Technical Report

IBM

Competitive Evaluation of Documentation: Surveys and Automated Readability Measures

James R. Lewis
Beth Tripi

February 28, 1992 (Revised August 5, 1993)

TR54.647

Unclassified

Competitive Evaluation of Documentation: Surveys and Automated Readability Measures

James R. Lewis
Design Center/Human Factors
IBM Boca Raton, FL

Beth Tripi
Information Development Editing and Test
IBM Boca Raton, FL

Abstract

User publications for IBM PS/2 (TM) systems have not scored well in recent surveys. To evaluate these results, we examined samples from IBM and competitors' publications using automated reading measures and a graphics/text ratio. Our goal was to answer two questions: Are there reliable differences in writing style between IBM and its competitors? Are these differences related to position in the surveys? The IBM samples did well for the measures based on average words per sentence and average syllables per word (reading grade level and fog index). The IBM samples did poorly for the cloudiness count, a measure based on the frequency of usage of abstract terms and passive structures. There were no significant differences among the samples for the graphics/text ratio. Only the cloudiness count was significantly related to position in the surveys. The results do not prove that cloudiness is the sole cause of position in the surveys, but do indicate that cloudiness may have some influence. We concluded that it is important to track and improve the IBM PS/2 cloudiness count by avoiding passive structures and abstract concepts. These conclusions appeared to be justified when Information Development adopted this requirement which, along with other library restructuring, led to:

- Infoworld rating of excellent for new publications, a 4-point jump from previous ratings of poor.
- 1992 Society for Technical Communications awards.
- Atlanta HelpCenter reported 80% drop in customer calls about publications.

Table of Contents

Introduction

Recent surveys (Dataquest, 1991; IBM, 1991) show that personal computer users rate the publications for the IBM PS/2 systems lower than most competitive publications. To improve the publications, Information Development has focused on the design of covers and publications layout (E. Musselman, personal communication) and customer surveys (R. Cherry, personal communication).

We hypothesized that there may be differences in competitive writing styles that automated readability analysis can detect. Readability formulas provide good indices of difficulty by using counts of language elements in the text (Bailey, 1989; Klare, 1974-1975). The IBM VM tool "READABLE" (Hardiman, 1991) analyzes the readability of text, and provides the following measures:

- Reading grade level (low score is better)
- Cloudiness count (low score is better)
- Flesch index (high score is better)
- Fog index (low score is better)
- British reading age (low score is better)
- Kincaid index (low score is better)

Analyses of the average length of sentences and the average number of syllables per word determine the reading grade level, Flesch index, fog index, British reading age and the Kincaid index. The frequency of abstract terms and passive structures determines the cloudiness count. (See Hardiman, 1991 for a detailed description of these measures.)

The goal of this study was to answer two questions. First, are there detectable differences in writing style between IBM and its competitors? Second, are any of these differences related to a publication's position in the recent surveys?

Method

We evaluated text samples from the publications for the following systems:

- o IBM Model 55
- o IBM Model 70
- o Compaq Deskpro 386
- o Epson Equity Ie
- o Apple Macintosh
- o Tandy 4025 LX
- o Dell System 316

Bailey (1989) recommends that analysts use at least five 100- to 150-word samples to estimate readability. We used a stratified random selection procedure to select ten text samples containing at least 200 words from each system's user publications. We divided each document set into ten equal parts and used a random number table to select a random sample within each part. Next, we typed each sample (excluding headings) into a file and analyzed it with READABLE. For each sample we also measured the square inches of text and the square inches of graphics and calculated the graphics/text ratio. After the analyses were complete, we entered the data into a database for analysis with the Statistical Analysis System (SAS).

Results

Correlations Among the Measures

Table 1 shows the correlations among the dependent measures. Note that the reading grade level, Flesch index, fog index, British reading age and Kincaid index are highly correlated. These indices are essentially measuring the same thing because they all are based on average words per sentence and syllables per word. Later analyses will focus on the reading grade level and the fog index from this variable set because IBM Boca Raton Information Development already uses them. Neither the cloudiness count (based on frequency of occurrence of abstract terms and passive structures) nor the graphics/text ratio (ratio of page areas devoted to graphics and text) are highly related to other measures. We will analyze both of these indices because they measure text characteristics that are different from the reading grade level and the fog index.

