A Preliminary Examination of the Readability of Consumer Pharmaceutical Web Pages

Michelle B. Kunz
Morehead State University

Peggy Osborne
Morehead State University

This paper examines the readability level of online consumer pharmaceutical web pages using standard readability indexes. Pages from sixteen online consumer pharmaceutical websites were downloaded and analyzed using the Flesch-Kincaid Grade Level and Reading Ease, Coleman-Liau, Gunning Fog and SMOG readability indexes. Results indicate the majority of web pages are written at a level above the average reading level of U.S. consumers. Most online pages were written at high school graduate or collegiate level. Differences were found in the readability level, based upon the type of material posted on the web page.

INTRODUCTION

The Internet is changing the way Americans engage with information. The 2008 Presidential election is evidence of that. Two major drivers for this change are broadband adoption and personal motivation (Fox, 2008). Like all others, the pharmaceutical industry has seen a significant change in how they communicate with their consumer. This includes the increase in the number of channels pharmaceutical companies now employ to vie for public attention as well as consumer demands for more and better quality health information. The growth of communication channels and methods includes direct-to-consumer advertising of prescription drugs and sales of medical devices and medications over the Internet. As of December 31, 2008, the U.S. is the largest e-market in the world, both in numbers of buyers as well as in volume of transactions; therefore, pharmaceutical companies must immediately address the desire for more online information.

In 2006, Fox reported eighty percent of American internet users went online to look for health information, and that 8 million American adults look online on a typical day. There are close to 17 million consumers age 62 or older in the U.S. (Burns, 2008), who spend collectively, an average of 44 minutes per day individually, on the Internet, and more than one-third, (38%), search for health and health-related information online. More recently, Burns (2009) reported the online population between age 70-75 had grown the most, from 26% of these users online in 2005, to slightly less than half, 45%, online now. Further data supports the continued use of the Internet as a source of health information, as more than three-fourths (82%) of Generation Xers are likely to get health information online. This "wired" generation is even more likely to retain their Internet connection as a source of information in the future.

The Pew Internet Project estimated between 75-80% of Internet users were online looking for health information (Fox, 2008), while Harris Interactive data reported that 81% of Internet users searched for health related information online, and up to 66% of all adults used the Internet to find health information. Choi and Lee (2007) reported that searching for health information (80 percent) had emerged as the third most common activity among on-line users. Similarly, Jones and Fox (Jones & Fox, 2009) reported that for seniors, searching for health information is the third most popular online activity.

In 2001 more than 70,000 web sites disseminated health information (Cling & Haynes, 2001). While the Internet offered widespread access to health information, even then, these authors were concerned about the ability of consumers to evaluate the information. Berland, et al. (2001) suggested that even if online materials are comprehensive and accurate, the ability of users to comprehend the material may not be appropriate. Although the Internet provides pharmaceutical companies the ability to provide information to consumers, there is no way to properly assess the consumer's ability to understand. These authors found the reading level of most Web-based material to be well above the average reading level of U.S. consumers. All the health-related sites in their study required at least a 10th grade reading level and more than half of the sites had material presented at the college level. Fifty-eight percent of respondents to a 2002 survey (Diaz, et al., 2002) reported they used online websites to investigate side-effects of drugs or complications of medical therapy. In addition, 41% reported searching for complementary or alternative medicine and second opinions about medical conditions. As the online communication channel becomes more critically integrated into the consumer's decision-making processes, companies must pay particular attention to how the information is perceived and used by the consumer. Of major concern to the pharmaceutical company is consumer health literacy.

CONSUMER LITERACY SKILLS AND CONTENT READABILITY

In 2001, Solomon reported the average reading level in the U.S. was approximately eighth to ninth grade, and that experts recommended critical information for daily living should be scaled back to the sixth-grade level. In a 2002 study (Sagaram, Walji, & Bernstam, 2002) examining consumer and alternative medicine information intended for consumers, of the sites examined, nearly three-fourths (73%) of the web pages had mean overall readability levels written at grade eleven, far higher than the recommended eighth grade level. Berland et al. (2001) found the average reading level of health information web sites to be collegiate grade 13.2, and ranged from 10th grade to graduate school level. Overall, they found the reading level of most web-based health related information to be quite high.

