	0	0	0
3.71	Original	0	0	0	6	0	0	2	0
	Modified	0	0	0	1	0	2	1	1
3.75	Original	0	0	1	3	1	0	0	0
	Modified	0	0	0	1	1	0	0	0
4.30	Original	3	0	1	2	1	1	0	1
	Modified	0	0	0	0	0	0	1	0
4.31	Original	3	0	0	3	1	0	1	1
	Modified	0	0	0	0	0	0	0	0
15.22	Original	0	2	2	0	0	0	2	0
	Modified	0	0	0	0	0	0	0	0
15.30	Original	1	0	0	1	0	0	2	1
	Modified	0	0	0	0	0	0	0	2
TOTALS	Original	14	8	9	36	6	9	26	12
	Modified	0	0	0	3	1	5	8	4

KEY: A=Nominalization; B="As to" phrases; C=Misplaced Phrases;
D=Vocabulary; E=Multiple Negatives; F=Full Passives;
G=Truncated Passives; H=Whiz/Complement Deletions

TABLE 11
A. JURY INSTRUCTIONS RECEIVED—BY TREATMENT GROUP

BAJI Instruction	TREATMENT GROUP 1			TREATMENT GROUP 2		
	Form Received— (Orig. or Modified)	Number Variables	Number Words	Form Received— (Orig. or Modified)	Number Variables	Number Words
1.00	Orig.	17	72	Mod.	19	96
1.01	Orig.	22	91	Mod.	24	107
1.02	Mod.	51	222	Orig.	30	154
2.21	Orig.	18	83	Mod.	27	109
2.40	Orig.	24	79	Mod.	18	80
3.01	Mod.	10	56	Orig.	22	98
3.11	Orig.	24	101	Mod.	24	114
3.50	Mod.	14	55	Orig.	9	38
3.71	Mod.	31	178	Orig.	24	129
3.75	Mod.	7	29	Orig.	8	27
4.30	Mod.	20	73	Orig.	30	133
4.31	Mod.	21	94	Orig.	15	79
15.22	Orig.	15	86	Mod.	16	94
15.30	Orig.	15	95	Mod.	13	71
TOTALS		289	1,314		279	1,329

B. ORIGINAL V. MODIFIED—BY NO. WORDS & NO. VARIABLES

Form of Instructions	Total No. Words	Total No. Variables	Words per Variable
Modified	1,378	295	4.67
Original	1,265	273	4.63

TABLE 12
FULL PERFORMANCE MEASURE ACROSS SUBJECTS
MODIFIED V. ORIGINAL INSTRUCTIONS

Instruction (BAJI)	Mean Fraction Correct Per Instruction		Δ Mean (Mod. Less Orig.)	Level of Significance (one-tailed z test) ⁱ	Percentage Change
	Original	Modified			
1.00	.348	.421	.07	$p < .06^*$	21%
1.01	.246	.347	.10	$p < .05$	41%
1.02	.201	.300	.10	$p < .05$	49%
2.21	.412	.446	.03	N.S.	8%
2.40	.293	.477	.18	$p < .001$	63%
3.01	.271	.542	.27	$p < .001$	100%
3.11	.255	.467	.21	$p < .005$	83%
3.50	.565	.441	-.12	N.S.**	-22%
3.71	.283	.258	-.02	N.S.**	-9%
3.75	.370	.452	.08	N.S.	22%
4.30	.254	.347	.09	$p < .05$	37%
4.31	.211	.407	.20	$p < .001$	93%
15.22	.253	.365	.11	$p < .01$	44%
15.30	.508	.708	.22	$p < .001$	39%
Mean of Means	.319	.427	.11	$p < .001^{***}$	41% (avg.)

