Evaluating the effects of document design principles

Janice C Redish, Daniel B Felker and Andrew M Rose
American Institutes for Research

The authors present a model of the process of designing documents. The model can be used as a job aid for writers and as a framework for testing document design principles. The final step in this model is evaluation – testing how easily readers can understand and use the document. They then describe an audience-centered evaluation in which a traditional bureaucratic document was compared with a shortened, reorganized, and rewritten version. Subjects using the revised document answered more questions correctly, were significantly better in identifying the correct section, took less time to answer questions, and rated the revised document as much easier to use.

Dr Janice C Redish is Director of the Document Design Center at the American Institutes for Research (AIR) in Washington, DC. Dr Daniel Felker and Dr Andrew Rose are Senior Research Scientists at AIR. Dr Redish is a linguist, Dr Felker is an instructional psychologist, Dr Rose is an experimental psychologist. They, and other members of the interdisciplinary team at the Document Design Center, are studying how to make public documents more understandable and more useful to readers.

Authors' address:
Janice C Redish
American Institutes for Research
1055 Thomas Jefferson Street, NW
Washington, DC 20007

This paper is based on work that was supported by the National Institute of Education. It does not necessarily reflect the views of the funding agency.

Information design journal
Vol 2, no 3&4, 236-243

Evaluation is a critical step both in the practical process of designing a document and in a research program to develop reliable and valid guidelines for document designers. In the Document Design Center, at the American Institutes for Research, we have developed a process model of document design which has been used very effectively in redesigning individual documents, in developing training materials, and in developing curricula.

The model, which is shown in Figure 1, describes a process which begins with an analysis of the rhetorical context (understanding the purpose, the audience, the audience's tasks and the designer's constraints) and with identifying problems both by expert analysis of the document and by audience-centered testing. The model continues beyond the design stage to an evaluation phase, stressing that the designer's task is not completed until the principles used in the design phase have been validated by testing users' performance with an audience and task that replicate the rhetorical context.

The process model is primarily a job aid for producing documents. As explained elsewhere (Redish, 1981), however, the process model can itself be a framework for testing document design principles. If the writer can appropriately implement document design principles, and if the writer specifies the purpose, audience, task, and the principles that were used to design the document, a well-planned evaluation serves to validate (or challenge) that set of principles in the particular context.

