



## Who Reads E-Books? An Analysis Of Key Factors In E-Book Reading In The U.S.

**Xin Gu**

Bentley University  
Waltham, MA 02452. U.S.A.

**Paul D. Berger**

Bentley University  
Waltham, MA 02452. U.S.A

### ABSTRACT

With the rapid increase in the number of e-book readers, researchers have become more and more interested in the factors that impact digital reading, because these factors are not only useful for marketing segmentation, but also important for optimizing marketing campaigns of an e-reader device company and the entire digital publishing industry. In this paper, we perform stepwise linear regressions and independent sample t-tests (for differences in means) to determine the key factors in e-book reading behavior, and analyze how these factors affect e-book reading.

**KEY WORDS:** e-books, digital reading, survey data, stepwise linear regression, independent sample t-test

### INTRODUCTION

The share of Americans who read e-books has been increasing in the last five years, from 17% in 2011 to 28% in 2016 (Perrin, 2016). With the rapid increase in the number of e-book readers, researchers have become more and more interested in the factors that impact digital reading, because these factors are not only useful for marketing segmentation, but also important for optimizing marketing campaigns of an e-reader device company and the entire digital publishing industry. In this paper, we perform stepwise linear regressions and independent sample t-tests for differences in means to determine the key factors in e-reading behavior, and analyzed how these factors affect it. The data used in this paper were collected by the Pew Research Center from March 7 to April 4, 2016 (Pew Research Center, 2016). The questionnaire was designed to collect data on the usage of public libraries, reading behavior and the usage of electronic devices such as smart phones, tablets and e-readers in the U.S. There are total of  $n = 1601$  respondents (observations) in the dataset.

### LITERATURE REVIEW

There has been a variety of research on the topic of digital reading and e-book readership. Perrin (2016) did a descriptive analysis and calculated the sample proportions of e-book readers with different education levels. The analysis of his results shows that college graduates are roughly four times as likely to read e-books compared with those who have not graduated from high school. Huang, Liang & Chiu (2013) considered gender differences in e-book reading. They focused on elementary school students and found that the attitudes toward e-books between boys and girls are not significantly different. Zhang & Ma (2011) studied the correlation between a user's education level and mobile reading behavior in China. They concentrated on a specific user group, college students, and concluded that those with a high

educational background are the main users of mobile reading, such as social media and e-books. Shahriza, Karim & Hasan (2007) compared digital reading behavior in Malaysia for different academic programs, such as IT-based programs versus arts-based programs. Their conclusion was that IT-based students rely more on electronic resources than arts-based students.

This paper concentrates on American residents at least 16 years old, and considers more factors than the papers cited above. In addition, the methodologies used in this paper are distinct, and the analysis more comprehensive, than the previous studies cited. Also, unlike the other cited studies, we focus exclusively on the reading of *e-books*, and do not lump that activity together with other digital-reading behavior.

### METHODOLOGY

First, we run several stepwise linear regressions on different clusters of questions in the questionnaire. This was to satisfy our initial objective, to discover the most influential cluster of questions on the questionnaire. The three clusters of questions we studied were:

*General Reading Behavior*

*Usage of Public Libraries and Their Websites and Apps*

*Demographics*

Our dependent variable, and the variables comprising each cluster above are defined as follows, along with the details of their measurement scales:

#### Variables

The dependent variable is:

**Q books2c:** read at least one e-book in the past 12 months

The relevant clusters, and variables within these clusters:

#### **Cluster 1 General reading behaviors:**

**Q books2a:** read at least one printed book in the past 12 months

**Q books2b:** used at least one audio book in the past 12 months

#### **Cluster 2 Usage of public library and its website or app:**

**Q lib.use-a:** visited public library

**Q lib.use-b:** visited public library website

**Q lib.use-c:** used public library app

#### **Cluster 3 Demographics:**

**Q sex:** sex

**Q age:** age

**Q educ2:** education levels

**Q emplnw:** occupations

**Q inc:** income levels

The coded scale of **Q books2c** (the dependent variable), **Q books2a** and **Q books2b** was:

1 Yes

2 No

8 Don't know

9 Refused to answer

For these three variables, “8 Don’t know” and “9 Refused” were treated as missing values (essentially, discarded), since, clearly, their numerical value was not consistent with an interval-scale interpretation. With only two categories thus remaining, the variable became a routine two-category dummy variable. And, we renamed and recoded Q books2c, Q books2a and Q books2b into “**Ebook**” “**Printbook**” and “**Audiobook**,” respectively, each with scale “1 = Yes” and “0 = No.”