* Marginally significant

** Computed using a two-tailed z test ⁱ

*** Computed using a one-tailed t (13df) test

TABLE 13
EDUCATION AS PREDICTOR OF PERFORMANCE
(FULL PERFORMANCE MEASURE)

TREATMENT GROUP 1 (N = 24) (13 Subjects High School, 7 College, 4 Post College)			
Instruction (BAJI) (Original Version)	Predicted Mean Fraction Correct (using Exp. I Coefficients)	Actual Mean Fraction Correct (Exp. II Results)	Difference (Predicted -- Actual)
1.00	.333	.348	-.015
1.01	.191	.246	-.055
2.21	.407	.412	-.005
2.40	.329	.293	.036
3.11	.260	.255	.005
15.22	.346	.253	.093
15.30	.486	.508	-.022
Mean of Means (Group 1)	.336	.331	.005*
TREATMENT GROUP 2 (N = 24) (12 Subjects High School, 6 College, 6 Post College)			
1.02	.234	.201	.033
3.01	.418	.255	.163
3.50	.528	.525	-.037
3.71	.270	.283	-.013
3.75	.318	.370	-.052
4.30	.215	.254	-.039
4.31	.256	.211	.045
Mean of Means (Group 2)	.320	.306	.014**
Mean of Means (Group 1 + 2)	.328	.318	.010***

* t(6df) = .279 Differences not statistically significant
 ** t(6df) = .492 " " " "
 *** t(13df) = .555 " " " "

TABLE 14
APPROXIMATION MEASURE ACROSS SUBJECTS
MODIFIED V. ORIGINAL INSTRUCTIONS

Instruction (BAJI)	Mean Fraction Correct Per Instruction		Mean Mod. Less Orig.	Level of Significance (one-tailed z test)	Percentage Change
	Original	Modified			
1.00	.474	.521	.05	N.S.	11%
1.01	.417	.560	.14	$p < .05$	33%
1.02	.313	.471	.16	$p < .05$	52%
2.21	.589	.651	.06	N.S.	10%
2.40	.464	.600	.14	$p < .05$	30%
3.01	.457	.589	.13	$p < .05$	28%
3.11	.396	.568	.17	$p < .05$	43%
3.50	.653	.736	.09	N.S.	14%
3.71	.430	.827	.40	$p < .001$	93%
3.75	.410	.438	.03	N.S.	7%
4.30	.488	.774	.28	$p < .001$	57%
4.31	.268	.489	.21	$p < .005$	78%
15.22	.312	.351	.04	N.S.	13%
15.30	.592	.711	.12	$p < .05$	20%
Means of Means c	.447	.592	.14	$p < .001^*$	35% (avg.)

* Computed using a one-tailed t(13df) test

TABLE 15
CONCORDANCE MEASURE ACROSS SUBJECTS
MODIFIED V. ORIGINAL INSTRUCTIONS

Instruction (BAJI)	Number Common Variables	Mean Fraction Correct Per Instruction		Mean Mod. Less Orig.	Level of Significance (one-tailed z test)	Percentage Change
		Original	Modified			
1.00	9	.458	.477	.02	N.S.	4%
1.01	12	.271	.413	.14	$p < .05$	52%
1.02	24	.224	.356	.14	$p < .01$	64%
2.21	13	.458	.481	.02	N.S.	4%
2.40	13	.314	.491	.18	$p < .005$	58%
3.01	8	.411	.552	.14	$p < .05$	34%
3.11	15	.267	.478	.21	$p < .005$	79%
3.50	7	.619	.626	.01	N.S.	2%
3.71	18	.248	.341	.09	$p < .05$	36%
3.75	6	.410	.438	.03	N.S.	7%
4.30	17	.360	.502	.14	$p < .05$	39%
4.31	10	.233	.458	.23	$p < .005$	100%
15.22	14	.268	.333	.07	$p < .08^*$	26%
15.30	13	.532	.715	.18	$p < .001$	34%
Mean of Means c		.362	.476	.11	$p < .001^{**}$	38% (avg.)

* Marginally significant

** Computed using one-tailed t(13df) test

TABLE 16
 CONCORDANCE MEASURE—SUBJECTS PAIRED BY ABILITY TO PERFORM
 PARAPHRASE TASK (20 SUBJECT PAIRS)

Subject Pair		Version			Version		
Group 1 v. Subj. #	Group 2 Subj. #	Orig.	Mod.	Δ No. (Mod. - Orig.)	Orig.	Mod.	Δ No. (Mod. - Orig.)
502	606	29	53	24	28	57	29
504	610	28	50	22	23	63	40
602	601	42	44	2	38	57	19
605	401	20	39	19	24	30	6
609	509	42	59	17	48	54	6
613	513	36	51	15	42	43	1
404	612	56	62	6	51	74	23
617	507	21	21	0	21	32	11
405	515	60	52	-8	51	57	6
410	505	21	23	2	16	31	15
603	415	20	24	4	22	27	5
607	406	20	37	17	29	22	-7
615	616	15	31	16	25	35	10
414	503	35	35	0	39	48	9
416	501	19	35	16	36	46	10
506	614	51	56	5	61	65	4
508	413	35	40	5	39	44	5
510	618	37	59	22	35	51	16
574	604	9	22	13	17	27	10
516	407	4	12	8	12	16	4
Mean		30.00	40.25	10.25*	32.85	43.95	11.1**

