Representation of Race and Sex of Characters in Children's Books

May 13, 2019

Abstract

Representation is a topic of great importance as the United States becomes increasingly diverse, particularly because of the impact on young children. In this paper, children's books from Drake Library in Grinnell, Iowa were analyzed to determine if one sex or race was overrepresented. We find that (a) there are a higher proportion of female characters than male characters ($p=0.0010$), (b) there are a higher proportion of white characters than characters of color ($p=0.0000$), (c) protagonist race and sex are not associated and are not confounded by publication year or author sex, and (d) protagonists are equally represented across each sex and race. We hope that this study will call to attention the overrepresentation of characters in children's books, and will prompt a change to equal representation among all sexes and races of characters and protagonists in picture books.

Introduction

Children's books–especially the images and sentiments presented in them–play a large role in shaping how children think about the world. Additionally, as census data reveals an increasingly racially diverse United States, the representations of race and gender in these books carry great importance (Chappell 2017). Though primary literature on the matter is limited, those offered suggest that people of color are less represented in children's transition books featuring more text than pictures (Hughes-Hassel et al. 2009). Grauerholz and Pescosolido (1989) also found that the prevalence of female and male characters vary over time in children's literature. More recently, a study led by Janice McCabe (2011) at Florida State University analyzed nearly 6,000 children's books, finding male characters are represented twice as much as females in titles of books and 1.6 times as much in central character roles. Bell et. al (2011) suggest in *Media Representations and Impact on the Lives of Black Men and Boys*, that many forces shape the lives of people of color, and among the most important for maintaining (or changing) societal barriers placed by historical, cultural, and political agenda, are mass media (Bell et al. 2011). This literature suggests that overrepresentation in children's can positively or negatively influence perceptions for people of different sex and race. Given these results, we are interested in the following: Is there a significant difference in representation of characters of a certain sex or race in picture books available to children in Grinnell, Iowa? We hypothesize that white characters and male characters will be overrepresented across the children's book section. Our hypothesis is based on Grinnell's 90.5% white population (US Census Bureau 2018) and McCabe's (2011) findings that male characters are represented twice as much as female characters. Additionally, we analyze not only total character counts but also race and sex of protagonists. This will provide an insight to the 'good guy' specifically, adding on to McCabe's findings about book titles and central characters. This will be an attempt to approximate sentiment analysis.Through sentiment analysis, characters in a book can be categorized into i.e 'good' or 'bad' according to textual cues and visuals in the book.

Methods

Data collection was conducted at Drake Library in Grinnell, Iowa. The sampled books were drawn from the library's Children's Books section, all of which were picture books. There, the books are categorized by the letter of the author's last name. To sample, we generated random letters online (http://www.dave-reed.com/Nifty/randSeq.html) to determine which shelf to pick the books from and blindly picked a book from the selected shelf (n=50). We counted the total number of white characters, characters of color, female characters, and male characters. The data collection was done including background characters and avoiding double-counts. We also categorized the race and sex of the protagonist. We believe that author sex and the year the book was published could be confounders. Author's sex could affect the representation of female and male characters--female authors might include more female characters and vice versa. The year of publication reflects the changes in society, which may translate to changes in the representation of women characters and characters of color. We dichotomized the publishing year as before the median year 2001 (coded as '0') and after 2001

Figure 1. The proportion of female and male characters.

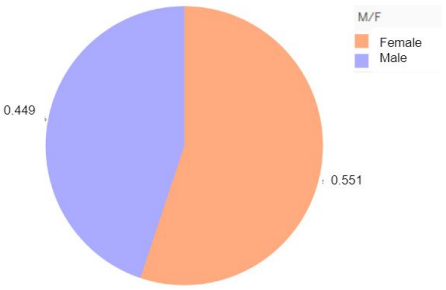
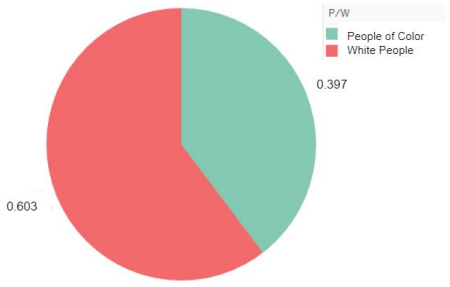


Figure 2. The proportions of characters of color and white characters.