The scale of the Q lib.use-a, Q lib.use-b and Q lib.use-c was

- 1 Yes, have done this at least once in past 12 months
- 2 Yes, have done this, but not in the past 12 months
- 3 No, have never done this
- 8 Don't know
- 9 Refused

For these three variables, “8 Don’t know” and “9 Refused” were similarly treated as missing values. And, since “3 No” represented never used a public library or its website or app, and both 1 and 2 represented “has used a public library or its website or app, we renamed and recoded

Q lib.use-a, Q lib.use-b and Q lib.use-c into “library”, “libweb” and “libapp,” respectively, with each scale “1 = Yes” and “0 = No.”

The scale of the Q sex was

- 1 Male
- 2 Female

For this variable, we recoded the scales to “1 = Male” and “0 = Female.”

The scale of the Q age was

- \_\_\_years [exact age 16-96]
- 97 97 or older
- 98 Don't know
- 99 Refused

For this variable, “98 Don’t know” and “99 Refused” were similarly treated as missing values. Other values were coded as recorded. No data values were in the 97 or older category. This assumes a possible linear relationship between the dependent variable and age. We will note in our limitations section that a superior choice might have been to code the variable as a set of dummy variables by age groups of some width. This would potentially allow for capturing a non-linear relationship between the dependent variable and age. This may also be true for the income variable - discussed a bit later.

The scale of the Q educ2: education level was

- 1 Less than high school (Grades 1-8 or no formal schooling)
- 2 High school incomplete (Grades 9-11 or Grade 12 with NO diploma)
- 3 High school graduates (Grade 12 with diploma or GED certificate)
- 4 Some college, no degree (includes some community college)
- 5 Two-year associate degree from a college or university
- 6 Four-year college or university degree/Bachelor’s degree (e.g., BS, BA, SB)
- 7 Some postgraduate or professional schooling, no postgraduate degree (e.g. some graduate school)

8 Postgraduate or professional degree, including master's, doctorate, medical or law degree (e.g., MA, MS, PhD, MD, JD)  
98 Don't know  
99 Refused

Again, "98 Don't know" and "99 Refused" were treated as missing values. The remainder of the scale (1 - 8) was left intact, and thus treated as interval scale and, in a sense, "linear."

The scale of the Q emplnw: occupation was

1 Employed full-time  
2 Employed part-time  
3 Retired  
4 Unemployed  
5 Have own business/self-employed  
6 Disabled  
7 Student  
8 Other  
98 Don't know  
99 Refused

Yet again, "98 Don't know" and "99 Refused" were treated as missing values. Since this variable has a nominal scale, in order to include different occupations in our analysis, we broke the variable into 5 different dummy variables (categories) - "Employed (including 1, 2 and 5 in the original scale)", "Retired", "Unemployed", "Student" and "Others (including 6 and 8 in the original scale)," each dummy variable with scale "1 = Yes" and "0 = No."

The scale of the Q inc: income level was

1 Less than \$10,000  
2 \$10,000 to under \$20,000  
3 \$20,000 to under \$30,000  
4 \$30,000 to under \$40,000  
5 \$40,000 to under \$50,000  
6 \$50,000 to under \$75,000  
7 \$75,000 to under \$100,000  
8 \$100,000 to under \$150,000  
9 \$150,000 or more  
98 Don't know  
99 Refused

As has been the case, "98 Don't know" and "99 Refused" were treated as missing values. The remainder of the scale stayed intact, and was thus treated as a routine interval scale. We do not believe that the results are materially affected by the fact that the widths of the intervals are not identical. In fact, it would not surprise us if the differences in width of the intervals lead to a scale that is more representative of linearity.

