

# Assessing the Readability of App Descriptions and Investigating its Role in the Choice of mHealth Apps: Retrospective and Prospective Analyses

Wu-Chen Su, MS<sup>1</sup>, Khyati Y. Mehta, MS<sup>1</sup>, Kirandeep Gill, MS<sup>1</sup>, Peng Yeh, MS<sup>3</sup>,  
Ming-Yuan Chih, PhD<sup>4</sup>, Danny T.Y. Wu, PhD, MS<sup>1,2</sup>

<sup>1</sup>Department of Biomedical Informatics, University of Cincinnati; <sup>2</sup>Department of Pediatrics, University of Cincinnati, Cincinnati, OH; <sup>3</sup>Department of Statistics, University of Kentucky; <sup>4</sup>College of Health Sciences, University of Kentucky, Lexington, KY

## Abstract

*People with low health literacy are more likely to use mobile apps for health information. The choice of mHealth apps can affect health behaviors and outcomes. However, app descriptions may not be very readable to the target users, which can negatively impact app adoption and utilization. In this study, we assessed the readability of mHealth app descriptions and explored the relationship between description readability and other app metadata, as well as description writing styles. The results showed that app descriptions were at eleventh- to fifteenth-grade level, with only 6% of them meeting the readability recommendation (third- to seventh-grade level). The description readability played a vital role in predicting app installs when an app had no reviews. The content analysis showed copy-paste behaviors and identified two potential causes for low readability. More work is needed to improve the readability of app descriptions and optimize mHealth app adoption and utilization.*

## Introduction

Health literacy is a critical factor affecting health behaviors and decisions and further impact health outcomes<sup>1</sup>. The readability level of patient education materials (PEMs), however, may not match the health literacy of the general public<sup>2</sup>. Additionally, studies have reported that 35% of adults in the United States (US) had basic or below basic health literacy<sup>3</sup> and 47% of European Union adult residents had limited health literacy<sup>4</sup>. To meet the needs of average Americans who read at the eighth-grade level<sup>5</sup>, the American Medical Association (AMA) and the National Library of Medicine at the National Institutes of Health (NIH/NLM) have suggested that PEMs be composed on a third to eighth-grade level, or even fifth to sixth grade level<sup>6,7</sup>. PEMs failing to meet this required reading level (RRL) may not effectively delivery health information to their readers.

With the advances of information technology, PEMs have been published online and widely accessed by lay persons. Overall, 63% of US users and 71% of European users search health information on the Internet for a variety of purposes<sup>8-10</sup>. People with low self-reported health literacy are more likely to use mobile apps and to get health information from social networking sites<sup>11</sup>. However, this population may have limited ability to read app descriptions and understand app purposes, which can negatively impact their app choices. Meanwhile, the large amount of healthcare mobile apps (“mHealth apps” hereafter) available on the markets imposes challenges to the choice of suitable apps. These mHealth apps may target multiple user groups, including students, patients, healthcare professionals, and policy-makers<sup>12,13</sup>, which adds to the challenge of app choices. Recent studies have explored the factors behind app choices and found that users from the US are more likely to download medical apps, and that price, app features, descriptions, reviews, and rating stars were the most important factors affecting app choices<sup>14,15</sup>. These factors can also be confounded with each other. For example, paid apps (price) are expected to provide better quality than free apps in most regards, and thus a less readable description may diminish the perceived value of paid apps<sup>16</sup>. While many studies have enhanced the understanding of these decision factors and their relationships, no study so far focuses on improving the readability of app descriptions to help app developers and vendors design a better app and improve app utilization.

In this study, we aimed to assess the readability of app descriptions and explore the role of readability in app choices. Readability assessment has been conducted on online health information<sup>17,18</sup> as well as other types of health information such as electronic health records. For example, a recent retrospective study revealed that free-text directions on electronic prescriptions (information about medications) can be less readable to the patients<sup>19</sup>. This study found that 51.4% of randomly sampled directions from 966 patients have at least one quality issue, and that the pharmacy staff had to transcribe these directions to make them more readable. Similar to the issue of understanding the directions of medication, many mHealth apps may confuse their potential users if the readability of app descriptions is low.

Here we had three research questions. Firstly, what were the RRL of app descriptions and their relationships with other app metadata in the major app markets, namely the Apple App Store and Google Play Store? Also, did the app readability meet the recommended readability guideline by AMA and NIH/NLM? We hypothesized that the app readability would not meet the recommended readability level similar to many online health information. Secondly, did the app readability play any role in app choices? The literature showed multiple factors, such as number of reviews and ratings, can play a significant role in app choices. We hypothesized that users were more likely to download and/or install an app if the RRL of app description is low (i.e., a lower readability score can predict a higher number of app installs). Finally, what would be the writing styles and mechanics contributing to the higher RRL of app descriptions? What were possible reasons for app developers to write a less readable description? Had any app developers taken any actions to improve the description readability? By answering these research questions, we aimed to bridge the knowledge gap, help increase the readability of app descriptions, and further assist the lay public to choose proper mHealth apps to improve their health.

