MEASURING THE READABILITY OF TRAINING MATERIALS BY THE PLATO IV COMPUTER... (U) SCHOOL OF APPLIED AEROSPACE SCIENCES CHANUTE AFB, IL G P SCHARF 15 SEP 74 CH-74-604 UNCLASSIFIED SBI-AD-F630 042
CHANUTE PROJECT REPORT 74-60L

MEASURING THE READABILITY OF TRAINING MATERIALS BY THE PLATO II COMPUTER-BASED INSTRUCTIONAL SYSTEM

17 September 1974

This document has been approved for public release and sale; its distribution is unlimited.

83 08 23 107
REPLY TO
ATTN OF
TTGH

SUBJECT: Request for Scientific and Technical Reports

TO: Administrator
Defense Technical Information Center
ATTN: DTIC-DDAB P83-0290
Cameron Sta BG 5
Alexandria VA 22314

1. I regret that it took so long to respond to your request. I was unable to locate the document at Chanute. A copy was finally located at our Headquarters' Tech Advisory Service Branch.

2. Notice that page 1 is missing from the document. As soon as I receive a copy I will send it to your office.

PAUL C. ASCHENBRENNER
Chief, PLATO Training Branch

1. DTIC-DDAB P83-0290 Ltr, 18 Jul 83
2. CH 74-604
DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY PRACTICABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.
ACKNOWLEDGEMENT

I wish to acknowledge the assistance of and express my thanks to the Army Corps of Engineers, the typed, the technical writers, and the training specialists who participated in this report. Special appreciation goes to Bill W. O. Dennis, O.G. and Chief, Dennis W. N. M., and A. W. and J. W. for their advice and assistance. Without their interest and guidance participation in this report would not have been possible.
MEASURING THE READABILITY OF TRAINING MATERIALS BY THE PLATO IV COMPUTER-BASED INSTRUCTIONAL SYSTEM

ABSTRACT

This report describes an experiment in which the PLATO IV computer-based instructional system was programmed to measure the reading grade levels of written training materials. The test, in the PLATO system that performs this task is called PRLD (platoon leader), which stands for PLATO Indicated Reading Level. In one of the PRLD experiments, samples of technical training materials from five different career fields were checked for their grade level by 14 technical writers using the PLATO system. The new samples were then checked by nine education/training specialists using the Plato system. The samples were then typed into the PLATO IV system by six clerks/typists and then by seven PLATO IV authors.

While the results showed some variation in grade levels within all four groups, the mean grade levels of each sample for all groups were quite close. An analysis of variance of the all-sample means of the four groups indicates no statistically significant differences among them. It is concluded that the PLATO IV lesson yields are an easily used and valid means for determining difficulty levels of written materials as a grade level.

GEORGE E. JONES, GS-11
Chief, Mat & Sp Plans Section

This report has been reviewed and is approved.

WILLIAM J. BLESCHER, Colonel, USAF
Team Leader
<table>
<thead>
<tr>
<th>Section</th>
<th>CONTENTS</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>BACKGROUND</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>METHOD</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>RESULTS AND DISCUSSION</td>
<td>7</td>
</tr>
<tr>
<td>E</td>
<td>CONCLUSIONS AND RECOMMENDATIONS</td>
<td>12</td>
</tr>
<tr>
<td>F</td>
<td>BIBLIOGRAPHY</td>
<td>22</td>
</tr>
</tbody>
</table>
instruction system to Chanute. The terminals are supporting a service test to see just how computer-assisted instruction might be used in teaching a technical training course. The project is monitored at Chanute by the Training Research Applications Branch. Because branch personnel had been interested in the problem of the reading difficulty of training literature for several years, it was proposed that PLATO IV be investigated for possibilities for automating the determination of reading grade levels of training materials. The remainder of this paper will discuss the results of this investigation.

SECTION C - METHOD

5. In deciding the ground rules for programming lesson PIRL in the PLATO system to determine reading difficulty levels, the experiences with past systems were used as the basis for the program parameters.

a. The PLATO IV terminal uses a keyboard similar to that of a regular typewriter. The first ground rule set up was that the readability determination system should require no variations from normal typing procedure. This ground rule has been modified to the extent that no periods are used in a sample being checked except to indicate the end of a sentence. Abbreviations and terms such as etc., i.e., and e.g. are typed into the system without the period.

b. Another rule adopted was that numbers would not be typed into the system; e.g., $1,000,000.

c. The third ground rule established was that the PIRL readout should be in grade level rather than on some other scale. In previous experiments, we have found that our technical writers (as well as training specialists and supervisors) can interpret a grade level, but do not relate well to a numerical scale.