Table 1. Correlations Among the Dependent Measures

	GRADE	CLOUDY	FLESCH	FOG	BRITISH	KINCAID	G/T
GRADE	1.00	0.36	-0.99	0.86	0.99	0.91	-0.08
CLOUDY	0.36	1.00	-0.37	0.09	0.31	0.16	0.08
FLESCH	-0.99	-0.37	1.00	-0.86	-0.99	-0.90	0.07
FOG	0.86	0.09	-0.86	1.00	0.90	0.94	-0.07
BRITISH	0.98	0.31	-0.99	0.90	1.00	0.95	-0.07
KINCAID	0.91	0.16	-0.90	0.94	0.95	1.00	-0.09
G/T	-0.08	0.08	0.07	-0.07	-0.07	-0.09	1.00

Analyses of Variance for Selected Readability Measures

Analyses of variance indicated significant main effects of publications for the cloudiness count ($F(6,63)=5.5$, $p<.0001$, see Figure 1), fog index ($F(6,63)=4.3$, $p=.001$, see Figure 2) and reading grade level ($F(6,63)=3.2$, $p<.008$, see Figure 3). Differences in the graphics/text ratios among the document sets were not statistically significant ($F(6,63)=1.3$, $p=.26$, see Figure 4). These results show that there are detectable readability differences among the publication sets. IBM publications were competitive for the reading grade level and fog index, but not for the cloudiness count.