## Methods

### Data Collection

In order to conduct large-scale app analyses, a request was sent to the authors who created a mHealth app repository<sup>13</sup>. This mHealth app repository contained rich metadata of ‘Medical’ (ME) and ‘Health and Fitness’ (HF) apps on Apple App Store and Google Play Store, leading to four subsets of data based on the two app categories and the two vendor markets. Table 1 shows the metadata collected in each of the markets. This repository was released in four time periods, namely, the second and fourth quarters in 2015 and 2016 (Q2/2015, Q4/2015, Q2/2016 and Q4/2016). In the present study, only apps with English descriptions and published on the US market were selected. In order to identify apps with English-only descriptions in the repository, a Python library called “langdetect” was used<sup>20</sup>. The metadata of the selected mHealth apps were processed and stored in our MySQL database.

**Table 1.** Extracted metadata of app

Vendor Market	Metadata List
Apple App Store	unique app id, developer id, average of user ratings, content rating, app description, price, category, number of user ratings.
Google Play Store	unique app id, average of user ratings, category, content rating, app description, developer id, number of installs, price, number of user ratings, video URL, number of screenshots, age of app (the difference between app release date, and data recorded date), the average ratings of top 4 most helpful reviews, the average ratings of top 10 most helpful reviews, the average ratings of top 10 recent reviews.

### Language Surface Metrics

The app description was characterized using surface metrics including average document length (ADL) and vocabulary coverage (VC) as listed in Table 2. A general English dictionary, GNU Aspell<sup>21</sup>, was used to compute the VC of app descriptions and has been used in previous studies<sup>18,22</sup>.

### Readability Measures

Two readability measures were utilized to assess the descriptions of healthcare apps, including Flesch-Kincaid Grade Level (FKGL)<sup>23</sup> and Gunning-Fog Index (GFI)<sup>24</sup>, which were chosen in the present study due to their simple calculation and easy interpretation of the scores. These readability measures are considered “classic” measures due to their general design. They have been widely used since the 1970s and recently been used to assess online PEMs. Microsoft Word has implemented FKGL to allow users to assess and report the readability level of a document<sup>25</sup>. To simplify the interpretation of RRL, an average readability score (ARS) of FKGL and GFI was used since both measures are strongly correlated<sup>2,26</sup>. Table 2 lists the definition of the readability measures.

**Table 2.** Definition of surface metrics and readability measures

Surface Metrics and Readability Measures	Definition
Average Document Length by Sentence (ADL-SENT)	Average number of sentences per document
Average Document Length by Lexicon (ADL-LEX)	Average number of lexicons per document
Vocabulary Coverage (VC)	Number of words covered by a dictionary normalized by the vocabulary size (percentage)
Flesch-Kincaid Grade Level (FKGL)	$0.39 * (\text{number of lexicons} / \text{number of sentences}) + 11.8$ (number of syllables / number of lexcons) – 15.59; a raw score can be rounded to the nearest integer below its current value to form a grade level.
Gunning-Fog Index (GFI)	$0.40 * [(\text{number of lexicons} / \text{number of sentences}) + 100 * (\text{number of difficult words} / \text{total lexicons})]$ ; a word is considered difficulty if it has 3 or more syllables.
Average Readability Score (ARS)	The average of the raw scores of FKGL and GFI. Similarly, an ARS score can be rounded to the nearest integer below its current value to form a grade level.

### Data Analysis

The data analysis included three components; each corresponding one of the three research questions. In the first analysis, a retrospective analysis was conducted to describe the mHealth apps statistically and compare the RRL among the groups using the surface metrics and the readability measures listed in Table 2. Specifically, the distribution of mHealth app groups by category, vendor market, and over four time periods were summarized. The ADL-SENT, ADL-LEX, VC were applied to characterize the app descriptions. Non-parametric (Mann-Whitney U) tests were used with the significance level of 0.05 to examine if there was any significant difference among the group medians<sup>27</sup>. We are particularly interested in the relationship between the RRL of app descriptions and app content rating, price, as well as the recommended readability level. Of note, we chose third- to seventh-grade level as the recommended threshold due to the fact that half of the US people cannot read eighth-grade level content<sup>5</sup>.