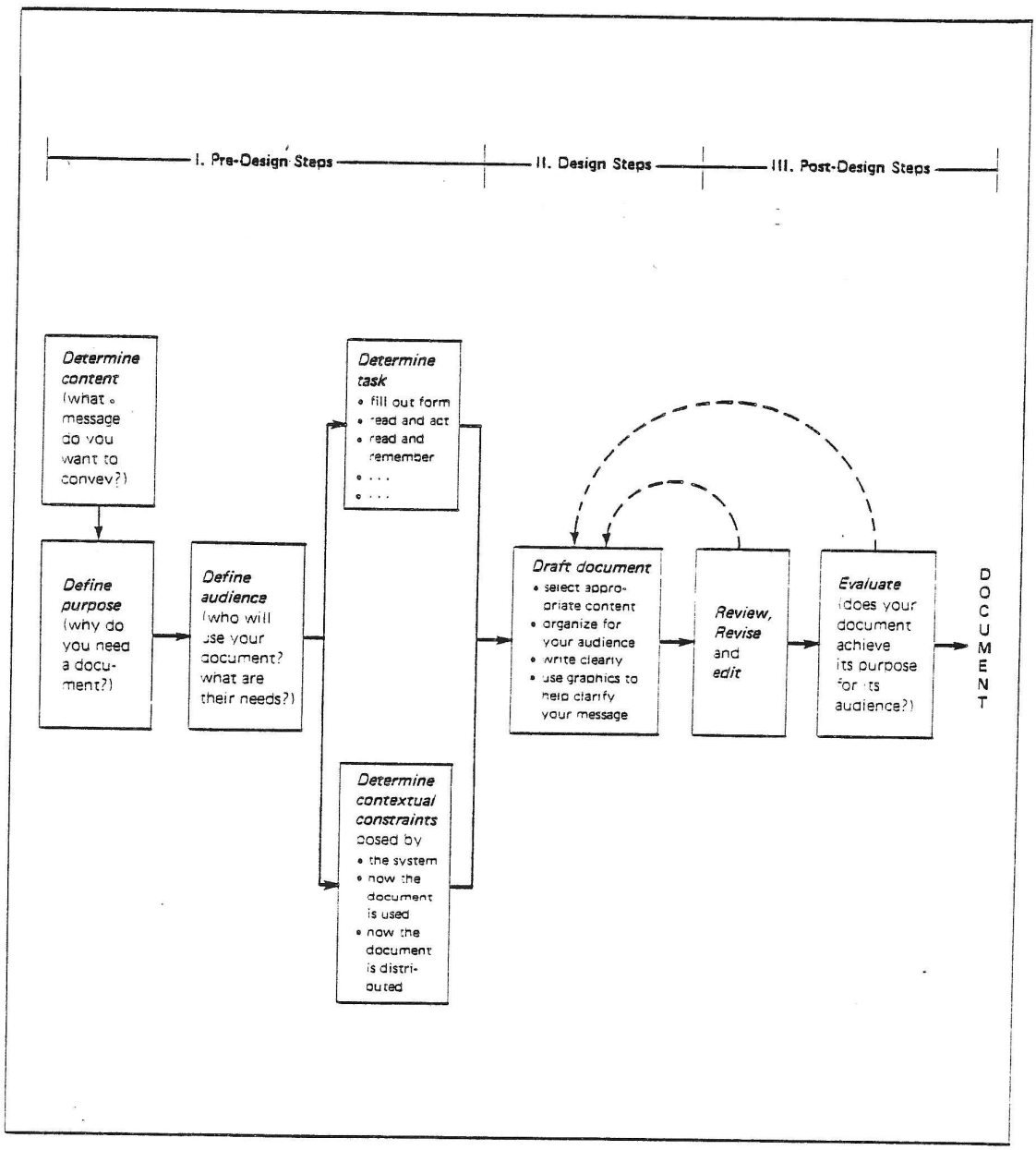
One of the major thrusts of the Document Design Center has been to evaluate the effects of document design principles on users' performance. Our insistence on audience-centered evaluation runs counter to the current practice in evaluating documents. The common method of evaluation in document design is to use a readability formula. However, a growing body of research results supports the intuition of linguists and psychologists that a readability formula is not a sufficient measure of how comprehensible or useful a document is. Furthermore, readability formulas cannot serve to validate a set of guidelines.

The case against readability formulas for evaluation

A spate of recent articles explains why readability formulas are insufficient as a measure of comprehension – particularly when applied to document revisions (Kern, 1979; Klare, 1979; Redish, 1980; Holland, 1981; see also the six short essays in Joenk, 1981). Researchers comparing

0142-5471/81/030236-8$02.00

Figure 1. The process model of document design



readability scores with other measures of comprehension have found that the formulas are *not* reliable predictors of behavioral data. Moreover, there is very low reliability in the predictions of different readability formulas.

Charrow and Charrow (1979) in a study of jurors' ability to comprehend jury instructions found that readability scores were not an accurate evaluation tool. That study tested jurors' understanding of (1) instructions as they are currently given, and (2) instructions rewritten following linguistically-based document design principles. They found that rewritten instructions which produced higher levels of comprehension also had improved readability scores only half the time.

Two rewritten instructions which showed large gains in comprehension had poorer readability scores than the corresponding original instructions. In four cases where the readability score was greatly improved, there was no corresponding significant increase in comprehension. In the Charrow and Charrow study, 'the correlations between readability scores and the performance on the modified instructions did not significantly differ from zero. 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.' (p 1341)

Kintsch and Vipond (1977) also found that readability (as measured by the Flesch Reading Ease scale) did not accurately predict behavioral data (reading time, recall, or question answering). Their stimulus material was four short prose passages.

Other studies have also shown that improving readability scores does not necessarily improve readers' comprehension of the material (Klare, 1976; Kern, 1979). Furthermore, several researchers, attempting to find readability scores for adult reading materials, in the last few years, have applied many formulas to the same passages. They have found that the same passage can have scores ranging over more than twelve grade levels, depending on the formula.

