

ISSN 2560-9815

Volume 6 | Issue 1 | Summer 2022

The Young Researcher



RSGC

Royal St. George's College

The Young Researcher
SUMMER 2022 VOLUME 6, NUMBER 1

A journal dedicated to the publication of original research from secondary school students

Editor-In-Chief
John Lambersky

Associate Editor
Matt Mooney

Royal St. George's College Editors

Jack Beatty
Lucas Hardie
Iain Playfair
Toby Salamon
Jack Skippon

Alumni Editors

Nicholas Bethlenfalvy
Ryan Hamilton
William Howard-Waddingham
Andrew Pyper

Financial Support

Royal St. George's College

Editorial Office

120 Howland Avenue
Toronto, ON, Canada M5R 3B5
www.theyoungresearcher.com

Layout and Design

Jonathan Cresswell-Jones

Consulting Editors

Suzanne Conklin Akbari
Becky Bakhshaei
David Baldesarra
Barrie Bennett
Anthony Campbell
Jeremy B. Caplan
Nathan Chow
Hance Clarke
Will Fripp
Michael Gemar
Jennifer Goldberg
Margaret S. Herridge
Ted Higginbotham
Tim Hutton
Ira Jacobs
Blake Lee-Whiting
Matt Mooney
Lori Loeb
Jaime Malic
William J. McCausland
Gurbir Perhar
Kate Schumaker
Eva Serhal
Michael Simmonds
Sydney Stoyan

Produced and Distributed by
Royal St. George's College

ISSN 2560-9815 (Print)
ISSN 2560-9823 (Online)

Table of Contents

- 5 The Effects of COVID-19 on the Mental Health of Senior Citizens and High School Students
Uma Kamath
- 20 Immigrant Parents and Academic Success
Breanna Villarreal
- 46 Average High School Teacher Salaries vs Student Performance on Standardized Tests: A Correlational Analysis of Northern New Jersey High Schools
Justin Sousa
- 60 Using Recurrent Neural Networks and Web Crawlers to Scrape Open Data from the Internet
Dhiraj Kanneganti
- 72 Assessing the Threat: Antibiotic Resistant Bacteria Near Pittsburgh Hospitals
Brendon Frankel
- 84 Application of Data Envelopment Analysis in a Small-Scale Transportation Network
Alexandra McWatters
- 94 Investigation on the Effect of Infill Orientation on the Flexural Properties in FDM Parts
Alexander Barry
- 106 Analyzing Twitter Data to Understand COVID-19 Vaccine Hesitancy Following the Federal Private Sector Mandate
Manisha Palaniappan
- 120 Unfounded Confidence? Gender Differences in Overconfidence Across Toronto Adolescents
Jessica Yu

Editorial

Research is the culmination of an exhaustive, systematic, and repeatable pursuit of objective knowledge. It is the culmination of consistency and the furtherance of humans' understanding of our place in this universe. Research is interdisciplinary and expansive. It is honest, objective, and deeply profound. Nevertheless, in the modern-day of free information, research finds itself under threat as never before. Subjectivity, partisanship, and misinformation are a plague corrupting the foundation of our society, and in the new age of misinformation, cracks are beginning to show in our longest-standing pillars. Free, objective and clear research has never before been as vital to the state of the world. In the humble pursuit of this freedom of knowledge, we, the editorial team, proudly present this edition of *The Young Researcher*.

The future of research rests in the hands of our youth. Sensationalist academia, emotional journalism, and cherry-picked news sources are trending towards prominence in both education and society. Many have resigned to their hopelessness in the face of this polarized landscape, and academic honesty is seeing itself enter into a slow period of decline. These wrongs, however, may still be righted. Students, taught the strategies of true research, can right the tipping of the metaphorical ship. Indeed, our editorial team's opinion is that, while professional fields may be increasingly infiltrated by academic dishonesty, younger academic fields remain rich with inquisitive vigor. Students, particularly secondary school students, have in these new times developed unique methods of parsing this influx of information to both satiate their curiosity and form nuanced opinions on the world, academia, and science as a whole.

Moreover, for many youths, the research process involves a profound personal connection and a need to strive for truth that older academics are often hailed for. Tragically, however, students have often found these efforts unacknowledged. For these purposes, we submit today the 2022 *Young Researcher*.

The Editors

Jack Beatty
Lucas Hardie
Iain Playfair

Toby Salamon
Jack Skippon

The Effects of COVID-19 on the Mental Health of Senior Citizens and High School Students

Uma Kamath

Mental health disorders are a leading health issue among adults and teenagers. Public health organizations and governments have highlighted an increase in mental health issues during the COVID-19 pandemic. This research uses a survey of 83 students and 95 senior citizens to understand the impact of COVID-19 on mental health. The results indicate that COVID-19 has negatively affected the mental health of students, seniors living in nursing homes and, to a lesser extent, seniors living at home. The impact is greater for external factors than internal factors for students and seniors living in nursing homes. These results have implications for future research on online education, social media usage, and other forms of technologies that limit in-person interactions. The results also indicate that it is important for public health policy to weigh the benefits of various restrictions against the detrimental effects such restrictions have on mental health.

Keywords: Mental health, COVID-19, students, senior citizens, internal factors, external factors

Introduction

Approximately 20% of adults experience a mental health disorder in some form, including depression, anxiety, and insomnia, regardless of external factors (Gerlach, 2021). Left untreated, mental health disorders can have serious impacts on people of all ages. In early 2020, many places in the United States underwent lockdown and experienced social distancing restrictions due to the COVID-19 pandemic. The restrictions resulting from the lockdown forced people to stay at home or be isolated from others. Various researchers (Abbott, 2021; Lee et al., 2020), public health organizations (Centers for Disease Control National Center on Health Statistics, 2020; World Health Organization, 2022), governments (U.S. Department of Health & Human Services, 2021; Washington State Department of Health, 2021), and news media out-

lets (CNBC, 2022; Reuters, 2022) have highlighted an increase in mental health symptoms among all age groups. The rapid increase in mental disorder symptoms raises questions and concerns about potential impacts of the pandemic on people susceptible to these symptoms.

There are very widely held perceptions about how mental health has been impacted among various individuals. These issues, manifested in the form of eating and sleeping patterns or the way people interact with others, could have unknown short-term and long-term effects on people's behavior. There have been and will continue to be numerous surveys and research studies to understand various aspects of the COVID-19 pandemic, including the effects on the general population. There have also been and will continue to be numerous studies to understand the psychological and physiological effects of the pandemic.

Despite these studies, the effect of the pandemic on mental health is not completely understood. This project aims to understand the extent of mental health problems specifically in senior citizens and high school students and gain an understanding of the contributing factors. This research project will help uncover answers to questions and concerns that a lot of people may have about possible changes and impacts to mental health over the past year and a half. The project will also help people with mental health concerns to better understand themselves and how the COVID-19 pandemic may have impacted them.

Review of Literature

Mental Health Overview

The World Health Organization defines mental health as “a state of well-being in which an individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and is able to make a contribution to his or her community” (World Health Organization, 2018). Mental health is the emotional, psychological, and social well-being that allows humans to think, emote, feel, act, and interact with each other (Center for Disease Control and Prevention, 2021; World Health Organization, 2018).

Poor mental health can be the result of several socioeconomic, biological, and environmental factors (World Health Organization, 2018). Some of these include genetic factors, physical health issues, unhealthy lifestyle including poor eating habits and lack of physical fitness, poor sleeping habits, physical violence, abuse, trauma, neglect, sexual violence, rapid social changes, gender discrimination, social exclusion, social or economic disadvantage such as poverty or debt, human rights violations, unemployment, loss of job, and stressful work environment. Being the victim of bullying has also been identified as a factor towards poor mental health in adolescents (American Psychological Association, 2010). Some of these potential causative factors, such as rapid social changes, social exclusion, poor sleeping habits, poor eating habits, and lack of adequate physical fitness, are likely to have been exacerbated due to the restrictions imposed during the COVID-19 pandemic.

Impact of Mental Health

Like physical health, good mental health is vital to humans and poor mental health can affect how we handle stress, relate to others, and make decisions (Center for Disease Control and Prevention, 2021). Some forms of poor mental health have been found to increase the risk of physical health conditions such as diabetes, heart disease, and stroke (National Institute of Mental Health, 2015). Poor mental health also manifests in the form of mental illnesses such as depression, anxiety, insomnia, post-traumatic stress disorder (PTSD) or major depressive disorder, and drug, opioid and substance use. Poor mental health can also lead to unhappiness and decreased enjoyment of life, family conflicts, relationship difficulties, social isolation, missed work or school, other problems related to work or school, risk-taking behaviors, human rights violations, legal and financial problems, poverty, and homelessness (World Health Organization, 2021). Children with mental health issues may underperform in schoolwork (Bowen, 2011) or be at risk of committing suicide (American Psychological Association, 2010). Healthy eating habits, alcohol and drugs avoidance, regular exercises, reduced time spent on social and online media, spending time with family and friends, volunteering, meditation, deep breathing and other relaxation exercises, setting realistic goals, seeking help when needed, therapy, and medical treatment are some of the remedies to improve mental health (Spielman, 2017; National Institute of Mental Health, 2021).

Impact of COVID-19 Pandemic on Mental Health

Mental health issues have always been prevalent across the world, even before the start of the COVID-19 pandemic. One in seven 10--19-year-olds globally is estimated to suffer from mental health issues, and suicide is the fourth leading cause of death among 15–19-year-olds (World Health Organization, 2021). 18.1% of U.S. adults ages 18 years or older are estimated to suffer from some form of mental health issue in any given year (U.S. Department of Health and Human Services, 2021).

Social distancing and other restrictions imple-

mented as a response to COVID-19 can most likely cause a negative impact on one's mental health (Javed, 2020). Initial studies conducted within the first three months after the onset of COVID-19 indicate an increase in various mental health issues among adolescents (Jones, 2020). Similarly, an increase in mental health issues was also found in adult age groups within the first five months after the onset of COVID-19 (Panchal et al., 2021). Approximately 24% of adults ages 65 and older reported mental health issues approximately 6 months after the onset of COVID-19 (Koma, 2020). Other studies, however indicated that compared with younger age groups, older adults were found to have experienced fewer mental health issues approximately 8 months after the onset of COVID-19 (Vahia, 2020).