7. In developing a formula for the PLATO system to use to determine grade level, the first consideration was to establish that the factors used relate to those found in other reading level indices. The element relating to sentence length (words per sentence) is identical to that found in most currently used indices such as the Flesch and Fry Formulas; so further verification for using this component in the PIRL formula was considered necessary. The verification of the relationship within the word structure was also virtually self-evident since an association does exist between the number of letters in a word and the number of syllables it contains. Also, the average number of letters per word apparently bears a fairly close relationship to the proportion of words included in a list of most common words, which is the basis of the Dale-
Chall readability formula. Consequently, the two factors used in the
PIRL formula (average number of letters per word and average number of
words per sentence) should provide a readability index valid with other
systems if proper weightings for each factor are determined.

a. To obtain a grade level reading, samples to be measured are
selected and the material is typed into the PLATO lesson called PIRL.
Detailed instructions on how to use this lesson are included in the
program. However, these are not called up unless the user indicates he
wishes them by pushing the "Help" key. The first screen display seen
with signing in PIRL is shown in Figure 1. After a sample has been
typed into the PIRL lesson, the "Next" key is pushed. The computer then
begins inspecting the words and assigns a point value to each word as
follows:

1. 1 or 2 letters ............... 1 point
2. 3 or 4 letters ............... 2 points
3. 5 or 6 letters ............... 3 points
4. 7 or more letters ............ 4 points

b. Following the word count, the computer counts the number of
complete thoughts/sentences as indicated by a period, question mark,
exclamation mark, colon, or semi-colon. The total number of word points
is then divided by the number of sentences. This answer is then divided
by three, which is then displayed on the screen as the PIRL, which is
interpreted as a grade level. An illustration of how this appears on the
screen may be seen in Figure 2.

c. If the "Stop" Next" keys are pressed, the material on the screen
is erased and a new sample may be typed into the system. However, if
a sample is typed into the system and the typist wishes to save the
information concerning the sample, the "Data" key is pressed. A data
collection file is then started and information on this first sample then
appears on the screen. When the "Next" key is pressed, this data is
saved while another sample is typed into the system. Data for up to
15 samples may be held in this way, with the mean of all the samples
automatically calculated whenever a new sample is added. A data
collection file screen display may be seen in Figure 3. This file
can only show the grade level of each sample and the mean of the
samples, but also shows the sentence count, total word count, and the
numbers of words with 1, 2, or 3 letters; 4, 5, or 6 letters; 7, 8, or
9 letters; and 10 or more letters.

d. The appropriateness of the weightings given to word lengths was
established by correlating grade levels calculated by PIRL with grade
P.I.R.L.

Plato Indicated Reading Level

TWO GROUND RULES:

1. Omit ALL periods EXCEPT at the end of sentences.

2. Numbers DON'T count, so omit them.
   (numbers = numerics i.e. 3, 187, 16 etc.)

-Figure 1.- First Display Seen When Signing Into Lesson PIRL
This report discusses an experiment in which the PLATO IV computer based instructional system was programmed to measure the reading (grade) levels of written training materials. The lesson in the PLATO system that performs this task is called PIRL, which stands for PLATO Indicated Reading Level. In this Air Force experiment, samples of technical training materials from five different career fields were checked for their grade level by technical writers using the Flesch count system. The same samples were then checked by nine education/training specialists using the Flesch system. The samples were then typed in the PLATO IV system by six clerk typists and then by seven PLATO authors. An analysis of variance of the all-sample means of the four groups indicates no statistical significant difference among them. It is concluded that the PLATO IV lesson PIRL is an easily used and valid means for determining and expressing the difficulty level of written materials as a grade level.

Reading level...15.3
Number of sentences.....7 Total number of words...158
-SHIFT NEXT- for new sample, -DATA- to store this data
Time... 7.7 minutes

Figure 2. Screen Display of Sample and Its Data
to "the form for 10 more samples before
due and collection time is finished.

<table>
<thead>
<tr>
<th></th>
<th>(4&lt;8)</th>
<th>(8&lt;12)</th>
<th>(12&lt;16)</th>
<th># Ureas</th>
<th>%SEM</th>
<th>P.T.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>0</td>
<td>1.2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>13</td>
<td>2</td>
<td>4.3</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>11</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>18</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>8</td>
<td>4</td>
<td>40</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

| 6       | 42       | 26       | 20         | 116      | 1    | ......TOTAL |
| 5       | 16       | 5        | 4          | 13       | 1    | ......MEAN  |

Year P.T. of 5 samples...12.5

This "to score...a rate
-CHART DATA...to omit this
6 to one short...from sample

Figure 3: Screen Display of Data
Collection File of Five Samples.
levels of the same samples established by other systems. Thus, the FIEL formula, like most other readability formulas, is derived from ratios representing word difficulty (number of letters per word) and sentence difficulty (number of words per sentence).