Morris, Thilman, and Myers (1980) scored four versions of an information sheet on valium with thirteen formulas. The variance in the scores ranged from a minimum of 7.4 grades to a maximum of 12.6 grades. Hartley, Trueman, and Burnhill (1980) applied ten formulas to two versions of technical prose. They found that the formulas were inconsistent from sample to sample within the same passage, and did not even agree on predicting whether the revised passage would be easier to read.

The case for audience-centered evaluation

A research approach in which a sample of potential users reads and answers questions about the document is a far more appropriate evaluation technique than a readability formula for assessing whether document design principles enhance comprehension. However, a major question remains. Should we test the effects of single guidelines (where attribution of differences in performance to specific guidelines would be relatively clear) or many guidelines at once (where attribution of the relative effect to any one guideline would be difficult if not impossible, but where we are more nearly approximating a situation that is realistic for document designers)?

Our own experience, and that of other researchers, shows that changing a single feature (or a few features) of a document will not yield significant differences in performance although subjects express clear preferences for the simpler document. In 1979, we conducted a study of product warranties in which we contrasted clearly written warranties that were solid blocks of prose with identically written warranties divided by informative headings (Charrow & Redish, 1980). None of the measures of difference in speed or accuracy was significant. On attitudinal and subjective questions (which warranty was easiest to read and understand? would you be more likely to keep and read warranty A or warranty B?) subjects selected a warranty with headings 90% of the time. We hypothesize that the warranties were too short and the difference of the

Figure 2. Example of 'old' marine radio rule

83.115 Retention of radio station logs.

(a) All station logs which are required under those provisions of this part pertaining to the particular classes of stations subject to this part shall be retained by the licensee for a period of one year from date of entry and for such additional periods as required by the following subparagraphs:

(1) Station logs involving communications incident to a distress or disaster shall be retained by the station licensee for a period of 3 years from date of entry;

(2) Station logs which include entries of communications incident to or involved in an investigation by the Commission and concerning which the station licensee has been notified shall be retained by the station licensee until such licensee is specifically authorized in writing by the Commission to destroy them;

(3) Station logs incident to or involved in any claim or complaint of which the station licensee has notice shall be retained by such licensee until such claim or complaint has been fully satisfied or until the same has been barred by statute limiting the time for the filing of suits upon such claims.

Note: See Part 42 of this chapter concerning preservation of records of common carriers.

(b) Station logs shall be made available to an authorized representative of the Commission upon request.

(c) Ship station logs shall be fully completed at the end of each voyage and before the operator(s) (or other person(s) responsible under the applicable provisions of this part) leave the ship. Unless otherwise authorized by the applicable provisions of this part, the radio log currently in use shall be kept by the licensed operator(s) of the station and during use shall be located at the principal radio operating room of the vessel. At the conclusion of each ocean voyage terminating at a port of the United States (includes Puerto Rico, and Virgin Islands), the original radio log (or a duplicate thereof) dating from the last departure of the vessel from a U.S. port shall be retained under proper custody on board the vessel for a sufficient period of time (not more than 24 hours) to be available for inspection by duly authorized representatives of the Commission. After retention on board the vessel as herein stipulated, the original log (and the duplicate log if provided) may be filed at an established shore office of the station licensee, and shall be retained as stipulated by paragraph (a) of this section.

Note: Duplicate logs are not required by the provisions of this paragraph, unless the original log is removed prior to opportunity for official inspection.

(d) Logs of ships of the United States containing entries required to be made by reason of the Great Lakes Agreement or 83.368(c) of this part shall be kept at the principal radiotelephone operating location while the vessel is being navigated. All entries in their original form required by said agreement or 83.368(c) shall be retained on board the vessel for a period od not less than one month from the date of entry. After retention on board the vessel as herein stipulated, the entries shall be filed at a place where they will be commission upon requent, and shall be retained as stipulated by paragraph (a) of this section.

single feature (headings) was too small to find significant differences in the performance of users. Similar results have been found in other studies using short clearly-written passages with and without informative headings (Swarts, Flower and Hayes, 1980) and in studies of plain English prose passages with inserted violations of single guidelines (Swaney, 1981).