(coded as ‘1’) to simplify the analyses and improve interpretation.

In order to find out whether there was a difference between the proportion of female and male characters as well as white characters and those of color, we utilized a randomization test for difference in proportions, for example, the total number of characters of color out of the total number of characters. To quantify any difference observed, we constructed confidence intervals through bootstrap distributions.To see if author’s sex and the year the books were published had an impact on the protagonist’s sex and race we use chi-squared test for association between author’s sex and protagonist’s sex author’s sex and protagonist’s race. Similarly, we tested for association between publication date (coded as: <2001=0, >2001=1) and the race and sex of the protagonist. To see if both races and both sexes were equally represented in the protagonists of these children’s books, we used a chi-squares goodness-of-fit test.

Figure 3

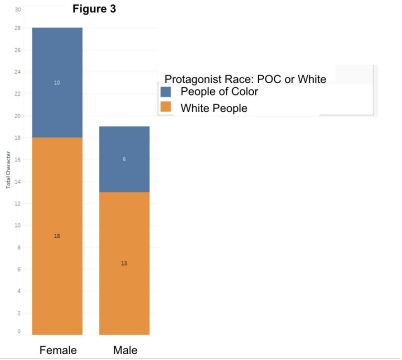
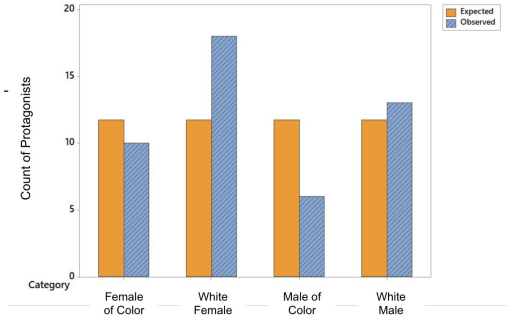


Figure 4
Chart of Observed and Expected Values



Results

Our first question was whether or not there was a higher proportion of males or females across the children’s books section. We decided to use the count of total characters as an indicator of this because it shows representation of sex across the book. Our second question was whether or not there was a higher proportion of white people than people of color. We decided to use the count of total white characters and count of total characters of color as an indicator of this because it would show the representation throughout the whole book. For both of our questions, we utilized randomization testing as the sample size was sufficiently large and the displayed proportions of male & female characters, as well as white characters & those of color, are roughly equal (figures 1 and 2). Thus the assumptions behind a randomization test were met. We conclude from our statistical analyses that white characters are represented more than characters of color, and female characters more than male characters. To quantify the true difference in proportions of female and male characters as well as the difference in proportions of white characters and characters of color, we used bootstrap sampling to generate 95% confidence intervals. Our 95% confidence intervals showed that the true difference between the proportion of female and male characters lies between 0.047 and 0.160, whereas the difference between the proportion of white characters and those of color lies between 0.096 and 0.214.

Overall, we can conclude that female characters and white characters are overrepresented in the children’s books at Drake’s Library.

To test whether the protagonist’s race had an association with the protagonist’s sex (figure 3), we performed a chi-square test for association to see whether or not these two variables were related. Before we would be able to do this, we had to make sure that author sex and publication year were not confounding variables, which meant we performed four chi-squared tests for associations: author sex vs. protagonist race, author sex vs. protagonist sex, publication year vs. protagonist race, and publication year vs. protagonist sex. Each of these tests were not-significant, using a Bonferroni adjusted significance level of $\alpha=0.0125$, so we can conclude that these variables are not associated. These results indicate that author sex and year of publication are not confounding variables, and a chi-squared test between

protagonist sex and race can be done. The results of this test show that protagonist race and protagonist sex are not associated, and the nature of this association is not affected by either publication year or author's sex (p=0.769). In sum, protagonist race does not have an impact on protagonist sex or vice versa.

We were also curious if the representation of characters was equal across each combination of sex and race. Thus we performed a chi-square goodness-of-fit test (figure 4, p=0.088) and conclude that female protagonists of color, male protagonists of color, white female protagonists, and white male protagonists are all equally represented.