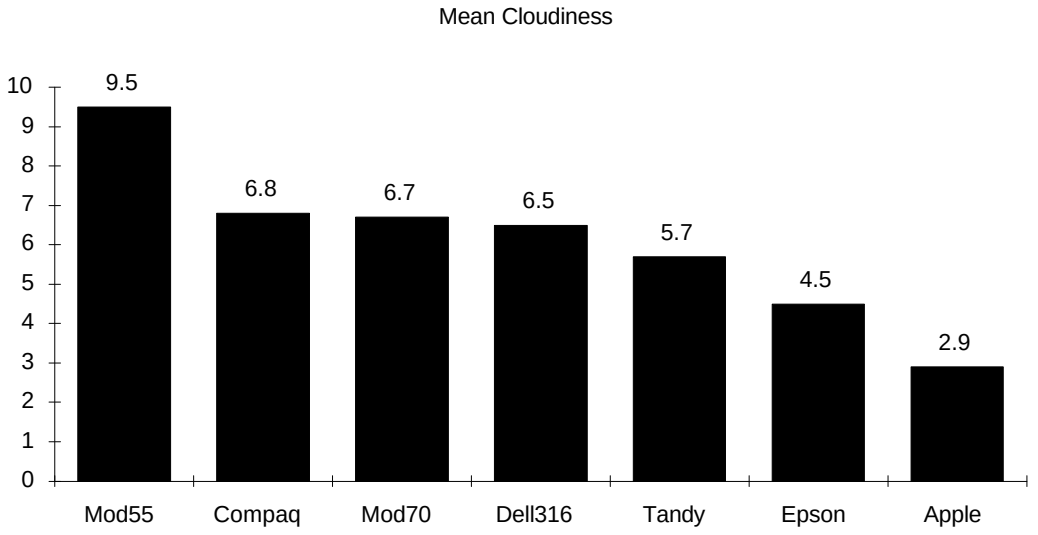

Figure 1. Differences in Publications: Cloudiness Count

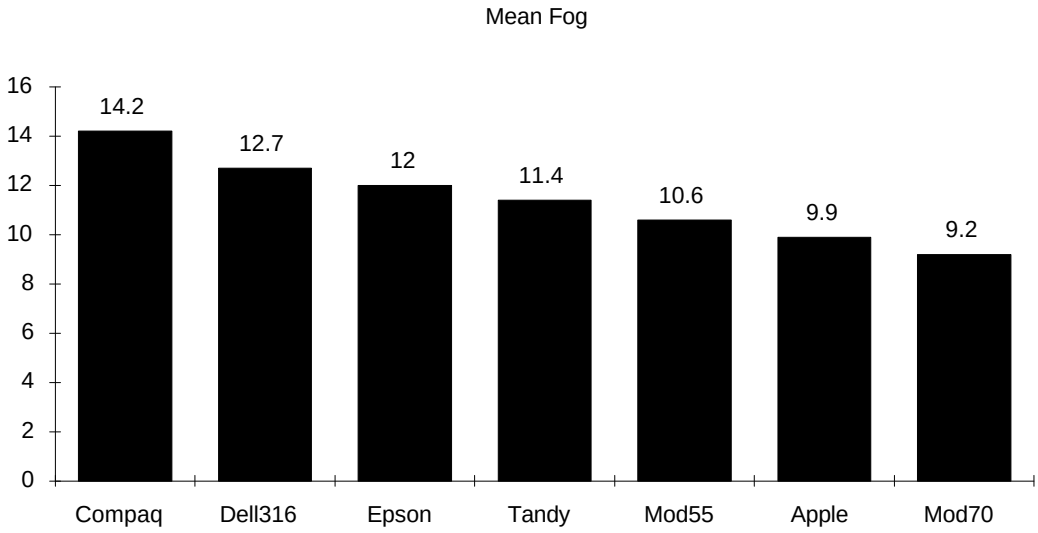

Figure 2. Differences in Publications: Fog Index

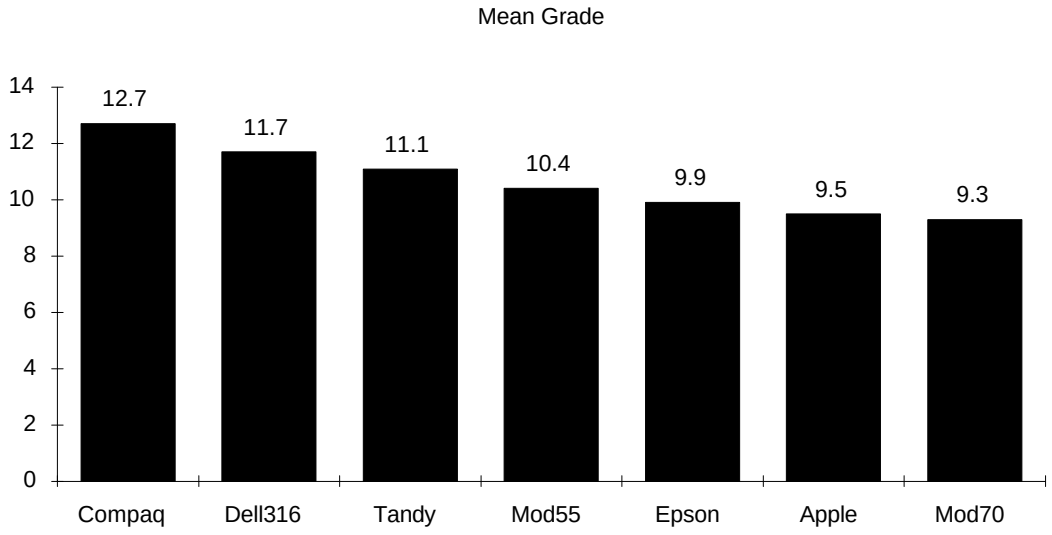

Figure 3. Differences in Publications: Reading Grade Level

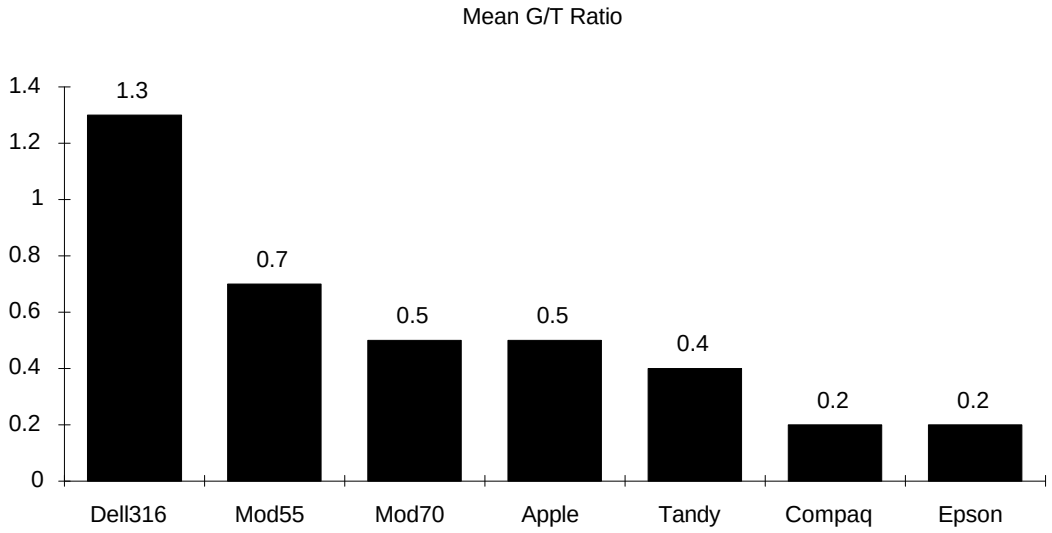

Figure 4. Differences in Publications: Graphic/Text Ratio

Position in Surveys: Dataquest and IBM

Table 2 shows the ratings and rankings for the publications produced by the companies included in this study. The relative positions of the companies in the two surveys were fairly consistent except for Tandy, which placed third in the Dataquest survey and sixth in the IBM survey.