Gap Analysis

Most studies were conducted within the first few months after the onset of COVID-19. Additionally, findings from these studies have yielded conflicting results. This project specifically aims to investigate the impact of COVID-19 on two groups of participants – high school students and senior citizens over the age of 65 -- and specifically after a longer time period with more time having passed for COVID-19 to have affected people. Mental health is a broad topic and previous studies have looked at the impact of COVID-19 on mental health more holistically. This project seeks to delve deeper into specific factors regarding social changes, social exclusion, thinking patterns, sleeping habits, eating habits, and physical fitness influencing mental health, classified into internal and external factors. Internal factors, such as eating patterns or stress levels, cause changes in the participants' thoughts, self-perception, emotions, emotion regulation, and feelings during the pandemic. External factors, such as participation in extracurricular activities or desire to spend time with others, cause changes in participants' interactions and communications with others, social behaviors, and desire to participate in social activities during the pandemic.

Methodology

Study Design

This study involved giving surveys to a group of high school students and a group of senior citizens. The survey included various statements related to the mental health of the participants, and each participant was asked to score on a scale of 1 - 5 how they felt after the pandemic for each of the statements, with 1 being the least relevant to the participants and 5 being the most relevant. As shown in Appendix A and Appendix B, survey statements were grouped into internal statements, external statements, and control statements. Internal statements included any changes in participants' own feelings, thought processing, and emotions during the pandemic. External statements included any changes in the way participants behave and interact with others during the pandemic. Control statements included participants' feelings, behaviors, thought processing, and emotions irrespective of an external stimulus. Thus, the control statements signify the participants' inherent traits and pre-pandemic scores and serve as a baseline to identify changes in their post-pandemic state, which are measured through the internal and external statements. Each internal and external statement included a corresponding control statement and the difference between the post pandemic score and the corresponding control score represented the impact of the pandemic on their mental health.

Participants

Participants for the high school group were drawn from students at Tompkins High School in Katy, Texas who were attending school remotely for one or both semesters during the 2020-2021 school year. Participants for the senior citizens group were people of age 65 and above in Houston and neighboring suburbs, either living at home or nearby nursing care facilities that were closed to visitors for about a year. Although a bigger sample size was desirable, COVID-19 and other limitations allowed for a sample size that consisted of 83 high school students, 53 senior citizens from nursing care facilities, and 42 senior citizens living at home. Interested high school participants were obtained by a few teachers making announcements in

their classes and word of mouth communications. All participants were clearly informed that their participation was completely voluntary and that they could withdraw at any time, including after having participated in the experiment. Additionally, the parent or guardian of each high school participant was required to provide written consent (shown in Appendix C) for participation. Several senior nursing care homes in Katy were approached to recruit participants, but only three senior nursing homes participated in the project. Seniors living at home were solicited through phone calls or face-to-face requests.

Survey Mechanism

The participants were grouped into two categories - high school students and senior citizens. Additionally, the senior citizens were sub-grouped by where they lived – in a nursing care facility or their home. Each participant was asked to fill out the survey online through a Google form but was also given the option to fill it out on paper. The survey provided a list of statements, and the participants were asked to rate on a scale of 1 - 5 how much each statement related to them. To prevent any bias, each participant was given the same survey within their group, and surveys were kept anonymous. The survey for high school students comprised 38 statements of which 11 pertained to the participant’s internal statements, 8 pertained to external statements, and the remaining 19 represented control statements corresponding to each internal and external statement. The survey for senior citizens comprised 34 statements of which 10 pertained to the participant’s internal statements, 7 pertained to external statements, and the remaining 17 represented control statements corresponding to each internal and external statement. For the most part, the high school students and senior citizens were asked the same survey questions so that the two groups could be compared in a reasonable manner. However, each test group was also asked a few questions that pertained specifically to their age group to understand differences between age groups. Copies of the survey statements are shown in Appendix A and Appendix B.

Care was taken to maintain complete confidentiality. Names and other personal information were not collected, recorded, or published. No gender, racial/ethnic, or other personal information were observed

or collected. All scores were kept anonymous, and it is impossible to link the scores to any of the individual participants. This process was reviewed and approved by an Institutional Review Board before administering the surveys.

Data Analysis and Hypotheses

The raw scores provided by each participant to each survey statement were tabulated. The statements were grouped by internal and external types and score averages were analyzed against equivalent control score averages. For each group of statements, the difference between the control scores and post pandemic scores represented the magnitude of the effect of the pandemic on their mental health. The analysis was similarly repeated for all the statements grouped together.

Change in internal group scores was also compared to change in external group scores. Change in scores for seniors living in nursing homes was compared to seniors living in their homes. Finally, change in scores for high school students was compared to change in scores for all seniors.

For comparing scores within each group, a paired sample t-test was used. Assuming the pandemic has impacted the mental health of participants, the null hypothesis is that the score after the pandemic is less than or equal to the control score.

$$\begin{aligned}
 H_0: \mu_{\text{PostPandemicScore}} &\leq \mu_{\text{ControlScore}} \\
 H_a: \mu_{\text{PostPandemicScore}} &> \mu_{\text{ControlScore}}
 \end{aligned}$$

For comparing the changes in internal group scores to changes in external group scores, a paired sample t-test was used. Assuming that isolation has impacted participants’ thoughts and emotions to a greater extent than their interactions with others, the null hypothesis is that the change in internal group scores is greater than or equal to the change in external group scores.

$$\begin{aligned}
 H_0: \mu_{\text{Internal}} &\geq \mu_{\text{External}} \\
 H_a: \mu_{\text{Internal}} &< \mu_{\text{External}}
 \end{aligned}$$

For comparing the change in scores for seniors living in nursing homes to seniors living at home, a two-sample t-test was used. Assuming seniors living in nursing homes had fewer interactions with oth-

EFFECTS OF COVID-19 ON MENTAL HEALTH – SENIOR CITIZENS AND STUDENTS

ers than seniors living at home, the null hypothesis is that the change in scores for seniors living in nursing homes is less than or equal to the change in scores for seniors living at home.

$$H_0: \mu_{\text{NursingHome}} \leq \mu_{\text{Home}}$$

$$H_a: \mu_{\text{NursingHome}} > \mu_{\text{Home}}$$

For comparing the change in scores for high school students to the change in scores for all seniors, a two-sample t-test was used. Assuming seniors were better able to cope with the pandemic restrictions, the null hypothesis is that the change in scores for high school students is less than or equal to the change in scores for all seniors.

$$H_0: \mu_{\text{HighSchoolStudents}} \leq \mu_{\text{Seniors}}$$

$$H_a: \mu_{\text{HighSchoolStudents}} > \mu_{\text{Seniors}}$$

The data was then analyzed using either a paired sample t-test or a two-sample t-test, as specified earlier depending on the groups being analyzed, in Mi-

crosoft Excel to identify any statistically significant inferences. A two-sample t-test is a statistical test used to infer whether the means of two data sets are significantly different (Mendenhall et al., 1995). The p-value or observed significance level, which is the smallest value of α for which test results are statistically significant (Mendenhall et al., 1995) will be compared against a significance level (α) of 0.05 to determine whether the null hypothesis can be rejected or not.

Results

Averages for internal, external, and all statements for each group of participants are plotted in Figures 1, 2 and 3. Each of these graphs indicate that the scores have increased after the pandemic. Change in scores appear to be higher for external statements than for internal statements for high school students and seniors living in nursing homes. These graphs also indicate that the change in scores is relatively smaller for seniors living at home.

Figure 1: High School Student Response Average Scores

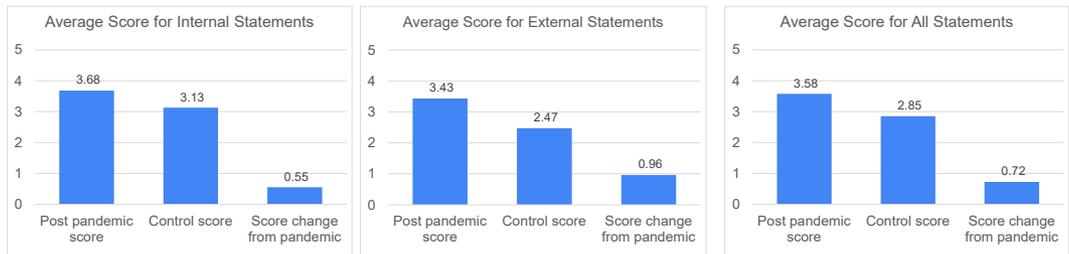
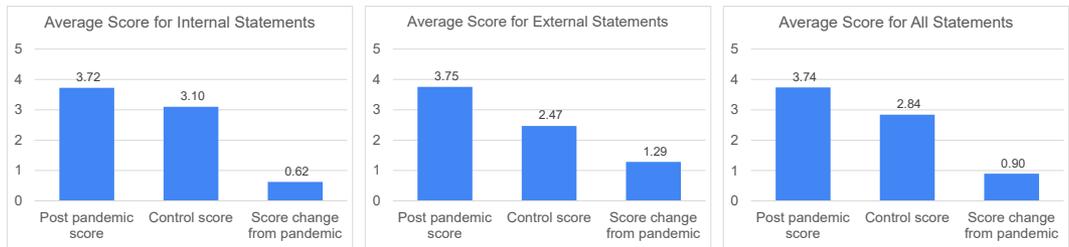
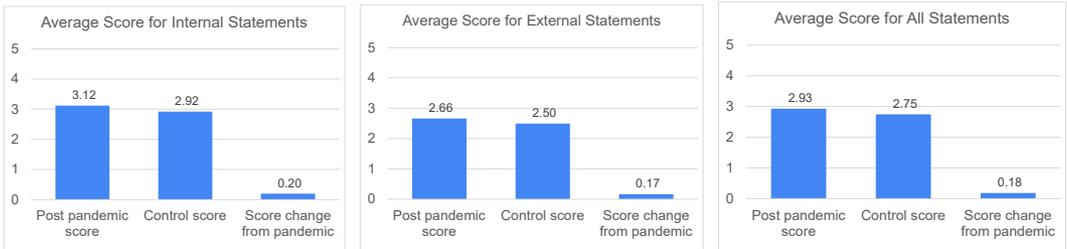


Figure 2: Seniors Living in Nursing Homes Response Average Scores



EFFECTS OF COVID-19 ON MENTAL HEALTH – SENIOR CITIZENS AND STUDENTS

Figure 3: Seniors Living at Home Response Average Scores



A t-Test was used in Microsoft Excel to confirm whether the score after the pandemic is less than or equal to the control score. Results of this test for high school student responses are shown in Tables 1, 2, and 3. For all three groupings of survey statements, the p-value is less than the significance level (α) of 0.05. As a result, the null hypothesis ($H_0: \mu_{\text{PostPandemicScore}} \leq \mu_{\text{ControlScore}}$) can be rejected. This indicates that the pandemic score is greater than the control score.