Our second analysis examined the relationship between the app description readability and app installs. The analysis focused on the free apps on the Google Play Store since this was the major group of apps with information about number of installs. Since the mHealth app repository did not contain detailed information about app downloads and installs, a prospective app analysis was conducted to collect the app data outside of the repository. Specifically, a total of 240 free Google Play apps listed in the mHealth app repository were crawled for three months (mid-April to mid-July of 2019) using a Python library called ‘play-scraper’<sup>28</sup>. The actual number of apps included in this analysis was determined by the alive apps, the capacity of the Python library, and the study period. The included apps, regardless of their categories (i.e., ME vs HF), were further separated into two groups based on their number of reviews. The apps without any review were put into one group while the apps with at least one review were put into the other group. Linear regression in the R statistical package was used to model the data in each group, with the number of installs as the dependent variable and key metadata as the independent variable (predictors)<sup>29,30</sup>. The selection of metadata (independent variables) were based on the literature, including but not limited to: number of reviews, number of review rating, number of screenshots, and readability scores<sup>15</sup>. The coefficient of determination (R-squared value) of each mode was reported to show the proportion of variance in the dependent variable explained by the independent variables.

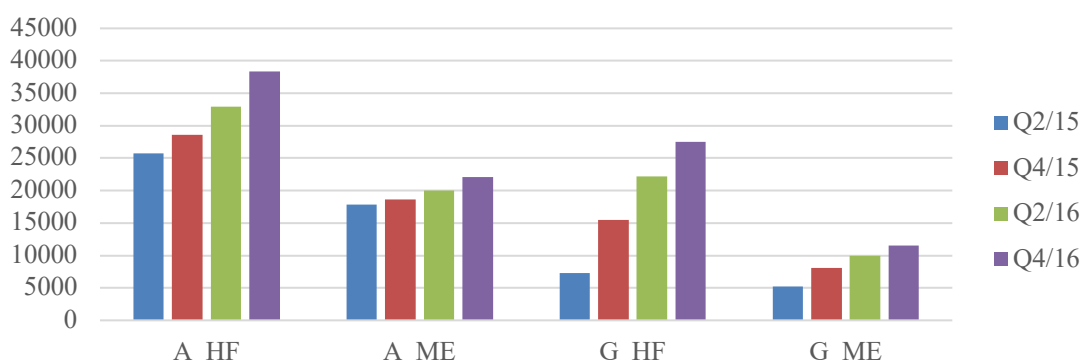
Lastly, a content analysis was conducted to understand the description writing styles. Of note, analyzing the concepts of the app descriptions was out of the scope of the present study. The first step in this analysis was to summarize the number of apps that each developer created in each market, assuming that a larger portion of apps were developed by a small group of app developers who created multiple apps. The second step followed the first and investigated “copy and paste” behaviors using a python library called ‘FuzzyWuzzy’<sup>31</sup>. A percentage was calculated by averaging the pairwise text similarity of the app descriptions of each app developer. The rationale was that copy-paste behaviors can exist if a developer creates multiple apps and uses a template to write the descriptions or just reuse existing descriptions. In this case, the readability of app descriptions would not change dramatically over time. The last step of the content analysis was to randomly select 50 apps that had a description with readability

at the eighth-grade level and above in each of the market, totaling 100 apps. These 100 app descriptions were manually reviewed by the research team to form suggestions.

## Results

### Descriptive Statistics and Group Comparisons

Figure 1 shows the numbers of mHealth apps based on their attributes in vendor markets (Apple App Store and Google Play Store), category (ME and HF), and published time period (Q2/15, Q4/15, Q2/16, and Q4/16). Each group for comparison was denoted using the abbreviation of the attributes. For example, Health and Fitness apps on the Apple App Store were denoted as “A\_HF”. Similarly, Medical apps on Google Play Store were denoted as “G\_ME”. Here bar charts instead of line charts were used to show the comparisons because of our focus on vendor market and app category. Overall, Apple App Store has more apps than Google Play Store. The average quarterly increase rate of the numbers of apps was more than 10% between 2015 and 2016 for apps in the Apple App store. For Google Play store, there was a similar tendency, although the average increase rate is more than 45%. In both markets, HF apps contributed more to the increase than ME apps; G\_HF apps outgrew A\_ME apps in Q4 of 2016.



**Figure 1.** Distribution of the selected mHealth Apps.

Table 3 shows the summary of surface metrics. Apps on the Google Play Store had more sentences and lexicons in their descriptions than those on the Apple App Store. However, there is no statistical difference of the average document length (ADL-LEX and ADL-SENT) among the app categories between the markets. The app descriptions were around 200 words in 10 sentences. The VC is high (>90%) in both markets, meaning the app descriptions were written in common vocabularies most of time rather than rare terms.