### **Analysis process**

After running a stepwise linear regression analysis for each cluster, we determined a "final model" of examining e-readership by performing another stepwise linear regression in which the eligible variables were those that survived the initial stepwise regression for each cluster. While, by definition of the stepwise regression technique, we "know" that there will be relatively little correlation among surviving variables within a cluster, the last step will

eliminate significant overlap/correlation between variables from different clusters, should that exist. Then, based on our results, we conducted various independent sample t-tests for differences in means, and analyzed both the t-test results and final model to see how these factors affect e-reading behavior.

### ANALYSIS AND DISCUSSION OF RESULTS

First, we ran a stepwise linear regression on Cluster 1 - General reading behavior.<sup>1</sup> From Table 1, we can see that both “Audiobook” and “Printbook” are significant variables with respect to our e-book reading dependent variable. From Model 2 (the last step), we can see that 8.1% of the variability in “Ebook” has been explained by the two independent variables.

**Table 1 Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.221 <sup>a</sup>	.049	.049	.475
2	.285 <sup>b</sup>	.081	.081	.467

a. Predictors: (Constant), Audiobook

b. Predictors: (Constant), Audiobook, Printbook

Then, we performed a stepwise linear regression on Cluster 2 - Usage of the public library and its website or app. From Table 2, we can see that “libweb” and “libapp” are the influential factors in e-book reading, while “library” did not enter at all. This indicates that once we consider “libweb” and “libapp,” “library” does not add significant predictive value about Y (e-book readership). It can also be seen that, while statistically significant, only 2.6% of the variability in “Ebook” has been explained by Model 2 in Table 2.

**Table 2 Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.129 <sup>a</sup>	.017	.016	.48278
2	.160 <sup>b</sup>	.026	.025	.48065

a. Predictors: (Constant), libweb

b. Predictors: (Constant), libweb, libapp

Lastly, we conducted a stepwise linear regression on Cluster 3 Demographics. From Table 3, we can see that Income, Age, EDUC, Student, Employed and Sex are included in the last step (Model 6), meaning that each of these variables adds incremental predictive value about Y. And, 5.9% of the variability in “Ebook” has been explained by Model 6 in Table 3.

---

<sup>1</sup> We are aware that we have a (0, 1) dependent variable and are, nevertheless, using linear regression, instead of the, perhaps more appropriate, logistic regression. We understand that several of our measures are, thus, not exactly, quantitatively accurate. However, we believe that the overall understanding of the message the data have for us is not altered, and that the coefficients arrived at with our linear regression approach may provide added clarity to many of the readers.

**Table 3 Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.152 <sup>a</sup>	.023	.023	.484
2	.199 <sup>b</sup>	.040	.039	.480
3	.228 <sup>c</sup>	.052	.051	.477
4	.236 <sup>d</sup>	.056	.054	.476
5	.240 <sup>e</sup>	.057	.056	.476
6	.242 <sup>f</sup>	.059	.057	.475

a. Predictors: (Constant), Income

b. Predictors: (Constant), Income, age

c. Predictors: (Constant), Income, age, EDUC

d. Predictors: (Constant), Income, age, EDUC, Student

e. Predictors: (Constant), Income, age, EDUC, Student, Employed

f. Predictors: (Constant), Income, age, EDUC, Student, Employed, sex

From the analysis above, we know that the "Cluster 1 General reading behavior" is the most predictive cluster for the digital reading behavior among all three clusters. We next performed one final stepwise linear regression, in which eligible variables were those which survived the three aforementioned clusters above.

Table 4a shows that of the 10 eligible variables, 9 of them "survived:" Printbook, Audiobook, EDUC, age, Income, libapp, Student, Employed, libweb are included in the final model. Only "sex" is not included, indicating that once we know the 9 variables in the final model, "sex" does not add any incremental predictive value about e-book readership, Y. Table 4b shows the ANOVA results and p-values of the models, and Table 4c illustrates the coefficients and p-values of the variables in the final least-squares equation.