Table 1: Paired Two Sample t-Test of Post Pandemic Score Compared with Control Statements for Internal Statements for High School Students

	Summary Statistics			Results		
	Mean	Std. Deviation	Observations	df	t-value	p-value
Post pandemic score	3.68	0.76	83	82	7.3721	< .0001
Control score	3.13	0.85	83			

Table 2: Paired Two Sample t-Test of Post Pandemic Score Compared with Control Statements for External Statements for High School Students

	Summary Statistics			Results		
	Mean	Std. Deviation	Observations	df	t-value	p-value
Post pandemic score	3.43	0.78	83	82	11.8956	< .0001
Control score	2.47	0.86	83			

Table 3: Paired Two Sample t-Test of Post Pandemic Score Compared with Control Statements for All Statements for High School Students

	Summary Statistics			Results		
	Mean	Std. Deviation	Observations	df	t-value	p-value
Post pandemic score	3.58	0.72	83	82	10.4427	< .0001
Control score	2.85	0.81	83			

EFFECTS OF COVID-19 ON MENTAL HEALTH – SENIOR CITIZENS AND STUDENTS

Results of t-test for responses from seniors living in nursing homes are shown in Tables 4, 5, and 6. For all three groupings of survey statements, the p-value is less than the significance level (α) of 0.05. As a result, the null hypothesis ($H_0: \mu_{\text{PostPandemicScore}} \leq \mu_{\text{ControlScore}}$) can be rejected. This indicates that the pandemic score is greater than the control score.

Table 4: Paired Two Sample t-Test of Post Pandemic Score Compared with Control Statements for Internal Statements for Seniors in Nursing Homes

	Summary Statistics			Results		
	Mean	Std. Deviation	Observations	df	t-value	p-value
Post pandemic score	3.72	0.75	53	52	7.4672	< .0001
Control score	3.10	0.92	53			

Table 5: Paired Two Sample t-Test of Post Pandemic Score Compared with Control Statements for External Statements for Seniors in Nursing Homes

	Summary Statistics			Results		
	Mean	Std. Deviation	Observations	df	t-value	p-value
Post pandemic score	3.75	0.55	53	52	12.3993	< .0001
Control score	2.47	0.90	53			

Table 6: Paired Two Sample t-Test of Post Pandemic Score Compared with Control Statements for All Statements for Seniors in Nursing Homes

	Summary Statistics			Results		
	Mean	Std. Deviation	Observations	df	t-value	p-value
Post pandemic score	3.74	0.61	53	52	10.7182	< .0001
Control score	2.84	0.86	53			

EFFECTS OF COVID-19 ON MENTAL HEALTH – SENIOR CITIZENS AND STUDENTS

Results of t-test for responses from seniors living at home are shown in Tables 7, 8, and 9. For all three groupings of survey statements, the p-value is less than the significance level (α) of 0.05. As a result, the null hypothesis ($H_0: \mu_{\text{PostPandemicScore}} \leq \mu_{\text{ControlScore}}$) can be rejected. This indicates that the pandemic score is greater than the control score.

Table 7: Paired Two Sample t-Test of Post Pandemic Score Compared with Control Statements for Internal Statements for Seniors Living at Home

	Summary Statistics			Results		
	Mean	Std. Deviation	Observations	df	t-value	p-value
Post pandemic score	3.12	0.61	42	41	2.2196	0.0160
Control score	2.92	0.89	42			

Table 8: Paired Two Sample t-Test of Post Pandemic Score Compared with Control Statements for External Statements for Seniors Living at Home

	Summary Statistics			Results		
	Mean	Std Deviation	Observations	df	t-value	p-value
Post pandemic score	2.66	0.91	42	41	3.6933	0.0003
Control score	2.50	0.98	42			

Table 9: Paired Two Sample t-Test of Post Pandemic Score Compared with Control Statements for All Statements for Seniors Living at Home

	Summary Statistics			Results		
	Mean	Std. Deviation	Observations	df	t-value	p-value
Post pandemic score	2.93	0.66	42	41	3.1714	0.0014
Control score	2.75	0.89	42			

EFFECTS OF COVID-19 ON MENTAL HEALTH – SENIOR CITIZENS AND STUDENTS

Results of t-test for changes in internal group scores compared with changes in external group scores are shown in Tables 10, 11, and 12. For high school students and seniors living in nursing homes, the p-value is less than the significance level (α) of 0.05. As a result, the null hypothesis ($H_0: \mu_{\text{Internal}} \geq \mu_{\text{External}}$) can be rejected. This indicates that the change in internal group scores is less than the change in external group scores. For seniors living at home though, the p-value is greater than the significance level (α) of 0.05. As a result, the null hypothesis ($H_0: \mu_{\text{Internal}} \geq \mu_{\text{External}}$) cannot be rejected and therefore we cannot confirm that the change in internal group scores is less than the change in external group scores.

Table 10: Paired Two Sample t-Test of Change in Score for Internal Statements Compared with External Statements for High School Students

	Summary Statistics			Results		
	Mean	Std. Deviation	Observations	df	t-value	p-value
Internal statements	0.55	0.68	83	82	-5.8650	< .0001
External statements	0.96	0.74	83			

Table 11: Paired Two Sample t-Test of Change in Score for Internal Statements Compared with External Statements for Seniors in Nursing Homes

	Summary Statistics			Results		
	Mean	Std. Deviation	Observations	df	t-value	p-value
Internal statements	0.62	0.61	53	52	-8.2881	< .0001
External statements	1.29	0.75	53			

Table 12: Paired Two Sample t-Test of Change in Score for Internal Statements Compared with External Statements for Seniors Living at Home

	Summary Statistics			Results		
	Mean	Std. Deviation	Observations	df	t-value	p-value
Internal statements	0.20	0.58	42	41	0.3322	0.3707
External statements	0.17	0.29	42			

EFFECTS OF COVID-19 ON MENTAL HEALTH – SENIOR CITIZENS AND STUDENTS

Results of t-test for the change in scores for seniors living in nursing homes compared with seniors living at home are shown in Tables 13, 14, and 15. For all three groupings of survey statements, the p-value is less than the significance level (α) of 0.05. As a result, the null hypothesis ($H_0: \mu_{\text{NursingHome}} \leq \mu_{\text{Home}}$) can be rejected. This indicates that the change in scores for seniors living in nursing homes is greater than the change in scores for seniors living at home.

Table 13: Two Sample t-Test of Change in Score for Internal Statements for Seniors Living in Nursing Homes Compared with Seniors Living at Home

	Summary Statistics			Results		
	Mean	Std. Deviation	Observations	df	t-value	p-value
Seniors living in nursing homes	0.62	0.61	53	90	3.4948	0.0004
Seniors living at home	0.20	0.58	42			

Table 14: Two Sample t-Test of Change in Score for External Statements for Seniors Living in Nursing Homes Compared with Seniors Living at Home

	Summary Statistics			Results		
	Mean	Std. Deviation	Observations	df	t-value	p-value
Seniors living in nursing homes	1.29	0.75	53	70	9.8955	< .0001
Seniors living at home	0.17	0.29	42			

Table 15: Two Sample t-Test of Change in Score for All Statements for Seniors Living in Nursing Homes Compared with Seniors Living at Home

	Summary Statistics			Results		
	Mean	Std. Deviation	Observations	df	t-value	p-value
Seniors living in nursing homes	0.90	0.61	53	88	6.9813	< .0001
Seniors living at home	0.18	0.38	42			

EFFECTS OF COVID-19 ON MENTAL HEALTH – SENIOR CITIZENS AND STUDENTS

Results of t-test for the change in scores for high school students compared with seniors are shown in Tables 16, 17, and 18. For all three groupings of survey statements, the p-value is greater than the significance level (α) of 0.05. As a result, the null hypothesis ($H_0: \mu_{\text{HighSchoolStudents}} \leq \mu_{\text{Seniors}}$) cannot be rejected and therefore we cannot confirm that the change in scores for high school students is greater than the change in scores for seniors.

Table 16: Two Sample t-Test of Change in Score for Internal Statements for High School Students Compared with Seniors

	Summary Statistics			Results		
	Mean	Std. Deviation	Observations	df	t-value	p-value
High school students	0.55	0.68	83	168	1.1758	0.1207
Seniors	0.44	0.63	95			

Table 17: Two Sample t-Test of Change in Score for External Statements for High School Students Compared with Seniors

	Summary Statistics			Results		
	Mean	Std. Deviation	Observations	df	t-value	p-value
High school students	0.96	0.74	83	176	1.4608	0.0729
Seniors	0.79	0.82	95			

Table 18: Two Sample t-Test of Change in Score for All Statements for High School Students Compared with Seniors

	Summary Statistics			Results		
	Mean	Std. Deviation	Observations	df	t-value	p-value
High school students	0.72	0.63	83	172	1.5018	0.0675
Seniors	0.58	0.63	95			

Discussion and Conclusions

This research reveals that the COVID-19 pandemic has affected the mental health of high school students, senior citizens living in nursing homes, and senior citizens living at home, although the extent of the impact for senior citizens living at home does not appear to be significant at a cursory glance. These results indicate that online and virtual interactions cannot substitute for the benefits of physical and in-person interactions and dispel the myth that some may have about high school students adequately coping with the mental health challenges through social media and online interactions. For high school students, the increased mental health issues could also be due to increased social media activity. For senior citizens in nursing homes, in addition to reduced social interactions, the pandemic restrictions may have also curtailed healthy physical activities.

For high school students and senior citizens living in nursing homes, the impact of COVID-19 is greater for external statements than for internal statements. This is understandable as the external statements require interactions with others and these interactions were restricted due to COVID-19 due to either online classes for high school students or restrictions on visitors in nursing homes. For senior citizens living at home though, there is no statistical difference between change in internal group scores and change in external group scores. This is most likely due to these senior citizens continuing to interact with their family members and possibly with others although on a very restricted basis.

The results indicate that mental health of senior citizens living in nursing homes has been impacted to a greater extent than the mental health of those living at home. While COVID-19 imposed a variety of restrictions at different times, nursing homes had some of the most stringent restrictions especially before vaccines were rolled out. Often senior citizens were confined to their rooms with very little interaction with anyone other than a couple of caretakers. Senior citizens living at home, however, most likely could choose to interact with others especially if they were living with other family members or friends.