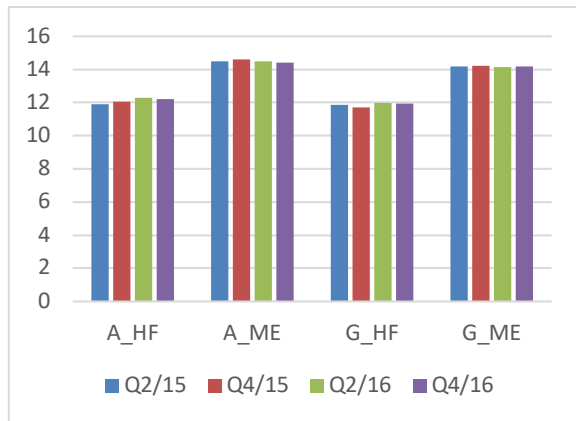
**Table 3.** Summary of surface metrics of app descriptions

<b>Apple App Store</b>	<b>Health and Fitness (A_HF)</b>		<b>Medical (A_ME)</b>	
Study Period	LEX <sup>a</sup> (std) <sup>b</sup>	SENT <sup>a</sup> (std)	LEX (std)	SENT (std)
Q2/15	188.56 (150.96)	9.88 (8.24)	180.29 (142.02)	8.73 (7.57)
Q4/15	185.28 (149.04)	9.65 (8.06)	179.03 (141.19)	8.67 (7.5)
Q2/16	179.14 (147.87)	9.27 (7.87)	179.36 (142.55)	8.68 (7.55)
Q4/16	173.05 (145.79)	8.99 (7.79)	180.05 (144.52)	8.68 (7.55)
<b>Google Play Store</b>	<b>Health and Fitness (G_HF)</b>		<b>Medical (G_ME)</b>	
Study Period	LEX <sup>a</sup> (std)	SENT <sup>a</sup> (std)	LEX <sup>a</sup> (std)	SENT <sup>a</sup> (std)
Q2/15	224.81 (166.67)	11.83 (9.81)	221.41 (166.79)	11.08 (9.34)
Q4/15	190.53 (155.10)	10.18 (8.78)	206.72 (159.76)	10.26 (8.8)
Q2/16	185.32 (158.80)	9.72 (8.6)	223.98 (222.45)	11.62 (13.78)
Q4/16	184.82 (164.57)	9.67 (8.76)	218.21 (215.21)	11.21 (13.17)

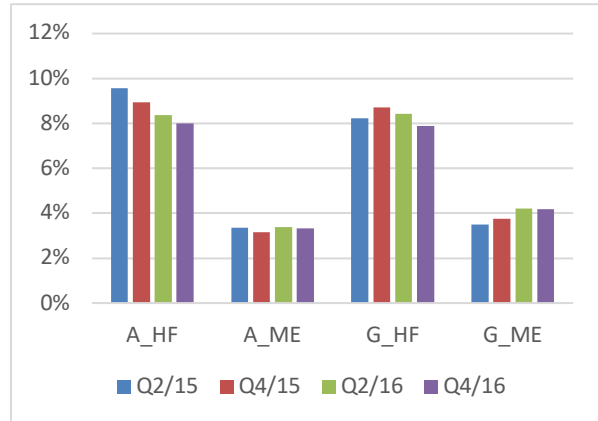
<sup>a</sup>LEX: average number of lexicons per document. SENT: average number of sentences per document.

<sup>b</sup>std: standard deviation.

Figure 2 shows that the ARS for these four groups of apps was between eleventh- and fifteenth-grade level. They were far beyond the recommended readability levels from AMA and NIH/NLM. The RRLs of app descriptions on the Apple App Store are statistically lower than these on the Google Play Store combining the categories and the time periods (Mann-Whitney U test,  $p < .01$ ). Further, the descriptions of ME apps had higher RRL than HF apps in both markets (Mann-Whitney U test,  $p < .01$ ). Additionally, only around 6% of apps on average on both markets met the readability recommendation, where more HF apps met the readability recommendation than ME apps (Figure 3).



**Figure 2.** Average readability scores (ARS) for different groups and time periods.



**Figure 3.** Percentage of Apps meeting recommended readability guideline (third- to seventh-grade level).

Next, the relationships between the RRL and the two metadata (i.e., price and content rating) were examined to understand the potential impact of description readability on app users. The analysis indicated that the paid apps did not have a lower RRL than the free apps. Meanwhile, suitable content may not always be readable to the targeted app users, which may have a negative impact on app choices and utilization. Specifically, apps with ‘4+’ content rating on the Apple App Store and ‘Everyone’ content rating in the Google Play Store were selected for analysis. As shown in Table 4, these apps categorized for any age group required their users with at least eleventh-grade level literacy to read the descriptions.

**Table 4.** Group mean readability scores and content ratings

	A_ME	A_HF	G_ME	G_HF
Age groups	4+	4+	Everyone	Everyone
Q2/15	14.07 (6.14)*	11.62 (6.52)	14.05 (6.86)	11.9 (6.4)
Q4/15	14.14 (6.32)	11.55 (6.4)	14.57 (6.74)	11.79 (5.49)
Q2/16	14.05 (6.36)	11.48 (6.18)	14.33 (6.68)	12.07 (7.78)
Q4/16	14.02 (6.34)	11.33 (5.94)	14.31 (7.11)	12.04 (7.02)

\* Medical apps on the Apple App Store with ‘4+’ content rating in the second quarter of 2015 had a mean readability score (ARS) of 14.06 with a standard deviation of 6.14.