Discussion

Our findings outlined above show that there is significantly higher proportions of female characters (p=0.0010) and white characters (p=0.000) in children's books available at Drake Library in Grinnell, Iowa. Thus, we fail to reject our hypothesis that men are overrepresented in children's books, yet the data supports our hypothesis that white people are overrepresented in children's books. Additionally, our chi-square test for goodness-of-fit found that there are no statistically significant differences amongst the number of protagonists that are white men, men of color, white women, and women of color (p=0.088, α=0.05). This does not support our hypothesis that white males will be overrepresented, but provides evidence of equal representation by sex and race. These results contradict the work of McCabe (2011) who found that men are 1.6 times more represented than women as central characters. However, McCabe(2011) had a much larger sample size (n = 5618) taken from popular publishers or award-winning books. Future studies should evaluate the relationship between the popularity of books and the representation of characters found in them, as popular books could confirm the implicit biases of the reader via skewed representation.

One important point is that due to the time and location constraints, we were unable to obtain a larger sample size from a more diverse set of locations. Therefore, we are only able to generalize our findings to the children's section of Drake Library, which is located in a town with a progressive college and is likely well-funded. We believe it would be interesting to study in more locations to determine if any factors, like lack of funding, price of the books, or population demographics impact character and protagonist representation.

Lastly, while we did find a statistically significant proportion of white characters, it is important to understand several points about Grinnell's population demographics and assumptions about representation that follow. The first is Grinnell's aforementioned 90.5% white population (US Census Bureau 2018). In this light, our 95% confidence interval for the percentage of white people (56.5%, 64.5%) and for people of color (35.4%, 43.2%) could be considered roughly representative of people of color. However, the conclusion from the previous statement ignores that we grouped the majority of the world's population in one racial group: people of color. Because of this, representation of specific racial groups could still be wildly underrepresented. Lastly, in our goodness-of-fit model, we assumed that the proportions of characters of color and white characters should be strictly equal. Determining what proportions of races should be represented is a lofty question. It is certainly beyond the scope of this study, and likely pulls from disciplines other than statistics to answer. However, it would be useful for a future study to investigate association between age of first noticing reading about a main character of a similar race to the reader and adult reading level to measure the effects of representation of children's literature on lifelong reading.

Bibliography

Bell, J.D., E.D. Janis, C. Moore, J. Bailin, J. Lerman, L. Siegel. 2011. “Social Science Literature Review: Media Representations and Impact on the Lives of Black Men and Boys.” The Opportunity Agenda. https://www.racialequitytools.org/resourcefiles/Media-Impact-onLives-of-Black-Men-and-Boys-OppAgenda.pdf?fbclid=IwAR1v2Wgx6hX4wcqzAqpz6tQQE4fg6HdL0LKUChOsU2oaOjbYDeZw0ACj-xQ

Chappell, Bill. 2017. “Census Finds a More Diverse America, as Whites Lag Growth.” National Public Radio. https://www.npr.org/sections/thetwo-way/2017/06/22/533926978/census-finds-a-more-diverse-america-as-whites-lag-growth

Grauerholz, E., B. A. Pescosolido. 1989. “Gender Representation in Children’s Literature from 1900-1984”. Gender and Society 113-125. SAGE Journals. https://journals.sagepub.com/doi/pdf/10.1177/089124389003001008

Hughes,Hassell, S., H.A. Barkley and E. Koehler. 2009. “Promoting Equity in Children’s Literacy Instruction Using a Critical Race Theory Framework to Examine Transitional Books.” American Association of School Librarians. https://files.eric.ed.gov/fulltext/EJ877497.pdf

McCabe, J., E. Fairchild, L. Grauerholz, B.A. Pescosolido, D. Tope. 2011. “Gender in Twentieth-Century Children’s Books: Patterns of Disparity in Titles and Central Characters.” Gender and Society 25: 197-226. SAGE Journals. https://journals.sagepub.com/doi/abs/10.1177/0891243211398358

United States Census Bureau. 2018. “Quick Facts: Grinnell, Iowa.” census.gov https://www.census.gov/quickfacts/grinnellcityiowa

Appendix.