On the other hand, we were able to observe significant differences in user performance when we tested a large document, in which an entire set of document design principles was used, against a traditionally-written public document (Felker and Rose, 1981). This audience-centred evaluation served not only to contrast users' performance on the documents, but also to validate the document design principles that were used to produce the rewritten document.

The document that we tested was a regulation from the US Government's Federal Communications Commission (FCC), governing the use of two-way radios on recreational boats. The Federal Communications Commission is the US Government agency concerned with regulating all non-government wire and radio communications. The FCC's sphere of influence and responsibility is immense: it oversees all public and commercial TV and radio broadcasts, it regulates interstate telephone rates, and it supervises all two-way radio services used by police and fire departments and used in aviation and marine operations. In carrying out its mandate, the FCC publishes, monitors, and enforces many thousands of rules and regulations aimed at audiences ranging from lawyers and engineers to the truck driver buying a CB radio. And, in common with other government agencies, most of the rules and regulations issued by the FCC are hard to read and understand.

The FCC has been in the forefront of Federal agencies to take serious steps to simplify some of its rules and regulations. The rules governing the citizens band radio service (CB radios) were rewritten into plain English in 1978. In a second

major effort, the FCC decided to revise the rules for radios on recreational boats.

Probably the most important guideline used in revising the FCC's marine radio rules for recreational boaters was one that would say 'select only the content that the audience needs.' The rules for recreational boaters were originally mixed in with rules for ocean liners and merchant ships and were loaded down with exceptions and rules to handle unusual cases. The set of revised rules is much shorter than the original.

After the FCC reorganized and rewrote the regulation, they asked the Document Design Center for assistance in planning and carrying out an evaluation of the rules. We reviewed the questions and the plan and we helped analyze and interpret the data.

The evaluation paradigm we used was a straightforward comparative, experimental design. Subjects were 53 experienced and 52 inexperienced recreational boaters. Half of each group were given the new marine radio rules; half were given the old rules. All subjects answered the same 13 test questions that concerned information they had read in the rules. Because some questions had multiple parts, a person could score up to 20 points in answering the 13 questions.

The 'old' rules consisted of 49 pages that contained information related to the use of marine radios in recreational boats. These pages were taken from a larger volume of the original regulations. An index was specifically created to help the subjects locate particular rules in the 49 page document. The new rules consisted of an 11-page booklet organized into 22 rules. The new rules cover the same general content areas and issues as the old rules, but were selected for recreational boaters and completely reorganized, redesigned, and rewritten. Figure 2 and Figure 3 show how the same issue is treated in the old rules and the new ones.

The results show that the new rules were significantly easier for subjects to use and understand. Subjects using the revised rules

Figure 3. Sample rule from FCC's revised marine radio rules for recreational boaters

VHF Marine Rule 15

Do I have to keep a radio log?

(a) You must keep a radio log. A radio log is a book in which you keep information about your radio. The radio log must be neat and orderly. Each page of the log must be numbered, signed by the operator, and show the name and call sign of your boat. You must keep your radio log for at least one year after the day of the last entry in the log.

(b) You must make the following entries in your radio log:

(1) Each distress (MAYDAY) message you send or hear;

(2) Each urgency (PAN PAN) or safety (SECURITY) message you send; and

(3) The installation and servicing of your radio.

(c) For more information on distress messages, urgency messages and safety messages, see VHF Marine Rule 19.

answered more questions correctly. They were significantly better in identifying the proper rules. They took less time to answer questions. They also rated the revised rules as easier to use.

Let us look at each of these results in greater detail. Figure 4 shows that subjects using the new rules answered more test questions correctly. Those using the new rules scored an average of 16.85 points (out of 20) compared to 10.66 points for those using the old rules. Both experienced and inexperienced boaters answered more items correctly with the new rules. As might be expected, the experienced boaters performed better with the old rules than inexperienced boaters, but this difference disappeared with the use of the new rules.

Figure 5 shows the average number of rules each group identified correctly when using the new or old rules. Subjects using the new rules correctly identified an average of 10.47 rules (out of 13) and those using the old rules an average of 6.78. Thus both experienced and inexperienced boaters were better able to locate and identify the correct rules needed for answering questions with the new rules.