### Modeling Readability and App Installs on Google Play Store

When we tracked apps prospectively, the total number of apps dropped from 240 (collected in the mHealth app repository) in mid-April to 218 (alive in the app stores when crawling) in mid-July. The modeling results (Table 5) show that the free Google Play apps with no review had three significant independent variables ( $p < .01$ ) to predict the number of installs. These significant independent variables (predictors) include the number of user ratings ( $\beta = 60.08$ ), the number of screenshots ( $\beta = -6.93$ ), and the average readability score (ARS,  $\beta = -4.62$ ), with an intercept of 140.64. However, the effect size of this model was weak ( $R\text{-squared}=0.203$ ). On the other hand, the free Google play apps with at least one review had six predictors significant ( $p < .01$ ), including the number of user ratings ( $\beta =$

39.52), the average user rating ( $\beta = -1308.85$ ), the number of screenshots ( $\beta = -142.35$ ), the average rating of top 4 most helpful reviews showing on the landing page ( $\beta = 1517.23$ ), the average rating of top 10 most useful reviews ( $\beta = -9544.52$ ), and the average rating of top 10 most recent reviews ( $\beta = 8511.05$ ), with an intercept of 10114.91. The effect size of this model was moderate ( $R\text{-squared}=0.5332$ ). The number of user ratings and screenshots were significant in both groups with and without a review, which speaks to the universal importance of these variables. When there was no review at all, the average readability score of the app descriptions seemed to play a significant role in predicting the number of installs of an app. The negative coefficient supported our hypothesis: the lower a RRL (ARS) is, the more the number of app installs would be. On the other hand, when there was at least one review, the predictors were dominated by the review ratings.

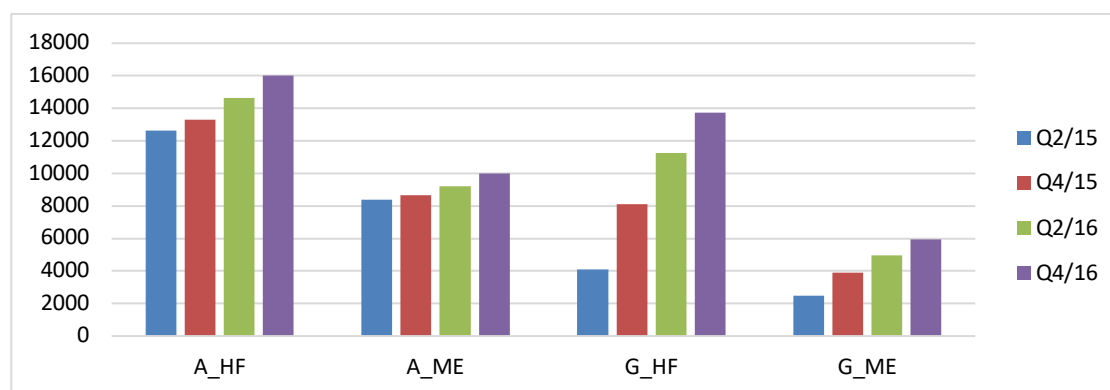
**Table 5.** Factors contributing to the number of app installs on Google Play Store.

Coefficients	App with no review		App with at least one review	
	Estimate	p-value	Estimate	p-value
(Intercept)	140.64	4.67e-6 ***	10114.91	7.61e-16 ***
Number of user ratings	60.08	< 2e-16 ***	39.52	< 2e-16 ***
Average user rating	7.14	0.0933	-1308.85	0.0014 **
Has video or not	-16.21	0.7066	-1012.84	0.1215
Number of screenshots	-6.93	4.06e05 ***	-142.35	4.28e-16 ***
App age (days on market)	0.22	0.7019	5.39	0.7268
Average readability score (ARS)	-4.62	0.0021 **	-5.33	0.5807
Number of reviews	NA	NA	-6.94	0.0635
Average rating of all reviews	NA	NA	-802.57	0.2593
Average rating of top 4 most helpful reviews	NA	NA	1517.23	0.0099 **
Average rating of top 10 most helpful reviews	NA	NA	-9544.52	0.0033 **
Average rating for top 10 most recent reviews	NA	NA	8511.05	0.0076 **

\*\* p<0.01 \*\*\* p < 0.001

### Content Analysis

Figure 4 shows the numbers of active app developers in each group over the four time periods. There were more app developers in the Apple market than in the Google market. Similar to the trend of app increase rate shown in Figure 1, the app developers grew steady in the Apple market while the developers grew two to three times faster in the Google market. In both markets, an app developer created two apps on average. Moreover, more than half of the apps were created by a small group of developers who had two or more apps. For example, for the Health and Fitness apps on the Apple App Store, 53% of the apps were created by 11% of the app developers with two or more apps, while 47% of the apps were created by 89% of the app developers with only one app.