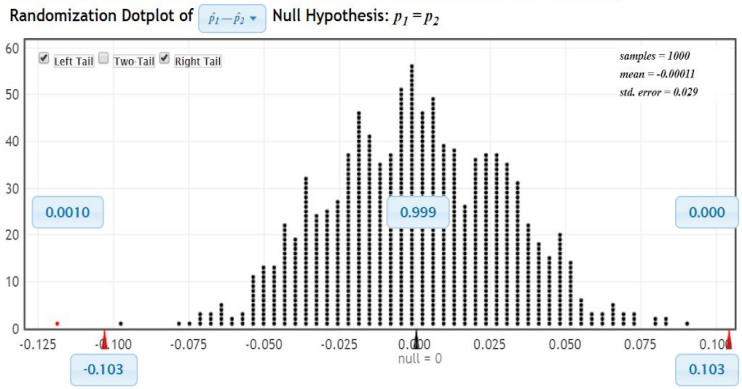


Figure 5. Randomization test for difference in proportions of female characters and male characters.

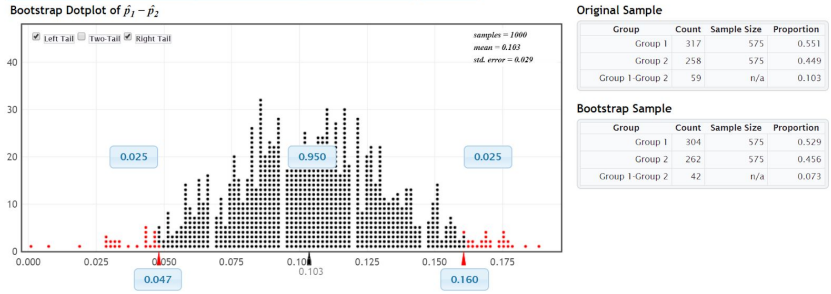


Figure 6. 95% confidence interval for difference in proportions of female characters and male characters.

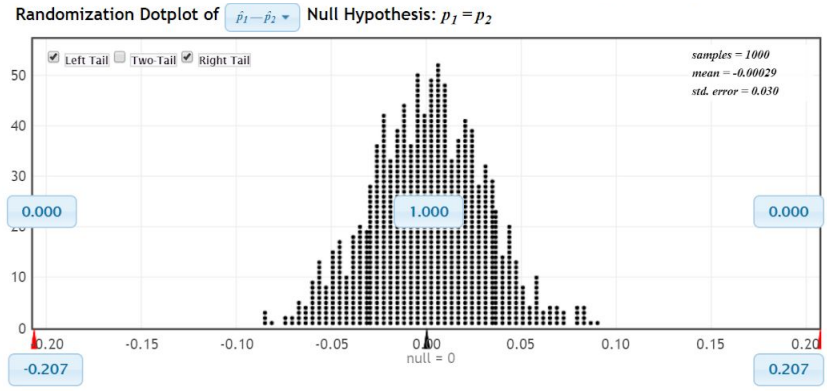


Figure 7. Randomization test for difference in proportions of white characters and characters of color.

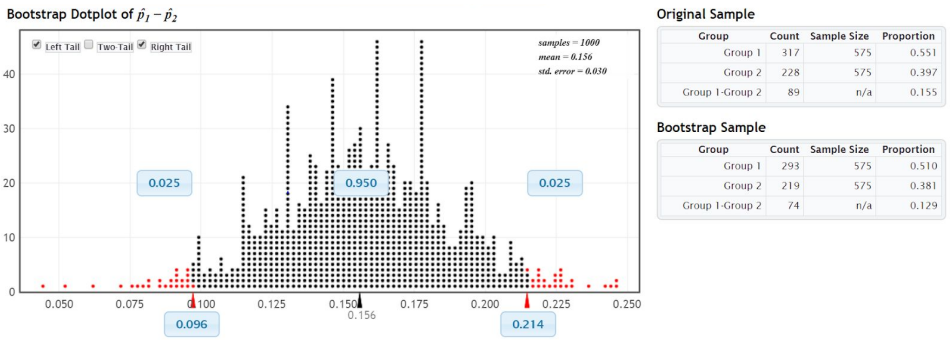


Figure 8. 95% confidence interval for difference in proportions of white characters and characters of color.