**Table 4a Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.212 <sup>a</sup>	.045	.045	.479
2	.284 <sup>b</sup>	.081	.080	.470
.....				
8	.365 <sup>h</sup>	.133	.131	.457
9 final model	.367 <sup>i</sup>	.135	.132	.456

a. Predictors: (Constant), Printbook

b. Predictors: (Constant), Printbook, Audiobook

.....

h. Predictors: (Constant), Printbook, Audiobook, EDUC, age, Income, libapp, Student, Employed

i. Predictors: (Constant), Printbook, Audiobook, EDUC, age, Income, libapp, Student, Employed, libweb

**Table 4b ANOVA**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regressio	33.603	1	33.603	146.518	.000 <sup>b</sup>
	Residual	712.111	3105	.229		
	Total	745.714	3106			
.....						
9final model	Regressio	100.441	9	11.160	53.563	.000 <sup>j</sup>
	Residual	645.273	3097	.208		
	Total	745.714	3106			

a. Dependent Variable: Ebook

b. Predictors: (Constant), Printbook

.....

j. Predictors: (Constant), Printbook, Audiobook, EDUC, age, Income, libapp, Student, Employed, libweb

**Table 4c Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.729	.028		25.587	.000
	Printbook	-.362	.030	-.212	-12.104	.000
.....						
9 Final Linear Model	(Constant)	.526	.043		12.211	.000
	Printbook	-.342	.029	-.201	-11.889	.000
	Audiobook	.201	.021	.162	9.477	.000
	EDUC	.027	.005	.104	5.366	.000
	age	-.003	.000	-.107	-5.706	.000
	Income	.017	.004	.084	4.515	.000
	libapp	.068	.028	.043	2.405	.016
	Student	.218	.064	.060	3.413	.001
	Employed	.043	.019	.042	2.250	.025
libweb	.040	.018	.041	2.226	.026	

a. Dependent Variable: Ebook

Based on Table 4c, our final linear regression equation is:

$$\text{Ebook}_{\text{predicted}} = 0.526 - 0.342\text{Printbook} + 0.201\text{Audiobook} + 0.027\text{EDUC} - 0.003\text{age} + 0.017\text{Income} + 0.068\text{libapp} + 0.218\text{Student} + 0.043\text{Employed} + 0.040\text{libweb}$$

The coefficient of determination, R<sup>2</sup>, of the final model is 0.135, which means (that we estimate that) 13.5% of the variability in “Ebook” has been explained by the variables in the final model.

#### FURTHER ANALYSIS OF THE SIGNIFICANT VARIABLES

Since the “Cluster 1 General reading behaviors” is the most predictive cluster for the e-reader behavior among all three clusters above, we decided to conduct independent sample t-tests on “Printbook” and “Audiobook,” variables which are included in the final model. We can see from Table 5a that for people who read e-books in the past 12 months, the mean values of

“Printbook” and “Audiobook” are both, respectively, different from those for the people who did not read e-books in the past 12 months, at 5% significant level (p-values are both zero to 3 digits).

**Table 5a Independent Samples t-Test**

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Printbook	Equal variances assumed	652.306	.000	12.26	3533	.000	.120	.010	.100	.139
	Equal variances not assumed			10.88	1908.3	.000	.120	.011	.098	.141
Audiobook	Equal variances assumed	678.374	.000	-13.32	3544	.000	-.179	.013	-.205	-.152
	Equal variances not assumed			-12.40	2254.7	.000	-.179	.014	-.207	-.150

Table 5b shows that the mean value of “Printbook” is higher for people who didn’t read e-books compared to those that did. The coefficient of “Printbook” in our final model is - 0.342. Therefore, both our final model and the t-test result support the fact that people who read at least one printed book in the past 12 months have lower probability of reading e-books than those who did not read a printed book<sup>2</sup>. This is not a totally "obvious" result. On one hand, it is clear by definition that people who read print books are less likely to also read e-books than those who do read e-books. However, it can be said about people who have read at least one printed book that at least they read books!! One who reads no books surely does not read e-books!!

**Table 5b Group Statistics**

	Ebook	Mean	Std. Deviation	Std. Error Mean
Printbook	No	.955	.208	.004
	Yes	.836	.371	.010
Audiobook	No	.129	.335	.007
	Yes	.308	.462	.013

Similarly from Table 5b, it can be seen that the mean value of “Audiobook” is higher for people who read e-books than for those who do not read e-books. The coefficient of “Audiobook” in our final model is, positive, 0.201. So, both the our final model and t-test result support that people who used audio books in the past 12 months have higher probability of reading e-books.