The results also indicate that there is no statistical difference in the extent of mental health impact on high school students than that on senior citizens. In

other words, the mental health of all groups of participants in this study have been affected by the restrictions imposed by COVID-19.

These results have implications for future research on online or virtual education, social media usage, and other forms of technologies that limit in-person interactions because of the impact that they may have on mental health. The results also indicate that it is important for public health policy to weigh the benefits of various restrictions such as lockdowns, isolation, online activities, travel bans, etc. against the detrimental effects such restrictions have on the mental health of various groups of the population.

Limitations and Future Research

Getting participants was a significant challenge due to ongoing restrictions from COVID-19. As a result, the sample size was smaller than what was originally desired. A follow-up study could be done to understand whether the results hold for other groups such as younger students and college students.

The results indicate that a lack of in-person interaction with others is a significant contributor to the effect on mental health. This could be further studied by performing similar research using participants that continued to interact with others during the pandemic such as healthcare workers and law enforcement personnel.

The study does not delve into any differences from participant gender or race. Further research would be required to understand whether the results are consistent for different genders or racial backgrounds.

All the high school participants were from the same high school and neighborhood and therefore probably have similar socioeconomic backgrounds. Similar biases may also have existed for the senior citizen sample. To eliminate any biases, students from other high schools and neighborhoods and senior citizens from other areas would be needed.

When conducting research with surveys, there may be “social response bias” or “subject bias” – respondents may not always answer honestly to present themselves in a positive manner or to support what they think the research is trying to show. This can be overcome by using a larger sample size, repeating the

research with a different sample, or measuring physiological responses, although they can be more difficult to define for mental health, because they may be less susceptible to demand bias.

Acknowledgements

I want to express my sincere thanks to Dr. Bryan Denny of Rice University for his guidance, without which I would not have been able to complete this project. I also want to express my deep appreciation to Ms. Colleen Thompson for agreeing to be my Project Advisor and consistently helping me throughout the project. I want to thank Mr. Mark Grisdale, Mr. Jason Daigle, and Ms. Kasey Nelson for approving my project as members of the IRB. Finally, I could not have done this project without the participants from Tompkins High School, Atria Senior Living, Sunrise Senior Living, OSA Heritage Homes, and seniors living at home, and I am deeply grateful to each one of them.

References

- Abbott, A. (2021). COVID's mental-health toll: how scientists are tracking a surge in depression. *Nature*. <https://www.nature.com/articles/d41586-021-00175-z>.
- American Psychological Association. (2010). Bullying: What Parents, Teachers Can Do to Stop It. <http://www.apa.org/news/press/releases/2010/04/bullying.aspx>.
- Bowen, L. (2011). Bullying may contribute to lower test scores. *Monitor on Psychology*, 42(9), 19.
- Centers for Disease Control National Center on Health Statistics. (2020). Anxiety and Depression: Household pulse survey. *CDC*. <https://www.cdc.gov/nchs/covid19/pulse/mental-health.htm>.
- Centers for Disease Control and Prevention. (2021). About Mental Health. <https://www.cdc.gov/mentalhealth/learn/index.htm>.
- CNBC. (2022). Last responders: Mental health damage from Covid could last a generation, professionals say. *CNBC*. <https://www.cnbc.com/2022/02/10/covid-pandemic-mental-health-damage-could-last-a-generation.html>.
- Gerlach, Lauren. (2021). Mental Health Among Older Adults Before and During the COVID-19 Pandemic. *University of Michigan*. <https://www.healthyagingpoll.org/reports-more/report/mental-health-among-older-adults-and-during-covid-19-pandemic>.
- Javed, B. (2020). The coronavirus (COVID-19) pandemic's impact on mental health. *Wiley Public Health Emergency Collection*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7361582/>.
- Jones, E. (2021). Impact of COVID-19 on Mental Health in Adolescents: A Systematic Review. *International Journal of Environmental Research and Public Health*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7967607/#:~:text=Globally%2C%20adolescents%20of%20varying%20backgrounds,during%20the%20COVID%2D19%20pandemic>.
- Koma, W. (2020). One in Four Older Adults Report Anxiety or Depression Amid the COVID-19 Pandemic. *KFF*. <https://www.kff.org/medicare/issue-brief/one-in-four-older-adults-report-anxiety-or-depression-amid-the-covid-19-pandemic/>.
- Lee, C. M., Cadigan, J. M., & Rhew, I. C. (2020). Increases in Loneliness Among Young Adults During the COVID-19 Pandemic and Association with Increases in Mental Health Problems. *Journal of Adolescent Health*. <https://doi.org/10.1016/j.jadohealth.2020.08.009>.
- Mendenhall, W., Beaver, R. J., & Beaver, B. M. (1995). Tests of Hypotheses for Means and Proportions. *A Brief Course in Business Statistics*, Pacific Grove, Duxbury Thomson Learning. 275-352.

National Institute of Mental Health. (2015). Chronic Illness & Mental Health: Recognizing and Treating Depression. <https://www.nimh.nih.gov/health/publications/chronic-illness-mental-health>.

National Institute of Mental Health. (2021). Caring for Your Mental Health. <https://www.nimh.nih.gov/health/topics/caring-for-your-mental-health>.

Panchal, N., Kamal, R., Cox, C., & Garfield, R. (2021). The Implications of COVID-19 for Mental Health and Substance Use. *KFF*. <https://www.kff.org/coronavirus-covid-19/issue-brief/the-implications-of-covid-19-for-mental-health-and-substance-use/>.

Reuters. (2022). Mental health issues in kids rose during pandemic; awareness and use of COVID treatments is low. <https://www.reuters.com/business/healthcare-pharmaceuticals/mental-health-issues-kids-rose-during-pandemic-awareness-use-covid-treatments-is-2022-04-28/>.

Spielman, R. M. (2017). Therapy and Treatment. *Psychology, Houston, Openstax*, 601-634.

U.S. Department of Health & Human Services. (2021). Mental Health and Mental Disorders. <https://www.healthypeople.gov/2020/topics-objectives/topic/mental-health-and-mental-disorders#1>.

U.S. Department of Health & Human Services. (2021). U.S. Surgeon General Issues Advisory on Youth Mental Health Crisis Further Exposed by COVID-19 Pandemic. *HHS*. <https://www.hhs.gov/about/news/2021/12/07/us-surgeon-general-issues-advisory-on-youth-mental-health-crisis-further-exposed-by-covid-19-pandemic.html>.

Vahia, I. (2020). Older Adults and the Mental Health Effects of COVID-19. *JAMA*. <https://jamanetwork.com/journals/jama/fullarticle/2773479#:~:text=Older%20adults%2C%20compared%20with%20other,reports%20from%20high%2Dincome%20countries>.

Washington State Department of Health. (2021). Statewide High-Level Analysis of Forecasted Behavioral Health Impacts from COVID-19. <https://doh.wa.gov/sites/default/files/legacy/Documents/1600/coronavirus//821-103-BHForecastSummary-20210831.pdf>.

World Health Organization. (2018). Mental Health: Strengthening Our Response. <https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response>.

World Health Organization. (2021). Adolescent Mental Health. <https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health>.

World Health Organization. (2022). COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide. <https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide>.

Appendix A: Survey Statements for High School Students

Please read each statement and rate how much the statement relates to you on a scale of 1-5, 1 being the least relevant and 5 being the most relevant. Compared to before the start of the COVID-19 pandemic in March 2020:

Internal Statements:

During the pandemic I prefer to stay at home more and do not go out.

During the pandemic I believe I criticize myself more for my mistakes.

During the pandemic when I do not feel my best, the amount of sleep I get each night significantly changes.

During the pandemic when I am feeling down, the amount of food I eat significantly changes.

During the pandemic I tend to experience more negative emotions when receiving criticism.

During the pandemic I have become more introverted and prefer to be alone more often.

During the pandemic I feel more depressed.

During the pandemic I feel more anxious.

During the pandemic I require more focus and effort on certain academic areas.

During the pandemic I feel more pessimistic towards people and daily activities and events.

During the pandemic I feel more stressed.

External Statements:

During the pandemic I have been spending very little time with friends and family by choice.

During the pandemic, I limit my time spent with others more to get my work done, even if I do not want to do so.

During the pandemic when I meet someone new, I feel more scared or nervous.

During the pandemic when I am feeling down, I am more likely to surround myself by people who feel the same way.

During the pandemic when I interact with others, I feel that they do not understand the message of what I am trying to communicate to them the way I understand myself.

During the pandemic my enjoyment in clubs or interactive/fun activities has decreased.

During the pandemic my number of friends has

decreased.

During the pandemic I feel lonelier using online classes and/or platforms than attending classes and/or activities in person.

Control Statements:

I prefer to stay at home and do not go out.

I criticize myself for my mistakes.

When I do not feel my best, the amount of sleep I get each night significantly changes.

When I am feeling down, the amount of food I eat significantly changes.

I tend to experience negative emotions when receiving criticism.

I am introverted and prefer to be alone.

I feel depressed.

I feel anxious.

I require more focus and effort on certain academic areas.

I feel pessimistic towards people and daily activities and events.

I am always stressed.

I spend little time with family and friends.

When I have a lot of things to do, I limit my time spent with others to get my work done, even if I do not want to do so.

When I meet someone new, I feel scared or nervous.

When I am feeling down, I surround myself by people who feel the same way.

When I interact with others, I feel that they do not understand the message of what I am trying to communicate to them the way I understand myself.

I have little enjoyment in clubs and/or interactive activities.

I have very few friends.

I feel lonely.

Appendix B: Survey Statements for Senior Citizens

Please read each statement and rate how much the statement relates to you on a scale of 1-5, 1 being the least relevant and 5 being the most relevant. Compared to before the start of the COVID-19 pandemic in March 2020:

Internal Statements:

During the pandemic I prefer to stay at home more and do not go out.

During the pandemic I believe I criticize myself more for my mistakes.

During the pandemic when I do not feel my best, the amount of sleep I get each night significantly changes.

During the pandemic when I am feeling down, the amount of food I eat significantly changes.

During the pandemic I tend to experience more negative emotions when receiving criticism.

During the pandemic I have become more introverted and prefer to be alone more often.

During the pandemic I feel more depressed.

During the pandemic I feel more anxious.

During the pandemic I feel more pessimistic towards people and daily activities and events.

During the pandemic I feel more stressed.

External Statements:

During the pandemic I have been spending very little time with friends and family by choice.