After conducting an analysis of readability statistics, Eysenbach et al. (2002) found that eleven studies used readability formulas to establish the reading level of a document based on the complexity and length of words and sentences, and nine studies used the Flesch-Kincaid (FK) Grade Level Index. Other formulas used include the SMOG Readability Formula, the Fry Readability Graph, the Gunning-Fog formula, and the Lexile Framework. Many readability indices do not reflect factors that affect comprehension such as frequency of use and explanation of medical jargon, writing style (use of active voice, non-patronizing language, motivational messages, tone/mood, how it relates to the audience), or use of culturally specific information.

The 2003 National Assessment of Adult Literacy completed a national representative assessment of English literacy among American adults age 16 and older ("National Assessment of Adult Literacy," 2003). The assessment was sponsored by the National Center for Education Statistics (NCES). It was designed specifically to measure adults' ability to use literacy skills to read and understand health-related information. Health literacy was defined by the NCES as the ability to understand and use health-related printed information in daily activities to achieve one's goals and to develop one's knowledge and potential. The NAAL health literacy component assessed responses to health-related tasks presented in written form. These tasks fell into three categories: clinical, prevention, and navigating the health system. Examples of health-related tasks might be to determine the correct dose of a prescribed medication from a prescription label; to understand from written material the health risks of obesity; or to determine the benefits of a health insurance plan. A report on literacy levels of older adults in the U.S. (Ownby, 2005)

showed that as many as 47% of adults aged 60 years or older have only basic reading skills. Results of most studies have shown that few elderly have basic health literacy skills yet most websites present health care information at college reading levels. Writing health-related material at a lower reading level is difficult, given the content required, including drug names, and medical terminology. Many specific and accurate medical terms are likely to be unfamiliar to consumers. Since these words may be longer and therefore more likely to be considered difficult, writers may face the problem of explaining complex medical concepts with simple language. Companies that face the daunting task of creating online health related information should consider the application of some readability index to measure the consumers' ability to understand the written word.

READABILITY INDEXES

The Flesch-Kincaid Readability Test (reading ease) is a government standard ("Flesh-Kincaid Readability Test," 2005). The formula rates text on a 100-point scale, with higher scores meaning the document is easier to read. As a rule of thumb, scores of 90-100 are considered easily understandable by an average fifth grader. Eighth and ninth grade students could easily understand passages with a score of 60-70, while passages with results of 0-30 are best understood by college graduates. *Reader's Digest,* for example, has a readability index of about 65, *Time* magazine scores about 52, and the *Harvard Law Review* has a general readability score in the low 30s. As another example, most states require insurance forms to score 40-50 using this index. The Flesch-Kincaid Grade Level rates text on a U.S. school grade level. Thus, a 6.0 rating indicates that a sixth grader should be able to understand the documents.

The Gunning Fog Index ("Internet Usage Trends in the United States of America," 2008) is a readability index designed to show how easy or difficult a passage is to read. It gives the number of years of education an individual hypothetically needs in order to understand the text he or she is reading. The Fog Index implies that short sentences written in plain English achieve a better score than long sentences written in a complicated language. For reference, the *New York Times* has an average Fog Index of 11-12 and *Time* magazine has an index of about 11. Typically, technical documentation has a Fog Index between 10 and 15, and professional prose almost never exceeds 18 (Illert & Emmerich, 2008).

The Coleman-Liau Index is a readability index that measures the difficulty of understanding text. This score is expressed as the grade level an individual in the United States would need to have completed in order to be able to read and understand the content. The Coleman-Liau Index uses the number of characters in the words to calculate the readability score ("Coleman-Liau Index," 2008).

The creator of the SMOG readability formula, G. Harry McLaughlin (1969) (McLaughlin, 1969) defines readability as: "the degree to which a given class of people find certain reading matter compelling and comprehensible (pg. 639)." This definition stresses the interaction between the text and a class of readers of known characteristics such as reading skill, prior knowledge, and motivation. The formula basically counts the words with three or more syllables in three 10-sentence samples, estimates the count's square root (from the nearest perfect square), and adds 3.