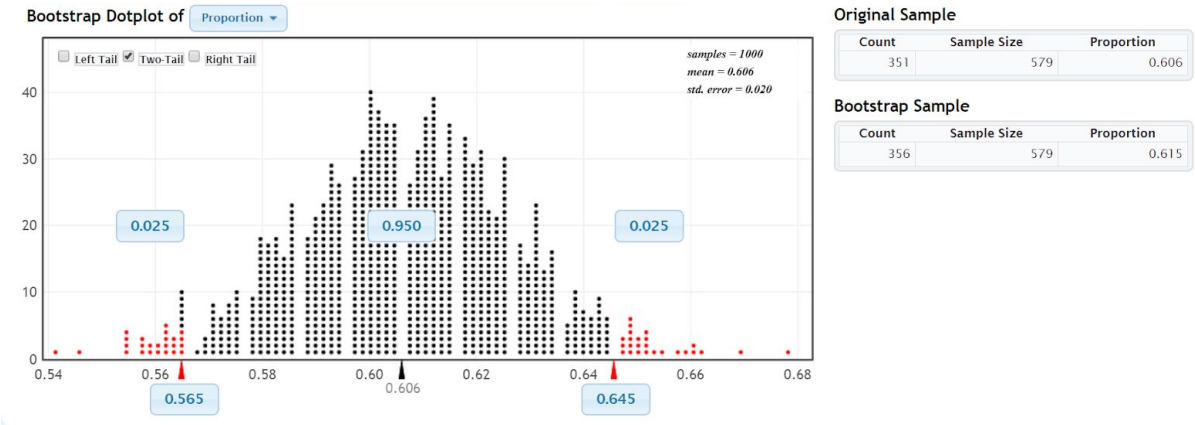


Figure 9. 95% confidence interval for proportion of white characters.

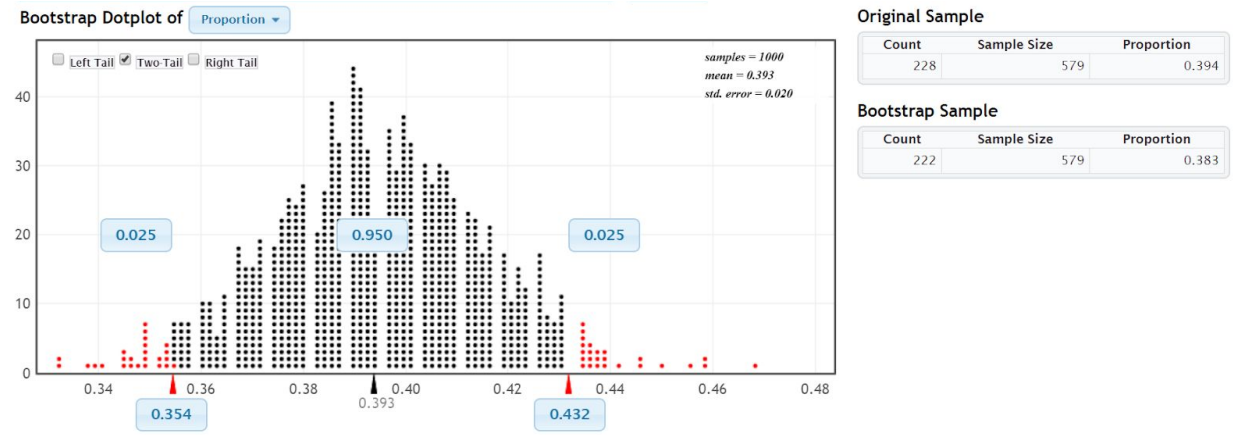


Figure 10. 95% confidence interval for proportion of characters of color.

Chi-Square Test for Association: Protag, ProtagRace

Rows: Protag Columns: ProtagRace

	POC	White	All
Female	10	18	28
	9.532	18.468	
Male	6	13	19
	6.468	12.532	
All	16	31	47

Cell Contents
Count
Expected count

Chi-Square Test

	Chi-Square	DF	P-Value
Pearson	0.086	1	0.769
Likelihood Ratio	0.087	1	0.769

Table 1. Chi-square test for association between protagonist sex and protagonist race.

Chi-Square Goodness-of-Fit Test

Observed and Expected Counts

Category	Observed	Test Proportion	Expected	Contribution to Chi-Square
FP	10	0.25	11.75	0.26064
FW	18	0.25	11.75	3.32447
MP	6	0.25	11.75	2.81383
MW	13	0.25	11.75	0.13298

Chi-Square Test

N	DF	Chi-Sq	P-Value
47	3	6.53191	0.088

Table 2. Chi-square goodness-of-fit for all combinations of sex and race for protagonists.