During the pandemic, I limit my time spent with others more to get my work done, even if I do not want to do so.

During the pandemic when I meet someone new, I feel more scared or nervous.

During the pandemic when I am feeling down, I am more likely to surround myself by people who feel the same way.

During the pandemic when I interact with others, I feel that they do not understand the message of what I am trying to communicate to them the way I understand myself.

During the pandemic my enjoyment in interactive/fun activities has decreased.

During the pandemic my number of friends has decreased.

Control Statements:

I prefer to stay at home and do not go out.

I criticize myself for my mistakes.

When I do not feel my best, the amount of sleep I get each night significantly changes.

When I am feeling down, the amount of food I eat significantly changes.

I tend to experience negative emotions when receiving criticism.

I am introverted and prefer to be alone.

I feel depressed.

I feel anxious.

I feel pessimistic towards people and daily activities and events.

I am always stressed.

I spend little time with family and friends.

When I have a lot of things to do, I limit my time spent with others to get my work done, even if I do not want to do so.

When I meet someone new, I feel scared or nervous.

When I am feeling down, I surround myself by people who feel the same way.

When I interact with others, I feel that they do not understand the message of what I am trying to communicate to them the way I understand myself.

I have little enjoyment in interactive/fun activities.

I have very few friends.

each statement relates to you.

Time commitments: Complete participation time is expected to take no more than 10 minutes.

Potential risks of study: There is minimal risk since humans are the subject of the experiment, but the risks should not be significant.

Benefits: Better understanding of yourself, opportunity to network with AP Psychology teacher, opportunity to contribute towards an understanding of effects of COVID-19 on mental health on high school students and senior citizens.

How confidentiality will be maintained: No names will be collected, recorded, or published. Participants will only be identified by random numbers. All data will be kept anonymous.

Right to Refuse or Withdraw: Please note that participation in this study is completely voluntary. You may decline to participate at any time without any negative consequences. Even after agreeing to participate, you may stop participating at any time without any repercussions.

Teacher Sponsor to contact if you have any questions about the study: Ms. Colleen Thompson, email: colleenthompson@katyisd.org.

By signing this form, you are attesting that you have read and understood the information above and you freely give your consent/assent to participate or give permission for your child to participate.

Participant Consent

Date: _____

Signature: _____

Name: _____

Parent/Guardian Consent

Date: _____

Signature: _____

Name: _____

Appendix C: Consent Form

Student Researcher: Uma Kamath

Title of Project: The Effects of COVID-19 on Mental Health of Senior Citizens and High School Students

You are being asked for your voluntary participation in this research project. Please read the following information about the project before providing your consent in the appropriate area below.

Purpose of the project: To understand the extent of the effects of COVID-19 on the mental health of senior citizens and high school students and the contributing factors.

If you participate, you will be asked to: Complete an anonymous survey that requires you to read a list of statements and rate on a scale of 1-5 how much

Immigrant Parents and Academic Success: Generational Status and Race on Academic Achievement

Breanna Villarreal

As the United States continues to get more diverse, the percentage of American students who are children or grandchildren of immigrants has increased. A survey was used to discover how the immigrant academic advantage affects both Black and non-Black Latino students in New York State. Second- and third-generation students between the ages of 14 and 18 completed a survey that assessed feelings of cultural identity, perceived and real academic success, and academic motivations. The results confirmed the widely accepted second-generation advantage but found that Latino students may not be at an academic disadvantage as much of the previous research concluded. Black Latino and non-Black Latino students had no significant academic differences, meaning both groups were equally benefited by the academic advantage. This research can lead to new conversations regarding Black Latino students and their academic attitudes, however, as there was a significant correlation between race and academic confidence.

Keywords: Children of immigrants, academic advantage, Black Latino

Introduction

The United States continues to become a more diverse country, with over 44.9 million immigrants in the U.S. as of 2019 (“Immigrants”, 2021). This increasing immigrant population has led to a growing percentage of adolescents who are children or grandchildren of immigrants. For the purposes of this research, these adolescents will all be referred to as children of immigrant parents (CIPs). Most previous research agrees that CIPs have an advantage academically over their peers with native-born parents. This study attempted to better understand how the academic advantage affects underrepresented communities, specifically Latino students of various racial and ethnic backgrounds. To this end, a survey was conducted using categorical demographic questions as well as

a series of Likert scale agreement questions and the results were analyzed through the statistical analysis program JASP.

Review of Literature

CIPs have distinctly different cultural traditions and family structures that affect how they learn and succeed academically. This has led many researchers to study the academic effects of having immigrant parents on American adolescents. Many have concluded that overall, CIPs are at an advantage academically and are more likely to succeed than their non-immigrant-related counterparts. Some have hypothesized that generational, ethnic, and racial differences can affect this advantage differently, and found

that certain groups have a greater advantage. No study, however, has combined the factors of generation, ethnicity, and race while focusing on Latinos of different backgrounds. This study will research how Latinos also identifying as Black Latino, African American, Afro-Latino, or Black, are affected academically by their CIP status as opposed to non-Black Latinos. All participants who identify as Black Latino, African American, Afro-Latino, or Black will be referred to as Black Latinos in this study.

Body of Knowledge

Immigrant Advantage

In a study published by the American Educational Research Association, psychologists and psychiatrists Duong et al. (2016) found that the immigrant advantage that had been hypothesized and researched in previous studies does exist, but that it is stronger among some subgroups. They found that first- and second-generation students have a greater advantage over third- and later-generation students and agreed with Qian et al. that the advantage was highest among Asian immigrants (Duong et al, 2016; Qian et al., 2018). Others agree that second-generation students have an advantage over other CIPs (Fuligni, 1997; Keller & Tillman, 2008; Mwangi et al., 2017; Thomas, 2009). A 2008 study built on the idea that CIPs have an advantage over students with native-born parents and found that immigrant parents tend to have higher expectations of their children in terms of success and higher education compared to native born parents, which leads to greater success for their children (Keller & Tillman, 2008). This study also agreed that second-generation students have a greater advantage over third- or later-generations, due to second-generation students being both strongly connected to their culture through their parents and connected to American culture and language because of their own native-born status (Keller & Tillman, 2008). Raleigh and Kao (2010) also found that immigrant minority parents often have “higher likelihoods of forming and maintaining college aspirations for their children compared to native born parents” (p. 17). Mwangi et al. (2017) and Fuligni (1997) agree that parental influence and placement of importance on education

is a strong academic motivator for CIPs and contributes to their advantage over non-CIPs. While Fuligni (1997) found that parental influence was much more important than socioeconomic status, Kim, Mok, and Seidel (2020) found that parental influence had little effect on academic achievement and socioeconomic status was a much more important factor. According to Duong (2016), Harris (2008), Ibarra (2004), and Rong (1992), the observed immigrant advantage is greater in Asian students than Latino students. Similarly, Duong, et al. (2016) found that Black students have an advantage over Latino students, and Ibarra (2004) found that white students have a greater advantage over Latino students as well. Harris, Jamison, and Trujillo (2008) found that socioeconomic status was a reason for this disadvantage, while Ibarra (2004) cited family structure as a cause. A 2006 study found that Latino students are at a disadvantage academically compared with other CIPs and need more academic support to succeed (Alfaro, et al., 2006). These students are more likely to live in impoverished areas than their white counterparts and are the most likely of all demographic groups to drop out of high school (Alfaro et al., 2006). This makes it much more difficult for Latino students to thrive academically. The different structure of the typical Latino family can leave students at a disadvantage and in need of more support from teachers and others in the classroom or academic sphere (Alfaro et al., 2006). Though these students can, and often do, still benefit from the immigrant academic advantage, they require more support to do so due to the obstacles they face in and outside the classroom. Urdan and Munoz agree that there is some immigrant advantage, but believe it has a smaller effect than previous research has stated; however, they adhere to the theory also discussed by Keller and Tillman claiming that there is a strong correlation between academic motivation and cultural identity (Keller & Tillman, 2008; Urdan & Munoz, 2012).

Black Latino Disadvantage

Within the group of Latino CIPs, there are subsets of different racial and ethnic backgrounds. Black Latinos, specifically, are overlooked in research and often grouped with non-Black Latinos or not studied at all. As noted in a 2020 study from Harvard University, the U.S. Census, and most, if not all, research including

Latino subjects take a “pan-ethnic’ view of the Latino community”, meaning that Latinos with different backgrounds and experiences are studied as one homogenous group (Godoy Peñas, 2020, p. 4). Most researchers on the subject believe that acknowledging Black Latinos as both Black and Latino, instead of overlooking one part of their identity, is most beneficial. Nolasco agrees with Burgos and Rivera that failure to recognize the intersection of Black and Latino identities makes it harder for Black Latinos to have positive cultural identities and contributes to many of the hardships they face, like colorism and racism, inside and outside the Latino community (Burgos & Rivera, 2009; Nolasco, 2020). Garcia-Louis and Nolasco agree that Black Latinos face unique hardships because they are discriminated against in both their own communities and other spaces (Garcia-Louis, 2018; Nolasco 2020).

Call to Research

Though there is a considerable amount of research on the topic of the immigrant academic advantage, there are many different variables affecting it that require further investigation. Previous studies have considered generational status, cultural identity, and racial and ethnic background, but most have focused on one, not the combined effect of these different factors. Some have concluded that generation is the most important factor and second-generation students have an advantage over all others. Others found that a student’s cultural identity and connection to their background is most important to increasing their advantage or that their racial or ethnic background most influenced their academic success. Most researchers agree that Black and Asian students have a greater advantage than Latino students, but no study has considered how Black Latino students are affected by the immigrant advantage. Black Latinos face many hardships, which presumably puts them at a disadvantage, but, according to previous studies, their Black identity could give them an advantage. This uncertainty caused by a gap in knowledge in the field leads to the focus of this research, which will attempt to determine how being a CIP affects academic success and motivations for Black Latino students versus non-Black Latino students. I hypothesize that because Black Latinos

are more often perceived as Black, their academic advantage will most closely resemble that of non-Latino Black students, giving them an advantage over non-Black Latinos.