The purpose of this current study was to examine the readability of online consumer pharmacy web site pages using standard readability indexes. The research questions to be addressed in this study were: 1) Are the home pages, privacy policies, patient/consumer information and drug information pages posted on the online pharmacy web sites readable by the average U.S. consumer? 2) Are there differences in the level of readability for online pharmacy web sites, by the type of page/information posted?

Five different indexes that measure overall readability were chosen for the purpose of this study. The Flesch Readability Test, Flesch-Kincaid Grade Level, Gunning Fog, Coleman-Liau, and SMOG index are reading level algorithms used to assess the overall reading ease and understanding of the various online pharmacy web pages reviewed in this study.

METHODOLOGY

Fox reported (2006) that consumers usually begin their search using a search engine. The 2005 study by Ownby sampled text from the first ten health related sites identified by search engine results. Approximately 100 words were extracted from the site main page, and two additional areas on the web site. In a similar analysis, Kunz and Osborne (2006) downloaded the privacy policies of web sites and analyzed the readability using the Flesch-Kincaid Grade Level and Reading Ease indexes. Their study found the privacy policies for 300 top-rated online retailers averaged 11+ grade level using FK Grade Level and 36-38 FK reading ease scores.

For the current study, the search term "online pharmacy" was searched using Google. The first ten online pharmacies identified in the results were used in this study, along with prominent U.S. pharmacies as listed in the *Stores* 2008 Top 100 Retailers (Schulz, 2008). Results for aggregators, alternative or non-prescription drugs/products were excluded from this study. A total of 16 sites were identified for this study. Seven of these were U.S. pharmacies, eight were Canadian, and one was a UK site. The home, patient/customer information, drug information pages, along with the privacy policy were copied and saved into a Word document. Each of the documents was analyzed using Storytoolz ("StoryToolz Resources for Authors," 2008). Storytoolz is a writing tool available online to assist authors in determining word counts and checks readability of information using several different measures. For this study the Flesch-Kincaid Grade Level, Flesch Reading Ease, Coleman-Liau, Gunning Fox, and SMOG indexes were employed.

RESULTS

Two of the sixteen web sites did not have all four of the pages sampled. The Mexrxone.com site did not have a privacy policy page, and the Riteaid site did not have a drug information page. Scores for the individual pages of each site are reported in Table 1. The range of readability scores was quite large, with fifth grade the lowest and post-graduate level at the high end. Mean scores for each of the individual indexes are relatively high—at college-level for all but the FK Grade Level, with a mean of about 10th grade, 60% above the 10th grade, and 15% at the 13+ grade level. The Coleman-Liau scores ranged from 7.7-19.1, with a mean score of 12.5. More than three-fourths, (79%) were at or above the 10th grade level and 32% were at the 13+ grade level. Overall Gunning Fog scores ranged from 8.3 to 22.1, with a mean score of 13.1. More than three-fourths (82%) were at or above the 10th grade level, and 45% 13.0 or higher. The SMOG scores ranged from 8.6-18.1, with a mean score of 11.9. Approximately three-fourths of the SMOG scores were at or above the 10th grade, with 23% above the 12th grade. The range of scores for the FK reading ease ranged from a low (difficult) of 31.3 to a high of 76.9. The mean reading score was 55 and 57% of the pages scored below 60, which is considered to be a readable level. See Table 2 for range of scores across all indexes.

To answer more specific questions, such as, are there differences in readability levels of consumer pharmacy web sites, several one-way ANOVAs were conducted. The first analysis was to compare the readability scores across the different web sites by company. No statistically significant differences were found. Since there were web sites from the U.S., Canada, and UK, a second analysis of variance was administered, and again, no statistically significant differences were found in the readability of sites based upon country of origin (see Table 3). Finally, the third ANOVA compared the readability scores across the different types of web page: home, drug information, patient information, and privacy policy. In this case all results were statistically significant as indicated in Table 4. As mentioned previously, the FK Grade Level scores ranged from 5.1 to 17.3, with a mean score of 9.75. Analysis of results, based upon the page type, (Table 5) indicates the average for patient pages was 8.006, which could be considered relatively readable, while privacy policies were written at almost the twelfth grade level. FK Reading Ease scores ranged from a minimum of 16.9 (very difficult to read) to 76.9—again relatively easy to read, with an average of 55.08. Review of the mean scores across page type revealed drug information and home pages in the mid-50 range, while patient information was slightly easier to read, with a mean of