* $t(19df) = 5.11, p < .001$

** $t(19df) = 4.73, p < .001$

TABLE 17
PERFORMANCE ACROSS INSTRUCTIONS BY CHANGES IN LINGUISTIC CONSTRUCTION

Construction	No. Variables Compared (n)	Mean Correct With Construction	Mean Correct Without Construction	Difference Between Means	Percentage Change	Significance
Nominalizations	9	0.313	0.453	.140	45	$p < .05$
"As to" Phrases	6	0.210	0.250	.040	19	N.S.
Misplaced Phrases	6	0.278	0.345	.067	24	N.S.
Whiz/Complement Deletion	7	0.266	0.421	.155	58	*
Problematic Lexical Items	17	0.340	0.498	.158	47	$p < .05$
Passives	22	0.301	0.447	.146	48.5	$p < .01$
Subord. Clause Passives	12	0.216	0.391	.175	81	$p < .01$
Main Clause Passives	10	0.404	0.514	.110	27	N.S.
Word Lists	11	0.428	0.462	.046	11	N.S.

* Interaction between Whiz/Complement Deletion and Passive Voice prevents meaningful statistical comparison.

TABLE 18
FULL PERFORMANCE BY NUMBER OF EMBEDDINGS IN SENTENCES WITH ONE MAIN CLAUSE FOR ORIGINAL AND MODIFIED INSTRUCTIONS

No. Embedded Clauses Per Sentence	Mean Full Performance Score—Original Instructions	Mean Full Performance Score—Modified Instructions
1	0.329	0.405 *
2	0.351	0.403 *
3	0.297	0.460 **
4	0.220	0.382 **

* Difference in Mean Scores Between Original & Modified Not Statistically Significant

** Difference in Mean Scores Between Original & Modified Statistically Significant, $p < .05$ (two-tailed z test).

FOOTNOTES TO TABLES

a. The standard deviation (σ) is a measure of variability. For example, suppose that there are two sets of scores with identical means (averages): Set A: 44, 45, 46; Set B: 25, 45, 65. The mean score for both sets is 45. However, the scores in Set A are much closer together than the scores in Set B. Hence, the variability, and thus the standard deviation associated with the scores, is much smaller for Set A (0.816) than for Set B (16.330). The data presented in Table 1 consist of 14 sets of scores (one for each instruction), and each set consists of 35 scores (one for each juror). The mean scores for BAJI 1.00, 2.40, and 15.22 are almost identical. However, the standard deviations associated with the three jury instructions are not identical. The jurors' scores for BAJI 15.22 ($\sigma=1.39$) are closer together than the scores for BAJI 2.40 ($\sigma=1.52$); and the scores for BAJI 2.40 are closer together than the scores for BAJI 1.00 ($\sigma=1.85$). The lower the standard deviation, the more reliable the mean will be for predicting the average performance of another group of subjects.

b. The "mean" for all subjects was computed by calculating that fraction of the total number of variables (273) that each subject correctly paraphrased, then adding the 35 fractions (one for each subject) and dividing the sum by 35.

c. The "mean of means" is the average of the 14 means, computed by adding the 14 mean scores together and dividing by 14.

d. The concept of correlation is widely used in the social sciences. We hear claims that certain traits or events go together: gasoline prices correlate with oil company profits; LSAT scores correlate with first-year performance in law school; intelligence correlates with performance in school.

A correlation may be positive, negative, or zero. For instance, height and weight are positively correlated; tall people tend to be heavier, short people tend to be lighter. If height and weight were perfectly correlated, that is to say if for all persons each additional 10 lbs. were associated with an additional inch in height, then the correlation would be perfect ($r = 1.0$). Of course, in reality height and weight are not perfectly correlated, because some short people weigh more than some tall people. Nevertheless, because more often than not an increment in height will correspond to an increment in weight, the comparison should yield a fractional positive correlation (e.g., $r \approx .5$). The more consistently the two change together, the greater that fraction will be.