**Figure 4.** Distribution of active app developers

The content analysis continued to investigate potential copy and paste behaviors since many apps were created by a small group of developers. The copy-paste behaviors were confirmed by the high (more than 50%) text similarity of the app descriptions for developers who created at least two apps in both markets. In addition, the analysis of app descriptions changes from time to time indicated that ME apps had a large variance of changes in both markets. Furthermore, only 13.86% of apps had a description change that led to a lower RRL. In other words, most description changes resulted in the same or even a higher RRL.

After reviewing the selected 100 apps which had a readability level of eighth grade or above (higher than the recommended level), two common factors contributing to the low readability of app descriptions were identified. 1) Lack of Writing Standard. Unlike online health education materials, there is no standard way to write app descriptions on either market. Therefore, the writing styles had high variations and app descriptions may contain repetitive content and unnecessary information. For example, an app description may contain a ‘To-Do list’ (Table 6 left, App 1) with different subjects or the change history of the app. In these cases, many difficult words (words with many syllables) were included, thus increasing the complexity of reading. App descriptions can be revised in a more succinct and organized way to improve the readability, which may favor the readability measure in improving the sentence length and the syllable count. 2) Search Optimization. Similar to many existing web search engines, the app stores have their own mechanism to retrieve relevant apps for users given a query. The search mechanism may consider app title, description, user ratings, and other metadata, and help users uncover the apps they need. Therefore, there are some App Store Optimization approaches<sup>32</sup> to boost the search results in the app stores. Obviously, some app developers would employ these strategies and incorporated many popular keywords (Table 6 right, App 2) in the descriptions, and expected that these approaches will bring more awareness and traffic to their apps. However, these approaches may also make the app descriptions less readable, or negatively impact the readability of app descriptions.

**Table 6.** App description examples.

Partial description of App 1*	Partial description of App 2*
48) How to Do Bojutsu Kamae Basic Postures	Mexican Fideo Soup
49) How to Do the Bojutsu Striking Drill	Creamy Asparagus Soup
50) How to Do the Upper Level Bojutsu Block	Chicken and Saffron Rice Soup
51) How to Do Goho from Bojutsu Training	Chicken, Wild Rice, and Mushroom Soup
52) Bo Furi Gata from Bojutsu Training	Mustard Greens Soup
53) How to Do the Muto Dori Technique	Soup- Ojai Valley Inn Tortilla Soup
54) How to Do the Kenjutsu Technique	Great collection of cheesecake recipe app,banana
55) How to Do Kusari-Fundo Techniques	bread,chicken breast recipes,pumpkin pie recipe,dessert
56) How to Use Hanbo Techniques	recipes,soup recipes,dinner recipes,easy chicken recipes,sweet
57) How to Use Metsubushi Techniques	potato pie,apple crisp recipe,apple butter recipe,banana
58) How to Throw a Shuriken	cake,italian recipes,peach pie,cookie recipes app,waffle recipe
59) How to Throw the Bo Shuriken	app,dessert recipes app,cupcake recipes app,Baking recipes
60) How to Use the Naginata	app,Chinese Recipes app,Chocolate recipes app,Delicious
61) How to Use the Yari	recipes app,chili recipe,Dessert Recipes app,Dinner recipe
62) How to Use the Kusarigama	app,Fish recipes app,Grill Recipes app,Indian recipes...
63) How to Use a Rope	app,Italian Recipes app,Kids recipes app...

\* The app descriptions were extracted from the mHealth app repository, not the most current information. Both apps were on the Google Play Store.

## Discussion

In this study, we characterized the mHealth apps in both Apple App Store and Google Play Store and summarized the readability of app descriptions. Although there were more apps on Apple App Stores, the Google Play Store was fast growing, especially in the Health and Fitness category. The overall RRL of app descriptions was between eleventh- and fifteenth-grade level, which was far beyond the recommended third- to seventh-grade level. This

finding was consistent with a recent study indicating that the privacy policies of mHealth apps were not very readable (sixteenth-grade level)<sup>33</sup>. Moreover, only around 6% apps that meet the recommended grade level highlights the need for improvement.

We also explored price and content rating and their relationships with description readability. Interestingly, paid apps did not provide better readability than free apps, and the content ratings of apps did not always match the corresponding grade levels. In other words, apps with suitable content may not be as readable to the target users as they are supposed to be and unfortunately become unsuitable. There could be other predictors. For example, expert involvement may increase the app installs<sup>34</sup>. Another key factor in the choice of apps may be the ability of a mHealth app to manage patient-generated health data given the current advances and wide adoption of smart watches and wearable devices. Developers and vendors should consider these factors carefully in app development and advertisement.