65.6, and privacy policies scoring a 45.6, more difficult to read. The Coleman-Liau scores ranged from 7.7 to 19.1, with a mean score of 12.5, an average of a high 12th-grade reading level. Examination across the different page types showed the average readability by C-L for the home page was higher, while the patient information page was slightly more readable, with an average score of 10.7. The Gunning Fog scores ranged from 8.3 to 22.1, with an average of 13.19. Examination of the GF scores across the page types found: patient information pages were written at an average GF score of 11.6, the home pages at an average of 12.6, and the drug information pages and privacy policies at 13+. SMOG readability scores are similar, with the patient information pages written at an average of 10.7, while the drug information and home pages 11+, and privacy policies the most difficult to read at an average SMOG score of 13.5.

CONCLUSION

The results of this study indicate the majority of information contained on consumer pharmaceutical websites is written at a level far higher than that which the average consumer can understand. Furthermore, this study found that there are significant differences in the readability level of website information, based upon the type of page information posted. The patient/customer information pages averaged the lowest readability level, while the privacy policy pages averaged the highest readability scores. Thus, as with previous studies, these results indicate the information posted for consumers on online pharmacy websites is written at a reading level much higher than that of the average U.S. consumer. The Internet has become ingrained in the consumer's search for information and products. So, where does the proliferation of the Internet in consumers' lives lead the pharmaceutical industry? It continues to face pressure from government agencies and impending legislation for direct-to-consumer advertising. Although the Internet allows pharmaceutical companies to better target their consumers, Vargas (2008) reported that Internet advertising was down 5% from approximately $167 million in 2006 to about $159 billion in 2007. Even though there was a decline in 2007, Bruce Grant, senior vice president of business strategy at Digitas Health, suggests that with the increasing presence of the Internet in consumers' lives, the future is the Web. Joan Mikardos of Sanofi-Aventis suggested that the industry as a whole was behind in working with the Internet as a promotional medium.

The Internet has shaped current consumers and their decision making processes. Regardless of the industry, if an organization intends to communicate directly with the consumer, the Internet will play a larger and larger role in that process. Illbert and Emmerich (2008) site Pfizer's reduction of their worldwide sales and marketing staff by 20 percent and Astra Zeneca's 10 percent cut of the U.S. sales force as evidence of the changes pharmaceutical companies are being forced to make in an effort to better reach today's consumer. It is likely that pharmaceutical companies will increase their use of blogs, chat rooms, online forums, disease symptom situations, promotional efforts, and other direct-to-consumer communications. The online marketspace will bring an increase in the number of customers; provide the ability to track demographic information, shopping habits, and other information useful in better targeting their market. Overall, the ability for consumers to conduct their transaction with a pharmacy in the online environment may actually lower their costs of operations, improve operating efficiencies, and increase overall revenues. Haubl and Trifts (2000) suggested that when making decisions in the online environment, consumers are often unable to evaluate all alternatives in depth and, as a result, use a two-stage process. First the consumer screens the information and identifies a subset of the best alternatives. Second, the subset is further evaluated, comparisons made, and purchase decision made. As result of this study, these authors recommend that online companies utilize interactive tools to assist the consumer in the initial screening and comparisons. The increasing dependence of consumers on the Internet for health related information will require pharmaceutical companies to be wary of the consumers' use of misleading, inaccurate, or inappropriate information, which may put consumers at unnecessary risk. Although many health professionals agree that the Internet is a boon for consumers because they have easier access to much more information than before, professionals also are concerned that the poor quality of some of the information on the Internet will undermine informed decision-making. Health related web sites must provide high quality health information that is accurate, current, valid, appropriate, intelligible,

and free of bias ("Healthy People 2010 Information Access Project," 2009). First and foremost however, the information posted must be written at a level that is appropriate and readable for the average U.S. consumer.

REFERENCES

Berland, G. K., Elliott, M. N., Morales, L. S., Algazy, J. I., Kravitz, R. L., Broder, M. S., et al. (2001). Health information on the Internet. *JAMA, 285*(20), 2612-2637.