<sup>2</sup> It must be noted that the interpretations are not exactly identical. The regression result indicates that there is a significant difference in the stated direction, holding the other variables in the model constant. The t-test result is indicating that there is a significant difference in the stated direction without controlling for any other variables. It is NOT pre-ordained that both results must match, either in terms of significance or in terms of direction. The spirit of this footnote holds true for all of the other t-test results we discuss.

Next, we performed independent sample t-tests on the “EDUC,” “age,” “income,” “libapp,” and “libweb” variables in the final model to see whether there is difference in mean between people who read e-books and those who didn’t. Table 6a shows that for all these variables, the mean values are different between people who read e-books and those who didn’t at 5% significant level (in each case, p-value is zero to 3 digits).

**Table 6a Independent Samples Test**

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower Upper	
EDUC	Equal variances assumed	6.449	.011	-9.13	3538	.000	-.587	.064	-.713	-.461
	Equal variances not assumed			-9.04	2800.9	.000	-.587	.065	-.714	-.460
age	Equal variances assumed	31.077	.000	7.10	3503	.000	4.605	.648	3.334	5.877
	Equal variances not assumed			7.29	3131.6	.000	4.605	.632	3.366	5.845
Income	Equal variances assumed	7.292	.007	-8.59	3170	.000	-.765	.089	-.939	-.590
	Equal variances not assumed			-8.65	2734.8	.000	-.765	.088	-.938	-.591
libapp	Equal variances assumed	235.164	.000	-7.63	3524	.000	-.079	.010	-.100	-.059
	Equal variances not assumed			-7.10	2236.4	.000	-.079	.011	-.101	-.057
libweb	Equal variances assumed	18.084	.000	-7.69	3531	.000	-.131	.017	-.165	-.098
	Equal variances not assumed			-7.67	2857.453	.000	-.131	.017	-.165	-.098

From Table 6b, we can note various outcomes -

People who read e-books have a higher mean value for EDUC than those who do not read e-books, and the coefficient of EDUC in our final regression model is 0.027. So, our final model result is consistent with the t-test result, that for people with higher education level, the probability of reading e-books is higher than for people with lower education. That e-books are more popular among more highly-educated people would seem to agree with common sense.

People who read e-books have lower mean age than those who do not read e-books, and the coefficient of age is -0.003 in our final regression model. So, again, the final regression model result is consistent with our t-test result that the older a person is, the lower the probability that he/she is an e-book reader. This result also would seem to make sense, as younger people tend to be more technologically proficient (as well as technology accepting!!)

People who read e-books have higher mean income than those who do not read e-books, and the coefficient of income is 0.017. So, yet again, the final regression model test result is

consistent with the t-test result, that for people with higher income, the probability of reading e-books is higher than for people with lower income. This again makes sense, as higher income is more likely to allow the purchase of an e-reader, as well as higher income correlates somewhat with higher education.

People who read e-books have higher mean values of both libweb and libapp. The coefficients of libweb and libapp are 0.040 and 0.068, respectively in the final regression model. So, we continue with consistency between the regression model results and the t-test results, that for people who use the website and apps of the public library, the probability of reading e-books is higher than, respectively, those who do not use the website or apps of the public library.

**Table 6b Group Statistics**

	Ebook	Mean	Std. Deviation	Std. Error Mean
EDUC	No	4.321	1.829	.039
	Yes	4.908	1.911	.0518
age	No	46.160	19.500	.421
	Yes	41.554	17.387	.471
Income	No	4.925	2.478	.057
	Yes	5.689	2.406	.0680
libapp	No	.071	.257	.006
	Yes	.151	.358	.010
libweb	No	.414	.493	.011
	Yes	.546	.498	.014

In addition, for "Q emplnw" variable, both "Student" and "Employed" are included in the final model. We recoded "Q emplnw" into EMPSTU with scale "1= Employed (including 1, 2 and 5 in the original scale)", "0=Student (including 7 in the original scale)" and conducted an independent sample t-test again to see which "occupation" has a higher probability of reading e-books. From Table 7a, we can see that at 5% significant level, the mean value of "Ebook" is significantly different between "Student" and "Employed" (p-value = .002), and Table 7b shows that the mean value of "Ebook" is higher for "Student," which means that students have a higher probability of reading e-books than those classified as "Employed." Since textbooks are often purchased in e-book form, along with other reasons, this result makes sense.