A more intuitive notion of the meaning of a particular coefficient can be obtained if the coefficient is squared. For example, consider the correlation between the attorneys' ratings of the conceptual complexity of the original instructions with subjects' actual scores. (See note 30 and accompanying text *supra*.) The correlation coefficient was -0.34 . By squaring that coefficient (i.e., $r^2 = (-0.34)^2 = .12$), we arrive at a number that indicates the fraction of the variation in one variable that can be accounted for by the variation in the other variable. In the example that we have been using ($r^2 = .12$), the differences in rated conceptual difficulty account for only 12% of the variation in subjects' performance. That is to say, 88% of the variation in the performance scores for the instructions is associated with factors other than the variation of the rated conceptual difficulty.

A "negative" correlation does *not* necessarily mean that there is actually the opposite of the correlation expected. It means simply that when the first variable goes higher, the second variable goes lower. Of course, with height and weight a negative correlation would be the opposite of that expected. For a comparison of weight of cars and gas mileage, however, one would *expect* a negative correlation; as the weight gets higher, gas mileage gets lower. Thus a correlation coefficient of $r = -.9$ would mean that there was a very "high" correlation between the weight of cars and poor gas mileage.

A correlation is zero when the two traits are absolutely unrelated one way or the other. Thus, persons high in the first trait are, over a sample, indiscriminately high, medium, and low in the other trait.

The most common means of measuring the correlation between two traits is by using the "Pearson" correlation coefficient (r). The coefficient is calculated by means of a formula too involved for full explanation here. (For information on calculation of the coefficient, see A. EDWARDS, *STATISTICAL METHODS* 101-03 (2d ed. 1967).) It will suffice for a rough understanding of the figures to know that the coefficient ranges from $+1$ to -1 , and that if the Pearson coefficient is equal to 1, then the two traits are perfectly correlated (either in the direction expected or in the opposite direction); if $r = 0$, then there is no correlation.

e. Throughout this Article reference will be made to the notion of "statistical significance." "Statistical significance" is a term of art that implies that the likelihood that two measurements differ by chance is less than some acceptable level. In most social science research the minimum acceptable level is .05, represented as $p < .05$. This figure means that the probability that the statistical event being measured (e.g., an improvement in scores or a correlation between two items) occurred by chance is less than 1 in 20 (5/100). Thus in Table 3, $p < .001$ means that the probability that the positive correlation between the full performance measure

and the approximation measure occurred by chance is less than 1 in 1000. There is more than one formula for calculating statistical significance. See, e.g., note 43 *supra*, explaining the "quasi-Fischer" ratio. For an explanation of the various basic methods of computing statistical significance, see A. EDWARDS, *supra* note d, at 175-367.

f. The F test or Fischer ratio is used to determine whether the variances associated with two or more samples differ significantly: the larger the F ratio the larger the level of significance. As far as the data in Table 4 are concerned, the very low F ratio of 0.0643 tells us that the differences in the mean fraction correct for each of the orders is not statistically significant. That is to say, the variability among the four mean fraction correct scores is very small when contrasted with the variability associated within each order.

g. F' is the symbol for the "quasi-Fischer" ratio, a method of calculating statistical significance. See note 43 *supra*. For a thorough discussion of F' in psycholinguistic research, see Clark, *The Language-as-Fixed-Effect Fallacy: A Critique of Language Statistics in Psychological Research*, 12 J. VERB. LEARN. & VERB. BEHAV. 335 (1973). See generally B. WINER, STATISTICAL PRINCIPLES IN EXPERIMENTAL DESIGN 375 (2d ed. 1971).

h. The numbers within the parentheses indicate the number of "degrees of freedom" associated with the Fisher distribution. For an explanation of the concept of degrees of freedom, see L. HOROWITZ, ELEMENTS OF STATISTICS FOR PSYCHOLOGY AND EDUCATION 171-74 (1974); A. EDWARDS, *supra* note d, at 226-43.

i. The z-test is a statistical test for determining whether or not the difference between the means of two groups of scores is statistically significant. Depending on the hypotheses sought to be tested, a z-test can be either one-tailed or two-tailed. For an explanation of the method of constructing such tests, see L. HOROWITZ, *supra* note h, at 278-86.