Method Alignment

To gather data for my research, I chose to survey Latino students between the ages of 14 and 18. Creating a new survey rather than using a pre-existing one allowed me to collect more specific data and tailor the questions to address each of the variables. In a similar study of university students who were CIPs, a survey was used to connect cultural identity to academic identity and motivation (Urdañ & Muñoz, 2012). Fuligni and Alfaro also used surveys to collect similar data among CIPs enrolled in high school and found more clear and reliable results than researchers using other methods with similar variables (Alfaro et al., 2006; Fuligni, 1997). My survey first asked participants their age, to make sure they were in the targeted age range of 14-18. Other studies in the same field focused on either a range of younger middle and high school students, from sixth to tenth grade (Alfaro et al., 2006; Fuligni, 1997) or university students from 18-25 (Mwangi et al., 2017; Urdañ & Muñoz, 2012). My research created a middle range that has not been studied extensively and is beneficial because many students in this age range are becoming more serious about academics and possibly applying to college soon, so they are the best to study on academic motivation and success. I then asked if the participant identified as Hispanic or Latino in order to again make sure they were part of the target demographic of my research. Next, my survey asked participants whether they, their parents, or their grandparents are immigrants and from what country in order to categorize participants by generational status and ethnic background. Participants could indicate if one parent/grandparent or two parents/grandparents are immigrants and if one of the options for two immigrant relatives was chosen, they were able to input two or more countries. Many Black Latinos are from mixed backgrounds which contributes to their lack of representation in research because they are often forced into one category, splitting their identity into either Latino or Black, not both. Then, my survey asked students to

choose a range that their GPA fits in and evaluate their feelings on a series of statements about their academic success and academic motivations using a 4-point Likert scale with anchors at Strongly Disagree and Strongly Agree. The Likert scale has been used in data collection for similar studies and allowed responses to be categorized more easily because there was no ambiguity. I used a 4-point scale without a neutral option because selection of the neutral option would not add any data and could hinder my data collection. Participants were then asked if they identify as Black/African American/Afro-Latino so I was able to separate results of Black Latinos and non-Black Latinos and compare them. All studies that considered race or ethnicity as a factor asked questions regarding broad racial or ethnic identification as well as country of origin, to get the most accurate data on how these factors affect academic success and motivation (Fulgini, 1997; Harris et al., 2008; Rong & Grant, 1992; Urdan & Munoz, 2012). Finally, participants were asked to evaluate their feelings on cultural identity and belonging using the same Likert scale. Participants could be asked one to two additional questions based on how many countries they input in the question inquiring about their cultural background and whether or not they identify as Black. These questions were used to assess participants' cultural identity and connection to their ethnic background. Some studies in the field have shown a correlation between cultural identity and academic success and other studies have hypothesized that Black Latinos may have a weaker cultural identity and sense of belonging due to their racial and ethnic identities often being separated and feelings of exclusion from both communities (Keller & Tillman, 2008; Nolasco, 2020; Urdan & Munoz, 2012). Asking participants questions on this can compare the cultural identities of non-Black Latinos with those of Black Latinos as well as see how that affects academic motivation and success. These questions were asked last so that participants' answers and thoughts that arise from this topic do not affect their answers on academic motivation or success.

Methods

To conduct my research, I first created a survey through Microsoft Forms. My research tested new variables that had not been studied before, so I created my own questions. Some questions, however, were inspired by previous studies that dealt with similar research questions and demographics. I distributed the survey online in an attempt to gather the greatest number of participants possible while also adhering to all COVID-19 safety guidelines as well as through people I was familiar with so that I could make sure the data I gathered was as reliable as possible. I collected responses from students ages 14 to 18 who self-identified as Hispanic or Latino and left the survey open to responses for three weeks so I could gather as much data as possible in the time I had available. Before beginning the survey, participants were assured that their anonymity and privacy would be protected, as I did not collect names and results were stored on a password-protected computer. After reading a short consent form and confirming their willingness to participate in the study, participants were asked questions that allowed me to categorize them as well as confirm that they were within the target demographic. Those initial questions split participants into the first two categories: second- or third-generation students. Participants were then categorized again based on whether they had one immigrant direct relative or two or more immigrant direct relatives. Participants with one relative were asked the country of origin for that immigrant, while participants with more than one relative were asked to input all the countries of origin for each immigrant relative. Next, both groups were asked a series of questions about their academic success and motivations, including their GPA and college aspirations. To further categorize my participants based on the variables my research considered, the next question asked participants if they identified as Black, African American, Afro-Latino, any combination of those identities, or none of the identities. Participants who indicated only one immigrant parent or grandparent and did not identify as Black, African American, or Afro-Latino were asked to rate their level of agreement with three statements regarding their cultural background. Participants who indicated one immigrant parent or grandparent and did identify as Black, African American, or Afro-Latino

IMMIGRANT PARENTS AND ACADEMIC SUCCESS

were asked to rate their level of agreement with four statements regarding their cultural background, the same three statements as the previous group along with an additional statement concerning their feelings of belonging in Black spaces or among Black people. The second broad group of participants, those who had two or more immigrant parents or grandparents, were also split between those who identified as Black, African American, or Afro-Latino and those who did not. Participants who did not connect with one of these identities were presented with four statements, the same base three as the two previous groups, along with an additional statement about their feeling of connection to each part of their cultural identity. The participants in this second group who did identify as Black, African American, or Afro-Latino were presented with five statements, the same four as the previous group and the additional statement concerning their feelings of belonging in Black spaces and among other Black people. I constructed the survey in such a way that questions addressing cultural identity and race were last, so that they would not affect participants' answers to academic evaluation questions or cause them to contemplate their answers too deeply to the basic academic questions and skew or corrupt the results of my research. After three weeks I closed the survey to responses and analyzed the responses I had collected. I created a spreadsheet with the responses to organize and analyze the information more easily. Any responses that were not complete, like participants who were not Hispanic or Latino, participants outside the 14 to 18 age range, or participants who were not second- or third-generation immigrants, were deleted so that I had a table of organized, usable data. To make my data easier to analyze, I changed all responses to Likert scale questions from the original response scale of strongly disagree to strongly agree to numerical values on a scale of one to four. This made it possible for me to input my collected data into a statistical analysis program known as JASP. I ran ANOVA tests on my data to find significance using questions like "What is your GPA?" or Likert scale agreement to statements like "I do well in school" and "I feel a strong connection to my cultural background" as independent variables to come to conclusions on the relationship between race or generational status and academic success, attitudes, or motivation. After running these statistical analysis tests, I looked

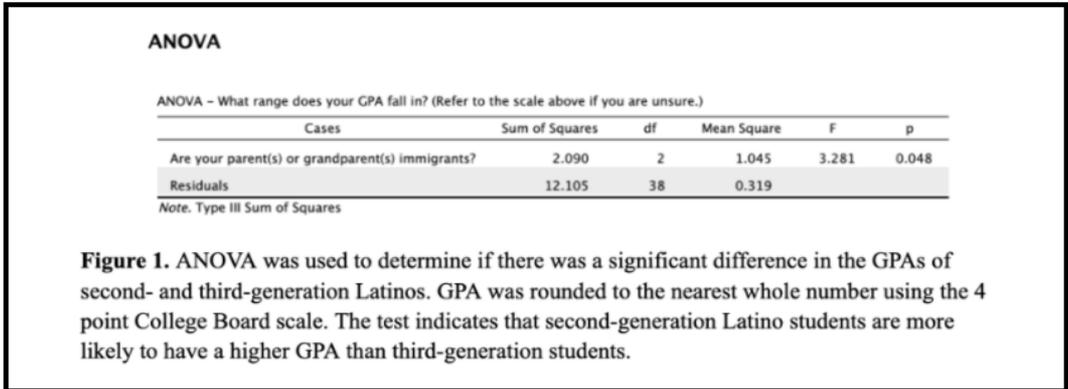
at p-values of the results as well as comparisons of the mean values between groups to determine which results were statistically significant or relevant enough to be included in the discussion of my results. A p-value of 0.05 or less was used as a marker for statistical significance and any tests resulting in p-values less than that were included and discussed. Tests where I expected a marker of significance and one was not present or where I expected a different outcome were also included when the results yielded were surprising or otherwise important or relevant. I also looked at the mean values given for each group and discussed those results, even if they were not shown to be statistically significant, because the substantial difference in answers of different racial groups was important to my research question and still needed to be examined. After selecting which results were significant or worthy of further discussion and analysis, I represented these findings through two-way ANOVA charts and descriptive charts containing the mean values of different independent variables, like GPA or Likert scale agreement, for the different participant categories, second- versus third-generation and Black versus non-Black Latino.

Results, Findings, and Analysis

I collected 64 responses from my survey. Of those responses, four were discarded because they were outside of the target age range, seventeen were discarded because they were not Hispanic or Latino, and two were eliminated because they did not have parents or grandparents who were immigrants. In the end, 41 results were analyzed. A combination of statistical analysis tests using the analysis software JASP and comparisons of the means of different data sets were used to come to conclusions about the correlations between the different variables recorded in my survey.

I first ran an ANOVA test through JASP to see if there was any correlation between the generational status of Latino students and their average GPAs. As seen in Figure 1, there is a correlation between the two factors, with the p-value being 0.048, a marker of statistical significance. Further analysis of the mean values and post-hoc tests showed that second-generation students, on average, had higher GPAs than third-

IMMIGRANT PARENTS AND ACADEMIC SUCCESS



generation students. Third-generation students had a mean GPA of 3.3 on a 4-point scale, while second-generation students had a mean GPA of 3.613. This conclusion aligns with my hypothesis as well as the conclusions of previous research, while also strengthening the claim against some researchers who say that there is a possible second generation decline in academic success or doubt the second-generation advantage. While the test did show a strong correlation be-

tween generation and GPA, no correlation was found between GPA and the number of immigrant direct relatives one had. Students with only one immigrant parent or grandparent showed no significant or consistent difference in mean GPA from students from the same generation with two immigrant parents or grandparents. There has not been much research on this specific variable, however, and further research with a larger sample size and more in-depth informa-

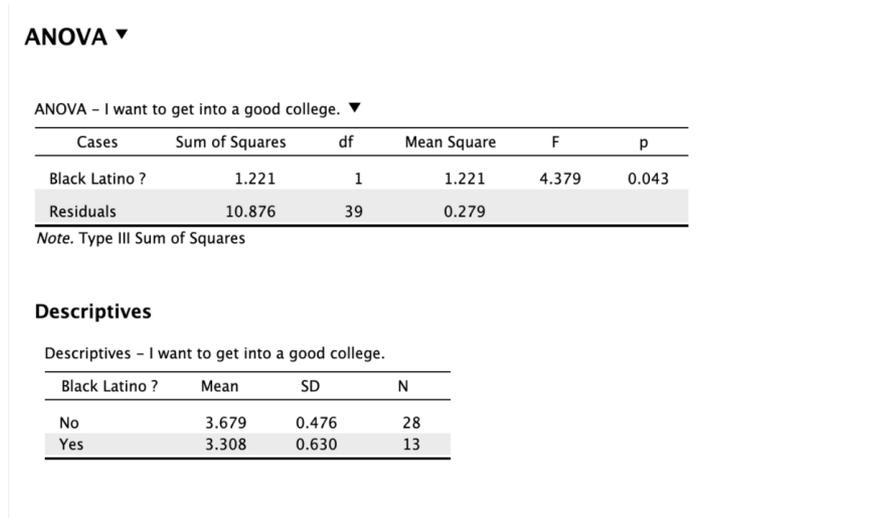


Figure 2. ANOVA was used to determine if there was a significant difference in college aspirations of Black and non-Black Latino students. College aspirations were measured using a 4-point Likert scale ranking agreement to the statement “I want to get into a good college.”. This test indicates that there is a significant correlation between the two factors.