**Table 7a Independent Samples Test**

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Ebook	Equal variances assumed	8.980	.003	3.134	2251	.002	.203	.065	.076	.330
	Equal variances not assumed			3.195	62.4	.002	.203	.064	.076	.330

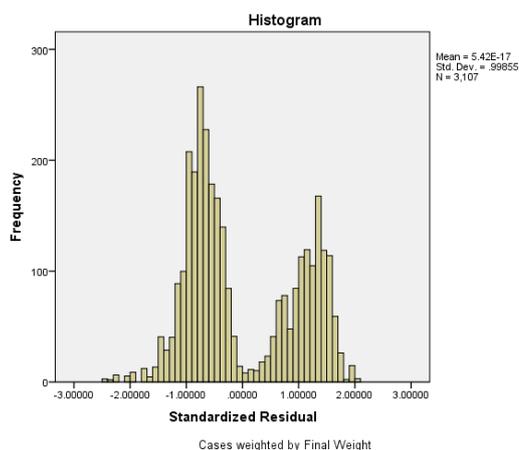
**Table 7b Group Statistics**

	EMPSTU	Mean	Std. Deviation	Std. Error Mean
Ebook	Student	.635	.485	.063
	Employed	.432	.495	.011

## CONCLUSIONS, LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

According to the analysis above, we concluded firstly, that people who use audiobooks, a website of a public library, and apps of a public library have higher probability of reading e-books than, respectively, those who do not. So, websites and apps of audiobook sellers or the public library should be good channels for e-books advertisements. Secondly, the probability of reading e-books is higher for students than those of other occupations. Lastly, younger people with higher income and a higher education level have a higher probability of reading e-books. The second and third conclusions may be useful for marketing segmentation of the digital publishing industry or e-reader device companies.

However, there are some limitations in this research. We already noted the fact that the analysis uses linear regression, even though the dependent variable is a dichotomous variable. Figure 1 displays that the distribution of Standardized Residuals of the final model is not normal (i.e., bell-shaped). This is not unexpected. Still, according to Gauss–Markov theorem, while the residuals are not normally distributed (a good case can be made that the residuals are a mixture of two normal distributions), our final model still contains the best linear unbiased estimators (Hill, Griffiths, & Grey 2011). Since the dependent variable is a binary variable, binary-logistic regression can be performed in a future study to see whether the analysis of e-reading behavior yields a different result. The authors do not believe that the fundamental significance and directional results would change.



**Figure 1 Histogram of Standardized Residual**

We also mentioned previously that some of our independent variables could have been set up as sets of categorical/dummy variables to avoid a built-in assumption of a linear relationship between the dependent variable and such variables as Age, Income, and Education.

In addition, the research relies on the analysis of 10 variables, among which 5 are general demographic characteristics, and only 5 are behavioral characteristics. Future research may wish to be more comprehensive, and include more behavioral variables, such as accessibility of e-books, usage of social media, and other variables that are related specifically with digital reading.

Finally, as noted earlier, this study had a sample that was taken totally from the United States. Future research may wish to extend this work to include data from other countries, with the objective of seeing if the results are the same or different, and if different, to try to gain insight by trying to determine why they are different.

**References**

Hill, R. Carter, Griffiths, William, & Guay Lim (2011). *Principles of Econometrics*, 62-63.

Huang, Y., Liang, T, & C. Chiu (2013). Gender Differences in the Reading of E-books: Investigating Children's Attitudes, Reading Behaviors and Outcomes. *Educational Technology & Society*, 16 (4), 97-110.

Perrin, Andrew (2016). Retrieved from <http://www.pewinternet.org/2016/09/01/book-reading-2016/>

Pew research center (2016). March 7-April 4, 2016 – Libraries (Dataset). Retrieved from <http://www.pewinternet.org/datasets/>

Shahriza, Nor, Karim, Abdul, & Amelia Hasan (2007). Reading habits and attitude in the digital age: Analysis of gender and academic program differences in Malaysia. *The Electronic Library*, 25(3), 285-298.

Zhang, Liyi, & Wei Ma (2011). Correlation analysis between users' educational level and mobile reading behavior. *Library Hi Tech*, 29(3), 424-435.