Table 2. Survey Ratings and Rankings

	Dataquest		IBM	
Company	Overall	Rank	% Dissatisfied	Rank
Dell	3	1	20	1.5
Apple	5	2	20	1.5
Epson	-	-	25	3
Compaq	10	4	31	4
IBM	23	5	32	5
Tandy	7	3	36	6

Relationship Among Measures and Survey Position

Table 3 shows the correlations among the measures and survey positions. Four correlations were large enough to merit attention. The largest correlation was between the reading grade level and the fog index (r=.89, p=.01). This was expected because these indices measure the same text characteristics using similar formulas. The next largest correlation was between the cloudiness count and the Dataquest survey position (r=.72, p=.10). Next, the Dataquest and IBM surveys correlated at r=.63 (p=.18), showing that the survey results were reasonably consistent. The fourth highest correlation was between the cloudiness count and the IBM survey (r=.45, p=.30). All other correlations were much too low to consider as evidence of a relationship between the variables. Although only the correlation between reading grade level and the fog index and the correlation between the cloudiness count and the Dataquest survey position are statistically significant (using α=.10), the correlation set suggests that the cloudiness count is more highly related to survey position than the other readability measures.

Table 3. Correlations Among Measures and Position in Surveys

SPEARMAN CORRELATION COEFFICIENTS
/ PROB > |R| UNDER H0:RHO=0

	DATAQ	IBM	CLOUDY	FOG	GRADE	G/T
DATAQ	1.00	**0.63**	**0.72**	-0.38	-0.38	-0.22
	0.00	0.18	**0.10**	0.46	0.46	0.67
IBM	**0.63**	1.00	**0.45**	-0.25	-0.02	-0.20
	0.18	0.00	0.30	0.58	0.97	0.66
CLOUDY	**0.72**	**0.45**	1.00	0.14	0.36	0.22
	0.10	0.30	0.00	0.76	0.43	0.64
FOG	-0.38	-0.25	0.14	1.00	**0.89**	-0.31
	0.46	0.58	0.76	0.00	**0.01**	0.50
GRADE	-0.38	-0.01	0.36	**0.89**	1.00	-0.09
	0.46	0.97	0.43	**0.01**	0.00	0.85
G/T	-0.22	-0.20	0.22	-0.31	-0.09	1.00
	0.67	0.66	0.64	0.50	0.85	0.00

Discussion

These results answer the questions posed in the introduction. There are detectable differences in writing style between IBM and its competitors. IBM publications have low reading grade levels and fog indices (indicative of good readability due to short words and sentences), but have high cloudiness counts (indicative of poor readability due to too many passive structures and abstract concepts). However, the only dependent measure that seems to be related to position in surveys is the cloudiness count.

These results do not prove that cloudiness is the sole cause of position in the surveys, but do indicate that cloudiness may have some influence, certainly more than reading grade level or the fog index.

Conclusions

IBM PS/2 publications do well with reading grade level and the fog index, but that does not seem to be good enough. The other Information Development activities to improve the quality of IBM PS/2 publications (customer surveys, cover design, and page layout) should continue. Information Development should continue to use READABLE to track the reading grade level and fog index to maintain IBM's competitive position. Also, Information Development should use READABLE to track and improve the cloudiness of IBM PS/2 publications.

Followup

Information Development adopted the strategy of reducing the cloudiness count, which resulted in a Model 95 cloudiness of 5.6 and Model 90 cloudiness of 4.3 – a substantial drop from the Model 55 cloudiness of 9.5. This, along with other library restructuring, led to (see March 1993 Currents):

- Infoworld rating of excellent for the new publications, a 4-point jump from previous ratings of poor.
- 1992 Society for Technical Communications awards.
- Atlanta HelpCenter reported an 80% drop in customer calls about publication.

Trademarks

Apple and Macintosh are registered trademarks of Apple Computer, Inc.

Compaq and Deskpro are registered trademarks of Compaq Computer Corporation.

DELL and Dell System are trademarks owned by Dell Computer Corporation.

Epson and Equity are registered trademarks of Seiko Epson Corporation.

IBM and PS/2 are registered trademarks of International Business Machines Corporation.

SAS is a registered trademark of the SAS Corporation.

Tandy is a registered trademark of Tandy Corporation.

References

Bailey, R. W. (1989). Human performance engineering: Using human factors/ergonomics to achieve computer system usability. Englewood Cliffs, NJ: Prentice Hall.

Dataquest, Inc. (April, 1991). Score report: Customer satisfaction – personal computers, Third Quarter 1990 through First Quarter 1991. San Jose, CA: Author.

Hardiman, R. J. (1991). READABLE: A tool to aid in the analysis of the readability of text (IBM paper available on VM). Hursley, England: International Business Machines Corp. IBM Internal Use Only.

IBM Corp. (1991). US customer satisfaction: Personal systems mid-year 1991. Purchase, NY: Author. IBM Confidential.

Klare, G. R. (1974-1975). Assessing readability. Reading Research Quarterly, 10, 62, 102.