Burns, E. (2008, May 6). How active is the 'Mature' market online? Retrieved March 6, 2009, from http://www.clickz.com/3629395/

Burns, E. (2009). Study: Sizing up the online generation gap. *Clickz.com*. Retrieved from http://www.clickz.com/3632579/

Choi, M., & Lee, W.-N. (2007). Understanding the impact of direct-to-consumer pharmaceutical advertising on patient physician interaction. *Journal of Advertising, 36*(3), 137-149.

Cling, R. J. W., & Haynes, K. M. (2001). Consumer health information seeking on the Internet: the state of the art. *Health Education Research: Theory & Practice, 16*(6), 671-692.

Coleman-Liau Index (2008). Retrieved May 6, 2008, from UsingEnglish.com: www.usingenglish.com/glossary/coleman-liau-index.html

Diaz, J. A., Griffith, R. A., Ng, J. J., Reinert, S. E., Friedmann, P. D., & Moulton, A. W. (2002). Patients' use of the Internet for medical information. *Journal of General Internal Medicine, 17*, 180-185.

Eysenbach, G., Powel, J., Kuss, O., & Sa, E.-R. (2002). Empirical studies assessing the quality of health information for consumers on the World Wide Web. *JAMA, 287*(20), 2691-2723.

Flesh-Kincaid Readability Test (2005, May 15). Retrieved June 1, 2005, from http://en.wikipedia.org/wiki/Flesh-Kincaid_Readability_Test

Fox, S. (2006, October 29, 2006). Online Health Search Retrieved March 5, 2009, from http://www.pewinternet.org/pdfs/PIP_Online_Health_2006.pdf

Fox, S. (2008). The Engaged e-Patient Population. Retrieved from http://www.pewinternet.org/pdfs/PIP_Health_Aug08.pdf

Haubl, G., & Trifts, V. (2000). Consumer decision making in online shopping environments: the effects of interactive decision aids. *Marketing Science, 19*(1), 4-21.

Healthy People 2010 Information Access Project (2009). Retrieved March 24, 2009, from http://www.healthypeople.gov/document/HTML/Volume1/11HealthCom.htm#_Toc490471352

Illert, G., & Emmerich, R. (2008). The need for new promotional materials. *Journal of Medical Marketing, 8*(1), 23-30.

Internet Usage Trends in the United States of America (2008). Retrieved March 24, 2009, from http://www.internetworldstats.com/stats2.htm

Jones, S., & Fox, S. (2009, January 28, 2009). Generations online in 2009 Retrieved March 5, 2009, from http://www.pewinternet.org/pdfs/PIP_Generations_2009.pdf

Kunz, M. B., & Osborne, P. (2006, March). *The Readability of Online Privacy Policies: Can Consumers Understand?* Paper presented at the Marketing Insights, Innovations and Ingenuity: Marketing in the 21st Century, Chicago, IL.

McLaughlin, G. H. (1969). SMOG grading--a new readability formula. *Journal of Reading, 22*, 639-646.

National Assessment of Adult Literacy (2003). Retrieved February 10, 2009, from http://nces.ed.gov/naal/kf_demographics.asp

Ownby, R. L. (2005). Influence of vocabulary and sentence complexity and passive voice on the readability of consumer-oriented mental health information on the Internet. *AMIA 2005 Symposium Proceedings*, 585-588.

Sagaram, S., Walji, M., & Bernstam, E. (2002). Evaluating the prevalence, content and readability of complementary and alternative medicine (CAM) web pages on the Internet. *AMIA Annual Symposium Proceedings*, 672-676.

Schulz, D. P. (2008, July). Top 100 Retailers: The nation's retail power players. *Stores, 90,* 19.

Solomon, S. (2001, November 20). Write Smart, Write Simple Retrieved June 2, 2005, from http://www.clickz.com/experts/design/cont_dev/article.php/925571

StoryToolz Resources for Authors (2008). Retrieved May 5, 2008, from BitPeg Software, Inc.: http://www.storytoolz.com/readability/text_entry

Vargas, S. (2009). Under pressure. *Pharmaceutical Executive, 28*(5), 87-91.