ANOVA ▼

ANOVA – I get good grades.

Cases	Sum of Squares	df	Mean Square	F	p
Black Latino ?	0.003	1	0.003	0.009	0.923
Residuals	13.558	39	0.348		

Note. Type III Sum of Squares

Descriptives

Descriptives – I get good grades.

Black Latino ?	Mean	SD	N
No	3.250	0.585	28
Yes	3.231	0.599	13

Figure 3. ANOVA was used to determine whether there was a significant difference in the perceived academic success of Black and non-Black Latino students. The perceived academic success of participants was measured using a 4-point Likert scale ranking agreement to the statement “I get good grades.”

tion may yield different, more statistically significant, results.

An ANOVA test was used again to find the significance of the correlation, if there was any, between college aspirations and the race of Latino students. Based on previous research, I hypothesized that there might be some differences between the academic successes and attitudes of Black and non-Black Latino students. One of the measures of academic attitudes and motivations in my survey was the level of agreement with the statement “I want to get into a good college” ranked on a 4-point Likert scale with anchors at “Strongly Disagree” (1) and “Strongly Agree” (4). The ANOVA test showed a correlation between the two factors, with a p-value of 0.043, which is less than 0.05, showing statistical significance. Further analysis of the mean values and post-hoc comparisons showed that Black Latino students were less likely to aspire to go to what they perceived as a good college than non-Black Latino students. This result did not align with the findings of previous studies, as Black students were seen to be affected by the immigrant academic advantage to a greater extent and therefore should have been equally as likely, if not more, to aspire to attend a “good college” as their non-Black Latino

counterparts. This also does not directly align with the other results of my survey, including the lack of correlation between GPA and race among my participants.

An ANOVA test was run using the independent variable of race, Black versus non-Black, and the dependent variable of perceived academic success measured here using a 4-point Likert scale which ranked participants’ level of agreement to the statement “I get good grades.” Other Likert scale statements measured levels of perceived academic success, including “I do well in school” and yielded similar results when analyzed in JASP. The test found that there was no correlation between the two factors, with the p-value of 0.923 indicating statistical insignificance. This result, while showing no correlation between the variables of my research, is still important because it adds to the body of knowledge on Latino students and the immigrant advantage. According to these findings, Black Latino students do not seem to be at an academic disadvantage, as some researchers hypothesized, and have no significant difference in academic success as measured by how good they think their grades are. This result could also be interpreted as opposing previous conclusions and may prompt further research. As referenced in the review of literature, most researchers in

ANOVA ▼

ANOVA – What range does your GPA fall in? (Refer to the scale above if you are unsure.)

Cases	Sum of Squares	df	Mean Square	F	p
Black Latino ?	0.440	1	0.440	1.246	0.271
Residuals	13.755	39	0.353		

Note. Type III Sum of Squares

Figure 4. An ANOVA test was conducted to find if there was a significant difference in the GPA’s of Black Latino students and non-Black Latino students. GPA was measured using the College Board 4-point scale and was rounded to the nearest whole number.

the field have come to the consensus that Latino students benefit the least from the immigrant academic advantage and may be at a disadvantage from other immigrant racial groups, including Black immigrants. The mean values of my data show that non-Black Latino students and Black Latino students do not differ significantly in their perceived academic success, and tend to have values close to four, the highest level of agreement allowed to the statement. This finding could show that Latinos are not at as much of a disadvantage as previously thought. This hypothesis could be better corroborated with further analysis of multiple racial and ethnic immigrant groups.

Another measure of academic success was GPA. I ran an ANOVA test using the variables of GPA and race of participants to find if there was any significant difference in the GPAs of Black and non-Black students and to see if the variables were significantly correlated in any way. The test resulted in similar findings as the previous measure of academic success; however, where Figure 3 measured perceived success, the data in Figure 4 was thought to be more reliable since it was a standardized measure of academic success. The results were still self-reported however, so even though GPA is a more standardized measure of success, it cannot be considered significantly more reliable data than the previous test. All answers did seem reasonable and aligned with the participants’ other answers to questions in the survey, so they were generally considered truthful and reliable results. While this test gave a lower p-value than the test in Figure 3, the p-value of 0.271 showed statistical insignificance. This result means that there is no significant difference between the GPAs of Black and non-Black Latino students, and again discredits the claim that

Black Latino students are at a disadvantage academically. This analysis, along with the previous test, shows that Black and Latino students perform very similarly academically. Though this cannot prove an immigrant academic advantage, because there are no results from a non-immigrant control, it does corroborate previous findings seeing as both groups have relatively high GPAs, which is indicative of an academic advantage among immigrants.

The results of my research seem to conflict, in that both groups are equally as likely to do well in school and have a high GPA, but Black Latinos are less likely to have high college aspirations. There was no significant difference in the academic success of these two groups, but there may be a difference in the attitudes surrounding academia. Though there was no statistically significant correlation between feelings of connection to cultural background and race (Figure 5), the difference in measured success and aspirations of Black Latino students could be tied to feelings of inadequacy or discomfort in their identity, which have been discussed and hypothesized in previous research. In addition, though the p-value in Figure 5 was above the threshold for statistical significance, this could have been influenced by the small sample size. The mean values of the Likert scale scores for this category for Black and non-Black participants differed by more than 0.3, a greater difference than most other categories in which differences were deemed insignificant. On average, Black Latinos felt less connected to their cultural background, which is consistent with conclusions made within the body of knowledge of research on this subject, which claimed that many Black Latinos may feel less connected to their cultural background because of the monolithic view of Latinos

IMMIGRANT PARENTS AND ACADEMIC SUCCESS

ANOVA

ANOVA – I feel a strong connection to my cultural background.

Cases	Sum of Squares	df	Mean Square	F	p
Black Latino ?	0.949	1	0.949	1.669	0.204
Residuals	22.173	39	0.569		

Note. Type III Sum of Squares

Descriptives

Descriptives – I feel a strong connection to my cultural background.

Black Latino ?	Mean	SD	N
No	3.250	0.752	28
Yes	2.923	0.760	13

Figure 5. An ANOVA test was used to find if there was a significant difference between feelings of connection to one’s cultural background, which was measured on a 4-point Likert scale, and the race of Latino students.

and because many Black Latinos come from multicultural or multiethnic backgrounds. These results could prompt further research seeking to answer this specific question and provide clearer data and answers.

From my analysis, Black and non-Black Latino students have no significant academic differences in motivations or success, disproving my initial hypothesis. This can further inform the parents and teachers of these students on how to best set them up for success. My results also add to the evidence of the second-generation advantage, which is again useful to teachers and parents because it can inform the encouragement they give students as well as how much help they may need to succeed. Finally, though the conclusion is not certain, my findings hint at a possible difference in academic attitudes and confidence, where Black Latino students may be at a disadvantage. These findings could provide a basis to be elaborated on in further research and could inform the parents and teachers of Black Latino students on how to best support them for maximum success.

Limitations

A main limitation of my research was that all academic data was self-reported, and that many survey questions used to analyze academic success were subjective, not standardized. The Likert scale agreement rankings, especially to statements like “I do well in school,” were opinion based, meaning the data I collected may not have been perfectly accurate because there was no explicit definition for what each ranking meant. Another limitation I encountered was that I surveyed a smaller sample size than I would have liked, and the split between Black and non-Black Latinos was not even. This may have been avoided if I had focused on dispersing my survey in an area with a larger and more diverse Latino population, such as an urban area. Finally, though my participants were from a variety of schools, almost all resided in middle-class areas in New York state and were mostly from Long Island and the New York City area; therefore, the findings may not apply to all Latino CIPs.

Implications

The findings of my research can inform how teachers and parents of Latino students support and encourage them throughout the learning process knowing what advantages or disadvantages students may be at depending on their background. Latino CIPs are not inherently at a disadvantage, as some previously believed, and need to be encouraged to the same degree as other immigrant groups that benefit from the academic advantage. It is also important for parents and teachers to note that while Black Latino students are not at a disadvantage academically, they may still need increased and specialized support to reach levels of academic confidence comparable to those of other immigrant student groups. The results of my research open a conversation for new research on Black Latino students specifically and their academic performance and confidence as well as further research on all CIPs to reevaluate which, if any racial or ethnic groups, are truly at an academic disadvantage.

Conclusion

Race has very little, if any, influence on high school academic success of second- and third-generation Latino students. Black Latinos are just as academically successful as their non-Black counterparts, even with the considerable challenges they face. Similarly, their Black identity does nothing to offset the challenges they may face as Latinos, so they do not benefit more from the academic advantage. Generation continues to be a strong indicator of success, with my results reinforcing the second-generation advantage. Increased efforts to connect with their cultural background as well as increased motivation from their second-generation parents may help third-generation students overcome this disadvantage. The academic confidence and attitudes of students, however, may be affected by their race, as Black Latino students seemed to have less confidence or less ambitious academic aspirations. This could be an effect of their more complex cultural identity and the difficulties they may feel creating a sense of belonging.