The content analysis of the less readable apps revealed that a small group of app developers created two or more apps and contributed to a significant portion of the apps. These app developers may copy and paste their app descriptions for efficiency and they have no incentive to improve the readability of app descriptions over time. The high RRL of app descriptions may be contributed by the lack of writing standard and the desire for search optimization, which is a similar pattern identified in a previous study analyzing cardiological apps in the Apple App Store in German<sup>35</sup>. More research is needed to demonstrate the best practice of writing app descriptions, balancing between readability, retrievability, and efficiency.

This study had several strengths. Based on our knowledge, this is the first study conducting both retrospective and prospective analyses with a focus on the readability of mHealth app descriptions in both Apple and Google market. Moreover, this study adopted multiple methods. Not only did it conduct statistical summary and group comparisons, but also it developed a model to predict user behaviors as well as reviewed app descriptions to understand potential causes for low readability. This study also had a few limitations. First, we only analyzed ‘Health and Fitness’ and ‘Medical’ categories based on the data obtained from the previous study<sup>9</sup>, acknowledging that there could be other related (or new) categories that we missed. Second, different app markets have different developer groups, description writing guidelines, and management policies, which affect our analysis results. Moreover, since we only focused on Apple App Store and Google Play Store, and the results may not be generalizable to other app markets (e.g. Amazon Kindle Store). Third, the study did not consider that fact that users install apps but delete them afterwards due to the lack of usefulness. Also, the study did not conduct a thorough content analysis to understand how the readability of app descriptions may impact the end users. Next, the modeling of app installs and their metadata only showed weak to moderate effect size and can be improved. In addition, the classic readability measures used in the present study have known limitations. For example, a word with multiple syllables may not be difficult and some short words can be very difficult to read and understand. Last but not least, the readability of app descriptions was not validated by app users. Future research can conduct user studies to validate the readability scores as well as understand user’s perspectives on less readable app descriptions.

## Conclusion

We assessed the readability of mHealth app descriptions and demonstrated its potential role in mHealth app choices. This study serves as the first attempt to address this issue and enabled future studies to develop solutions to improve the readability of app description and further help lay persons select suitable mHealth apps for their health needs.

## Acknowledgements

We thank Dr. Welong Xu and Dr. Yin Liu for their willingness to share the mHealth app repository. We also thank Mr. Karthikeyan Meganathan at the University of Cincinnati College of Medicine for his assistance on the statistical analysis and Ms. Anunita Nattam for her efforts in proofreading the manuscript.

## References

1. Ferguson LA, Pawlak R. Health Literacy: The Road to Improved Health Outcomes. *J Nurse Pract.* 2011;7(2):123-129. doi:10.1016/j.nurpra.2010.11.020
2. Hansberry DR, Agarwal N, Baker SR. Health literacy and online educational resources: an opportunity to educate patients. *AJR Am J Roentgenol.* 2015;204(1):111-116. doi:10.2214/AJR.14.13086
3. America’s Health Literacy: Why We Need Accessible Health Information. Published online 2008. <https://www.ahrq.gov/sites/default/files/wysiwyg/health-literacy/dhhs-2008-issue-brief.pdf>