1. To obtain a relatively broad base for verifying the FIEL formula, initial work was done with five, 100-word samples. Grade levels were determined by the Flesch count, the fog count, and by FIEL and found to have a high degree of correlation. Then, to verify the FIEL lesson in an operational situation, the following experimental design was developed.

2. Fifty-word samples of random paragraphs from existing technical training materials were selected. Each sample was from an entirely different career field; i.e., weather, missiles, pneumatics, i.e., engines, and automotive.

3. Twenty-seven technicians/authors from the five training departments were asked to perform a fog count of the samples and record their findings. However, no technician performed a fog count on material originated in his department. As it turned out, there were 21 author/technicians who did a fog count on each sample.

4. Five typists, unfamiliar with the PLATO system, were asked to type the five samples into the PLATO system and to determine and record the grade level of each sample.

5. Several authors were asked to type the five samples into the PLATO system and to determine and record the grade level of each sample.

6. Nine Education and Training Specialists with curriculum experience were asked to perform a Flesch count of the five samples. This count was then converted to its grade level using the table in Flesch's book, How to Test Readability.

SECTION D - RESULTS AND DISCUSSION

1. The results achieved by the four groups of people used in this study are shown in Table 1 through...

   a. Fog count results are shown in Table 1. Differences in the grade levels obtained by different writers can be easily seen in the range for each sample.
The grade level of each sample was determined by the typists, and the results are shown in Table 1. Because typing is essentially a mechanical procedure, one would expect a very high correlation of the typists' determinations of grade levels. However, the PIRL method does depend upon typing accuracy, and with the relatively small samples used in this experiment, typing errors did cause noticeable grade level variations. The typing errors, in most cases, were probably caused by the different touch of the PIRL terminal keyboard.

Results achieved by the PLATO authors using lesson PIRL are shown in Table 1.

---

<table>
<thead>
<tr>
<th>Sample</th>
<th>Grade Level Range</th>
<th>Mean Grade Level</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11-12 = 1.1</td>
<td>13.2</td>
<td>1.0</td>
</tr>
<tr>
<td>II</td>
<td>13-14.2 = 3.2</td>
<td>13.6</td>
<td>1.2</td>
</tr>
<tr>
<td>III</td>
<td>14.1-14.5 = 7.8</td>
<td>15.5</td>
<td>1.7</td>
</tr>
<tr>
<td>IV</td>
<td>6.3-11.7 = 5.5</td>
<td>8.8</td>
<td>1.0</td>
</tr>
<tr>
<td>V</td>
<td>11.6-15.0 = 2.5</td>
<td>12.6</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Table 1. Results of the Manual Fog Count of Five Samples by 21 Technician/Authors.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Grade Level Range</th>
<th>Mean Grade Level</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.7-13.5 = 1.1</td>
<td>13.5</td>
<td>1.3</td>
</tr>
<tr>
<td>II</td>
<td>13.6-14.8 = 1.7</td>
<td>14.5</td>
<td>1.5</td>
</tr>
<tr>
<td>III</td>
<td>14.2-14.3 = 8.6</td>
<td>14.3</td>
<td>1.6</td>
</tr>
<tr>
<td>IV</td>
<td>14.7-11.2 = 11.7</td>
<td>13.7</td>
<td>1.7</td>
</tr>
<tr>
<td>V</td>
<td>11.2-14.6 = 4.2</td>
<td>12.5</td>
<td>2.1</td>
</tr>
</tbody>
</table>

Table 2. Results of PIRL Grade Level Determinations of Five Samples by Six Clerk/Typists.
<table>
<thead>
<tr>
<th>Sample</th>
<th>Grade Level Range</th>
<th>Mean Grade Level</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1.0-1.9 = 1</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>II</td>
<td>1.0-1.9 = 1</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>III</td>
<td>1.0-2.9 = 1</td>
<td>2.0</td>
<td>0.0</td>
</tr>
<tr>
<td>IV</td>
<td>1.0-2.9 = 1</td>
<td>2.0</td>
<td>0.0</td>
</tr>
<tr>
<td>V</td>
<td>1.0-2.9 = 1</td>
<td>2.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 1. Results of PLAT score level determination of five samples by Seven PLAT Examiners.

For the PLAT samples, variations in grade level in "typing errors" probably occurred because they are not produced during a "typing". Comparing Tables I and II, however, it doesn't seem to be statistically possible to discern that there was practically any grade level variation among the typists using PLAT than among the counters. Both the ranges and the standard deviations are from less than a PLAT expert users.