Figure 6 shows the average amount of time taken by the two groups to answer each test question using the old and new rules. Included in this is the time taken to read the questions, locate the proper rule, read the rule, and answer the questions. Subjects using the new rules took significantly less time to answer test questions (1.62 minutes on the

Figure 4. The average number of questions answered correctly (out of 20) by using old and new recreational boating rules

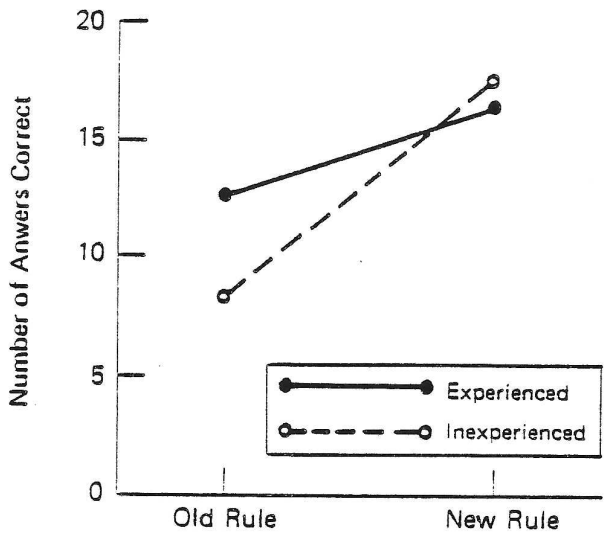
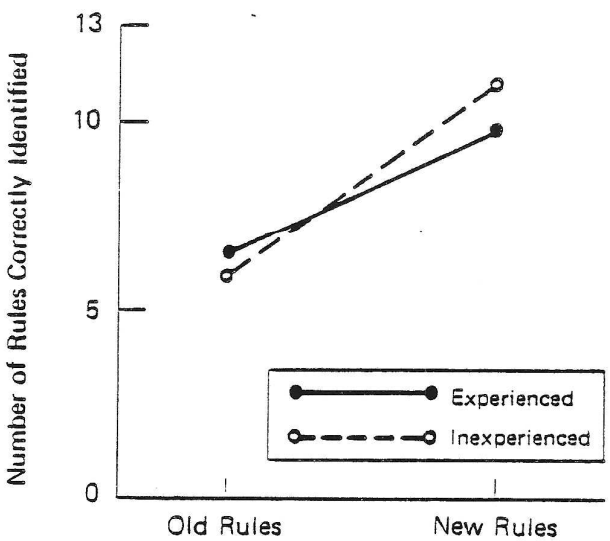


Figure 5. The average number of rules identified correctly (out of 13) by using old and new recreational boating rules



average, vs. 2.97 minutes). This result is predictable because the 49 pages of the old rules are inherently more cumbersome and time-consuming to use than the 11 pages of the new rules.

All subjects rated how difficult it was to use the particular set of rules that was assigned to them. The rating scale ranged from 5 (hard to use) to 1 (easy to use). Figure 7 shows the differences in the ratings given by the two groups. It is clear from the figure that both experienced and inexperienced boaters judged the new rules to be much easier to use.

This audience-centered evaluation validates both the process model (which requires attention to audience and purpose) and the set of document design guidelines that were used to revise the regulation (including guidelines on selecting content and organizing for the audience as well as guidelines for clear writing and attractive layout). The results were also used diagnostically to further revise the two rules that several subjects misinterpreted. These results clearly demonstrate that the use of simpler language, distinctive headings, logical presentation, and the elimination of technical jargon and superfluous information does make a difference. This evaluation, conducted by government writers with the help of researchers, also shows that non-research trained document designers can plan and conduct a useful empirical evaluation of a public document.