4. Sørensen K, Pelikan JM, Röthlin F, et al. Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). *Eur J Public Health*. 2015;25(6):1053-1058. doi:10.1093/eurpub/ckv043
5. Strauss V. Hiding in plain sight: The adult literacy crisis. *The Washington Post*. <https://www.washingtonpost.com/news/answer-sheet/wp/2016/11/01/hiding-in-plain-sight-the-adult-literacy-crisis/?noredirect=on>. Published November 1, 2016. Accessed March 7, 2021.
6. Weiss BD. Health Literacy: A Manual for Clinicians. Published online 2003. Accessed March 7, 2021. <http://lib.ncfh.org/pdfs/6617.pdf>
7. Question: How can I locate materials on MedlinePlus that are easy to read and what is their reading level? MedlinePlus: Trusted Health Information for you. Accessed March 7, 2021. <https://medlineplus.gov/faq/easytoread.html>
8. Hesse BW, Nelson DE, Kreps GL, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. *Arch Intern Med*. 2005;165(22):2618-2624. doi:10.1001/archinte.165.22.2618
9. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, et al. European citizens' use of E-health services: a study of seven countries. *BMC Public Health*. 2007;7:53. doi:10.1186/1471-2458-7-53
10. Fox S. The Social Life of Health Information 2011. Published online May 12, 2011. Accessed March 7, 2021. [https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2011/PIP\\_Social\\_Life\\_of\\_Health\\_Info.pdf](https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf)
11. Manganello J, Gerstner G, Pergolino K, Graham Y, Falisi A, Strogatz D. The Relationship of Health Literacy With Use of Digital Technology for Health Information: Implications for Public Health Practice. *J Public Health Manag Pract JPHMP*. 2017;23(4):380-387. doi:10.1097/PHH.0000000000000366
12. Wu-Chen Su. A Preliminary Survey of Knowledge Discovery on Smartphone Applications (apps): Principles, Techniques and Research Directions for E-health. Published online 2014. doi:10.13140/RG.2.1.1941.6565
13. Xu W, Liu Y. mHealthApps: A Repository and Database of Mobile Health Apps. *JMIR MHealth UHealth*. 2015;3(1):e28. doi:10.2196/mhealth.4026
14. Zhang C. Why do We Choose This App? A Comparison of Mobile Application Adoption Between Chinese and US College Students. Published online 2018. Accessed March 7, 2021. [https://scholarworks.bgsu.edu/media\\_comm\\_diss/56](https://scholarworks.bgsu.edu/media_comm_diss/56)
15. Lim SL, Bentley PJ, Kanakam N, Ishikawa F, Honiden S. Investigating Country Differences in Mobile App User Behavior and Challenges for Software Engineering. *IEEE Trans Softw Eng*. 2015;41(1):40-64. doi:10.1109/TSE.2014.2360674
16. Hsu C-L, Lin JC-C. What drives purchase intention for paid mobile apps? – An expectation confirmation model with perceived value. *Electron Commer Res Appl*. 2015;14(1):46-57. doi:10.1016/j.elerap.2014.11.003
17. McInnes N, Haglund BJA. Readability of online health information: implications for health literacy. *Inform Health Soc Care*. 2011;36(4):173-189. doi:10.3109/17538157.2010.542529
18. Wu DT, Hanauer DA, Mei Q, et al. Assessing the readability of clinicaltrials.gov. *J Am Med Inform Assoc JAMIA*. Published online August 11, 2015. doi:10.1093/jamia/ocv062
19. Zheng Y, Jiang Y, Dorsch MP, Ding Y, Vydiswaran VGV, Lester CA. Work effort, readability and quality of pharmacy transcription of patient directions from electronic prescriptions: a retrospective observational cohort analysis. *BMJ Qual Saf*. Published online May 25, 2020. doi:10.1136/bmjqs-2019-010405
20. Danilk M. *Langdetect: Language Detection Library Ported from Google's Language-Detection*. Accessed March 7, 2021. <https://pypi.org/project/langdetect/>
21. Atkinson K. *GNU Aspell*.; 2018. Accessed March 7, 2021. <http://aspell.net/>
22. Zheng K, Mei Q, Yang L, Manion FJ, Balis UJ, Hanauer DA. Voice-dictated versus typed-in clinician notes: linguistic properties and the potential implications on natural language processing. *AMIA Annu Symp Proc AMIA Symp AMIA Symp*. 2011;2011:1630-1638.
23. Kincaid J, Fishburne RJr, Rogers R, Chissom B. Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel. Published online February 1975.
24. Gunning R. *The Technique of Clear Writing*. McGraw-Hill; 1968.
25. Microsoft Corporation. Get your document's readability and level statistics. Accessed March 7, 2021. <https://support.microsoft.com/en-us/topic/get-your-document-s-readability-and-level-statistics-85b4969e-e80a-4777-8dd3-f7fc3e8b3fd2>

26. Cheng C, Dunn M. Health literacy and the Internet: a study on the readability of Australian online health information. *Aust N Z J Public Health*. 2015;39(4):309-314. doi:10.1111/1753-6405.12341
27. McKnight PE, Najab J. Mann-Whitney U Test. In: Weiner IB, Craighead WE, eds. *The Corsini Encyclopedia of Psychology*. John Wiley & Sons, Inc.; 2010:corpsy0524. doi:10.1002/9780470479216.corpsy0524
28. Liu D. *Play-Scraper: Scrapes and Parses Application Data from the Google Play Store*. Accessed March 7, 2021. <https://pypi.org/project/play-scraper/>
29. Prabhakaran S. Linear Regress in R. Accessed March 7, 2021. <http://r-statistics.co/Linear-Regression.html>
30. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2020. <https://www.R-project.org/>
31. Cohen A. *FuzzyWuzzy: Calculate the Differences between Sequences in a Simple-to-Use Package.*; 2020. Accessed March 7, 2021. <https://pypi.org/project/fuzzywuzzy/>
32. Patel N. What is App Store Optimization? (ASO). Accessed March 7, 2021. <https://neilpatel.com/blog/app-store-optimization/>
33. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. *J Am Med Inform Assoc JAMIA*. 2015;22(e1):e28-33. doi:10.1136/amiajnl-2013-002605
34. Pereira-Azevedo N, Osório L, Cavadas V, et al. Expert Involvement Predicts mHealth App Downloads: Multivariate Regression Analysis of Urology Apps. *JMIR MHealth UHealth*. 2016;4(3):e86. doi:10.2196/mhealth.5738
35. Albrecht U-V, Hasenfuß G, von Jan U. Description of Cardiological Apps From the German App Store: Semiautomated Retrospective App Store Analysis. *JMIR MHealth UHealth*. 2018;6(11):e11753. doi:10.2196/11753