On the other hand, the PLAT "typing error" was found ranging from 1.0 to 2.9 by the present study. These higher level ranges were shown for this part of

<table>
<thead>
<tr>
<th>Sample</th>
<th>Grade Level Range</th>
<th>Mean Grade Level</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1.0-2.9 = 1</td>
<td>2.0</td>
<td>0.0</td>
</tr>
<tr>
<td>II</td>
<td>1.0-2.9 = 1</td>
<td>2.0</td>
<td>0.0</td>
</tr>
<tr>
<td>III</td>
<td>1.0-2.9 = 1</td>
<td>2.0</td>
<td>0.0</td>
</tr>
<tr>
<td>IV</td>
<td>1.0-2.9 = 1</td>
<td>2.0</td>
<td>0.0</td>
</tr>
<tr>
<td>V</td>
<td>1.0-2.9 = 1</td>
<td>2.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 2. Results of PLAT score level determination of nine pairs among typing and training specialists.
the study because of their background and experience in curricula, their educational level, and the fact they should be more skilled at identifying syllables, which is the heart of the Flesch system. Even higher level people are not infallible, and in one 100-word sample, there was a count variation of 40 syllables between the highest and lowest counts. It is also important that one be aware that the Flesch system has a range rather than one value to convert to grade levels. For example, a Flesch raw score of 42-46 approximates a grade level of 11. The range from 47-51 is a grade level of approximately 13. The range of 51-57 is approximately 11.5. It is this range factor that accounts for the fact that for sample IV every person using the Flesch system came up with a grade level of 11.0 for the sample, even though there were variations in their syllable counts.

12. Possibly the most significant table for validating the PIRL formula, however, is Table 5. Here is shown the mean of all the five samples added together for each method. In other words, this is the mean grade

<table>
<thead>
<tr>
<th>Method and Groups</th>
<th>Mean Grade Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Grade Level for 21 Technician/Writers Using the &quot;fog count&quot; method</td>
<td>12.50</td>
</tr>
<tr>
<td>Mean Grade Level for Six Clerk/Typists Using the PLATO IV lesson called PIRL</td>
<td>11.62</td>
</tr>
<tr>
<td>Mean Grade Level for Seven PLATO IV Authors Using the lesson called PIRL</td>
<td>11.80</td>
</tr>
<tr>
<td>Mean Grade Level for Nine Education and Training Specialists using the &quot;Flesch&quot; count</td>
<td>12.75</td>
</tr>
</tbody>
</table>

Table 5. Mean of Grade Levels for the Five Test Samples Treated as One, Large Sample.

Levels of an equivalent 1,000 word sample. As can be seen, the means for all methods used to determine grade levels are quite close. An analysis of the variance of the grade levels found by the four groups is shown in Table 6, and indicates that there is no statistically significant difference in the grade level determinations. All computations for this study were performed using statistical packages available in the PLATO IV system.

\[
P = .312 \quad \text{d.f.} = 3, 39 \quad F = 0.0167
\]

Table 6. Analysis of Variance of the Grade Levels of 2,000 Word Samples by Four Groups.
### CONCLUSIONS AND RECOMMENDATIONS

It should be kept in mind that there are several factors concerning readability which formulas cannot measure. Such things as imagery, subject matter, word order, and line organization are also quite (if not more) important than word and sentence length in determining readability. In addition, formula scores can be inaccurate due to errors in sampling or in their application. However, difficulty levels based upon word and sentence length serve as good indicators of the understandability of the material. High grade levels can alert a writer that there may be a need for improvement in his or her material in order to improve its quality of being easily understood. This is especially relevant in Air Force technical writing where experience has shown reading aptitude may be rather low even when the student is a high school graduate and has high mechanical or electronic A&E scores. Thus, readability ratings can and do serve a useful purpose as a tool to help make one's writing clearer and more understandable.

In a part of the analysis and explanation of this study, the use of PIRL in the "PIRL II" system appears to be adequately used and will make for determining and expressing the understandability of written materials at a grade level. There are now PIRL terminals at three AF research centers, Lomay and Shippens. It is recommended that training departments and technical writers in these cases take advantage of this and use lesson PIRL to check the reading levels of their training materials.
SECTION F - BIBLIOGRAPHY


5. Flesch, Rudolph, How to Test Readability, Harper and Brothers, N.Y., 1951.


*New AFM 13-2, Guide for Air Force Writing. The full length method of grade level determination has been deleted from this pamphlet.

The FIAL for this report is 11.6, a reading level of "difficult" in the Flesch system. Less than 30% of the adult population in the United States can read at this level.