TABLE 1
SCORES FOR MEAN INDEX MEASRUES BY COMPANY

Company	Country	N	FKGrade	FKRead	ColeLiau	GunnFog	SMOG
Drugstore.com	US	4	9.125	55.525	13.525	12.700	11.600
Medrx-one.com	UK	3	8.700	60.900	10.267	11.900	11.067
Walgreens.com	US	4	9.450	55.150	13.950	12.950	11.675
Planetdrugsdirect.com	CA	4	9.700	55.925	12.925	12.575	11.425
24X7pharmacy.com	CA	4	8.400	60.700	11.300	11.925	11.100
Canadadrugs.com	CA	4	11.450	42.025	14.550	15.750	13.275
Drugdelivery.ca	CA	4	11.275	53.900	11.500	14.775	12.625
CVS.com	US	4	7.050	70.850	9.150	9.750	9.450
Canadianpharmacymeds.com	CA	4	10.600	49.825	13.175	14.300	12.750
Onlinecanadianpharmacy.com	CA	4	10.800	53.125	12.350	14.075	12.500
Pharmacy-online.ca	CA	4	9.050	56.775	11.550	12.300	11.425
Edrugstore.md	US	4	10.275	51.025	14.650	14.325	12.800
Pharmacyrxworld.com	CA	4	9.275	55.500	12.125	12.825	11.725
Riteaid.com	US	3	11.967	42.367	14.567	14.500	12.900
Walmart.com	US	4	8.925	61.450	11.475	12.650	11.500
Meijer.com	US	4	10.250	54.450	13.100	13.700	12.250
Total		62	9.750	55.076	12.513	13.187	11.876

TABLE 2
MEANS FOR INDIVIDUAL INDEX MEASURES

Scale	Min	Max	Mean	S	n
FK Grade	5.1	17.2	9.750	2.9004	62
FKRead Ease	16.9	76.9	55.076	15.2021	62
Cole-Liau	7.7	19.1	12.513	2.9712	62
Gunn-Fog	8.3	22.1	13.187	3.2427	62
SMOG	8.6	18.1	11.876	2.1786	62

TABLE 3
MEAN BY COUNTRY

Country	N	FKGrade	FKRead	ColeLiau	GunnFox	SMOG
Canada	32	10.069	53.472	12.434	13.566	12.103
US	27	9.489	56.330	12.856	12.881	11.696
UK	3	8.700	60.900	10.267	11.900	11.067
Total	62	9.750	55.076	12.513	13.187	11.876

TABLE 4
ANALYSIS OF VARIANCE FOR INDEXES BY PAGE TYPE

Index	Source of Variation	Df	Sum of Squares	Mean Square	F	p
FK Grade	Between Groups	3	114.522	38.174	5.554	.002
	Within Groups	58	398.633	6.873		
	Total	61	513.155			
FK Read Ease	Between Groups	3	3138.501	1046.167	5.537	.002
	Within Groups	58	10958.873	188.946		
	Total	61	14097.374			
ColeLiau	Between Groups	3	91.940	30.647	3.980	.012
	Within Groups	58	446.570	7.699		
	Total	61	538.510			
GunnFog	Between Groups	3	122.660	40.887	4.571	.006
	Within Groups	58	518.770	8.944		
	Total	61	641.430			
SMOG	Between Groups	3	58.975	19.658	4.945	.004
	Within Groups	58	230.559	3.975		
	Total	61	289.534			

TABLE 5
MEAN READABILITY SCORES NY PAGE TYPE

Page Type	N	FK Grade	FKRead	ColeLiau	GunnFog	SMOG
Drug	15	9.760	54.020	12.273	13.273	11.840
	sd	2.7555	15.2478	2.8986	3.1997	1.9650
Home	16	9.531	54.425	13.338	12.613	11.500
	sd	1.7876	10.5690	2.7344	1.8945	1.2992
Patient	16	8.006	65.594	10.675	11.575	10.794
	sd	2.4294	10.8082	2.4267	2.6625	1.7831
Privacy	15	11.833	45.607	13.833	15.433	13.467
	sd	3.3367	17.4768	3.0291	3.9213	2.7189
Total	62	9.750	55.076	12.513	13.187	11.876
	sd	2.9004	15.2021	2.9712	3.2427	2.1786