Figure 6. The average time taken to answer each test question using old and new rules

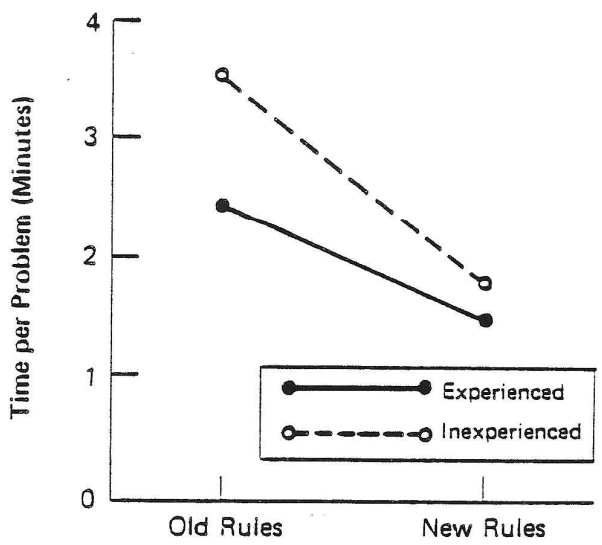
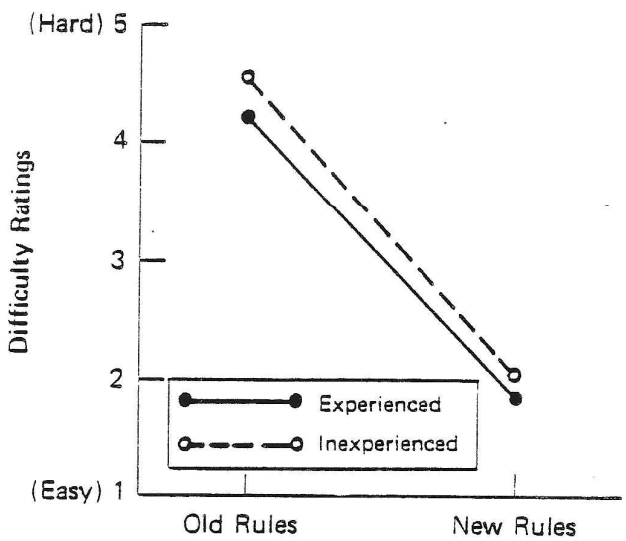


Figure 7. The average ratings of how difficult it was to use the new and old rules



Charrow RP and Charrow VR 1979
Making legal language understandable: a psycholinguistic study of jury instructions
Columbia Law Review, 79/7

Charrow VR and Redish JC 1980
A study of standardized headings for warranties
Document Design Project, Technical Report no. 6
American Institutes for Research, Washington DC

Felker DB and Rose AM 1981
The evaluation of a public document: the case of FCC's marine radio rules for recreational boaters
Document Design Project, Technical Report no.11
American Institutes for Research, Washington DC

Hartley J, Trueman and Burnhill P 1980
Some observations on producing and measuring readable writing
PLET, 17, 164-174

Holland VM 1981
Psycholinguistic alternatives to readability formulas
Document Design Project, Technical Report no.12
American Institutes for Research, Washington DC

Joenk RJ (ed) 1981
IEEE Transactions on Professional Communication: special issues on making information usable, PC-24, 1, March
The Institute of Electrical and Electronic Engineers, New York

Kern RP 1979
Usefulness of readability formulas for achieving Army readability objectives: research and state-of-the-art applied to the Army's problem
Technical Advisory Service, US Army Research Institute, Fort Benjamin Harrison, Indiana

Kintsch W and Vipond D 1977
Reading comprehension and readability in educational practice and psychological theory
Paper presented at conference on memory, University of Uppsala, June

Klare GR 1976
A second look at the validity of readability formulas
Journal of Reading Behaviour, 8, 129-152

Klare GR 1979
Readability standards for army-wide publications
Evaluation Report 79-1
US Army Administrative Center, Fort Benjamin Harrison, Indiana

Morris LA, Thilman D and Myers AM 1980
Application of the readability concepts to patient-oriented drug information
American Journal of Hospital Pharmacy, 37, 1504-1508

Redish JC 1981
Research in the junior year writing course
in MM Marcuse (ed)
Inaugural conference for the junior year writing program
University of Maryland, College Park, Maryland

Swaney J 1981
personal communication to authors

Swarts H, Flower LS and Hayes JR 1980
How headings in documents can mislead readers
Document Design Project, Technical Report no.9
American Institutes for Research, Washington DC