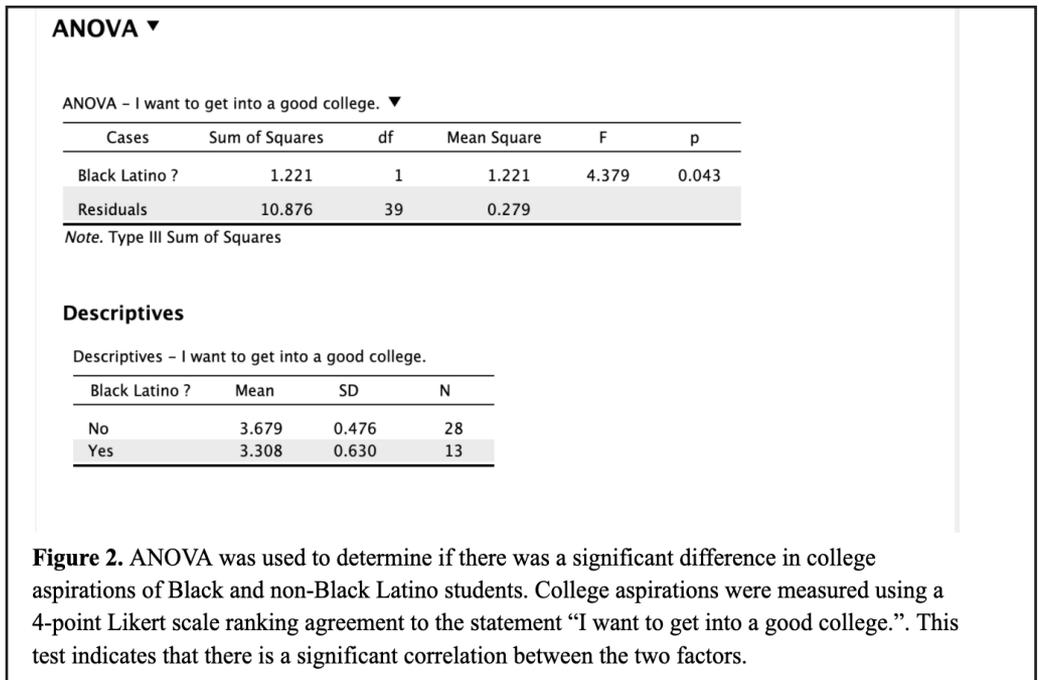
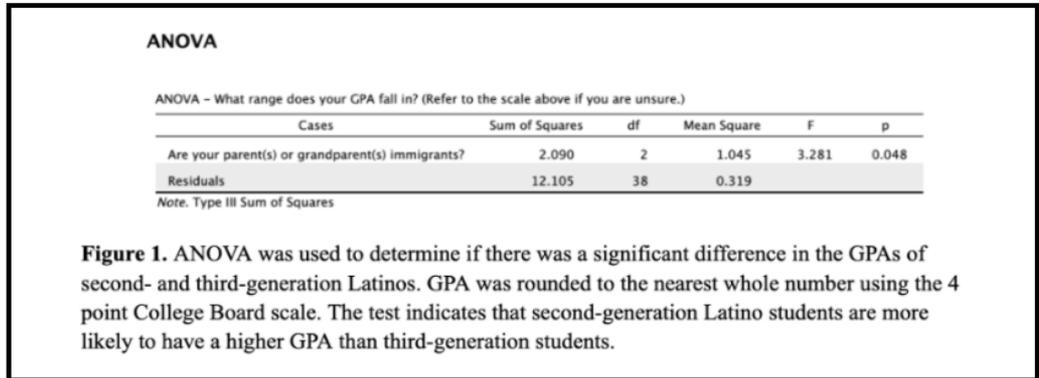
Future Directions

Future researchers should study larger populations and focus on urban areas, in order to increase the possibility of finding statistical significance in results and survey larger numbers of Black Latinos, as well as Latinos from a wider range of countries or ethnic backgrounds. Additionally, many of my variables may not have shown statistical significance because I measured GPA in a range, grouping together data and making it less specific. Future researchers may want to include every GPA on the College Board scale without rounding to get more precise data and results.

References

- Alfaro, E. C., Umaña-Taylor, A. J., & Bámaca, M. Y. (2006). The Influence of Academic Support on Latino Adolescents' Academic Motivation. *Family Relations*, 55(3), 279–291. <http://www.jstor.org/stable/40005313>
- Burgos, G., & Rivera, F. (2009). The (In)Significance of Race and Discrimination among Latino Youth: The Case of Depressive Symptoms. *Sociological Focus*, 42(2), 152–171. <https://doi.org/10.1080/00380237.2009.10571348>
- Duong, M. T., Badaly, D., Liu, F. F., Schwartz, D., & McCarty, C. A. (2016). Generational Differences in Academic Achievement Among Immigrant Youths: A Meta-Analytic Review. *Review of Educational Research*, 86(1), 3–41. <http://www.jstor.org/stable/24752868>
- Fulgini, A. J. (1997). The Academic Achievement of Adolescents from Immigrant Families: The Roles of Family Background, Attitudes, and Behavior. *Child Development*, 68(2), 351–363. <https://doi.org/10.2307/1131854>
- Garcia-Louis, C. (2019). Ni Latino, Ni Negro: The (In)Visibility of Afrolatino Males in Higher Education Research. *JCSCORE*, 4(1). <https://doi.org/10.15763/issn.2642-2387.2018.4.1.96-122>
- Godoy Peñas, J. A. (2020). Are You Black or Latino? Being Afro-Latino in the U.S. *Estudios Del Observatorio/Observatorio Studies*. Published. <https://doi.org/10.15427/or062-06/2020en>
- Harris, A. L., Jamison, K. M., & Trujillo, M. H. (2008). Disparities in the Educational Success of Immigrants: An Assessment of the Immigrant Effect for Asians and Latinos. *The Annals of the American Academy of Political and Social Science*, 620, 90–114. <http://www.jstor.org/stable/40375812>
- Ibarra, R. A. (2004). Chapter Six: Academic Success and the Latino Family. *Counterpoints*, 253, 113–132. <http://www.jstor.org/stable/42979576>
- “Immigrants in the United States”. American Immigration Council. (2021, September 21). <https://www.americanimmigrationcouncil.org/research/immigrants-in-the-united-states>
- Keller, U., & Tillman, K. H. (2008). Post-Secondary Educational Attainment of Immigrant and Native Youth. *Social Forces*, 87(1), 121–152. <http://www.jstor.org/stable/20430852>
- Kim, Y., Mok, S. Y., & Seidel, T. (2020). Parental influences on immigrant students' achievement-related motivation and achievement: A meta-analysis. *Educational Research Review*, 30. <https://doi.org/10.1016/j.edurev.2020.100327>
- Mwangi, C. A.G., Daoud, N., English, S., & Griffin, K. A. (2017). “Me and My Family”: Ethnic Differences and Familial Influences on Academic Motivations of Black Collegians. *The Journal of Negro Education*, 86(4), 479–493. <http://www.jstor.org/stable/10.7709/jnegroeducation.86.4.0479>
- Nolasco, V. J. (2020). Doing Latinidad While Black: Afro-Latino Identity and Belonging. *Graduate Theses and Dissertations*. <https://scholarworks.uark.edu/etd/3713>
- Qian, Y., Buchmann, C., & Zhang, Z. (2018). Gender differences in educational adaptation of immigrant-origin youth in the United States. *Demographic Research*, 38, 1155–1188. <http://www.jstor.org/stable/26457073>
- Raleigh, E., & Kao, G. (2010). Do Immigrant Minority Parents Have More Consistent College Aspirations for Their Children?*. *Social Science Quarterly*, 91(4). <https://doi.org/10.1111/j.1540-6237.2010.00750.x>
- Rong, X. L., & Grant, L. (1992). Ethnicity, Generation, and School Attainment of Asians, Hispanics, and Non-Hispanic Whites. *The Sociological Quarterly*, 33(4), 625–636. <http://www.jstor.org/stable/4121400>
- Thomas, K. J. A. (2009). Parental Characteristics and the Schooling Progress of the Children of Immigrant and U.S.-Born Blacks. *Demography*, 46(3), 513–534. <http://www.jstor.org/stable/20616478>
- Urduan, T., & Munoz, C. (2012). Multiple contexts, multiple methods: a study of academic and cultural identity among children of immigrant parents. *European Journal of Psychology of Education*, 27(2), 247–265. <http://www.jstor.org/stable/43551109>

Appendix



IMMIGRANT PARENTS AND ACADEMIC SUCCESS

ANOVA ▼

ANOVA – I get good grades.

Cases	Sum of Squares	df	Mean Square	F	p
Black Latino ?	0.003	1	0.003	0.009	0.923
Residuals	13.558	39	0.348		

Note. Type III Sum of Squares

Descriptives

Descriptives – I get good grades.

Black Latino ?	Mean	SD	N
No	3.250	0.585	28
Yes	3.231	0.599	13

Figure 3. ANOVA was used to determine whether there was a significant difference in the perceived academic success of Black and non-Black Latino students. The perceived academic success of participants was measured using a 4-point Likert scale ranking agreement to the statement “I get good grades.”

ANOVA ▼

ANOVA – What range does your GPA fall in? (Refer to the scale above if you are unsure.)

Cases	Sum of Squares	df	Mean Square	F	p
Black Latino ?	0.440	1	0.440	1.246	0.271
Residuals	13.755	39	0.353		

Note. Type III Sum of Squares

Figure 4. An ANOVA test was conducted to find if there was a significant difference in the GPA's of Black Latino students and non-Black Latino students. GPA was measured using the College Board 4-point scale and was rounded to the nearest whole number.

IMMIGRANT PARENTS AND ACADEMIC SUCCESS

ANOVA

ANOVA – I feel a strong connection to my cultural background.

Cases	Sum of Squares	df	Mean Square	F	p
Black Latino ?	0.949	1	0.949	1.669	0.204
Residuals	22.173	39	0.569		

Note. Type III Sum of Squares

Descriptives

Descriptives – I feel a strong connection to my cultural background.

Black Latino ?	Mean	SD	N
No	3.250	0.752	28
Yes	2.923	0.760	13

Figure 5. An ANOVA test was used to find if there was a significant difference between feelings of connection to one’s cultural background, which was measured on a 4-point Likert scale, and the race of Latino students.



Immigrant Parents and Academic Success

Student Researcher: Breanna Villarreal

Title of Project: Immigrant Parents and Academic Success

Purpose of Project: To expand the research on Latino students and find if race, ethnicity, generational status, and cultural identity have any effect on their academic success and motivation.

If you participate you will be asked to: Answer a series of questions on your cultural background, identity, academic motivation, and success.

Time required for participation: About 5 minutes

Risks: The survey has minimal risk.

Benefits: Participation in this study will help add to the body of knowledge surrounding Latino students of different racial and ethnic backgrounds and their academic advantages.

How confidentiality will be maintained: Participants will be kept anonymous, and results will be kept on a password protected computer.

If you have any question about this study, feel free to contact Katie Moltz or Jeanette Azzaretto at moltzk@vschsd.org and azzarrej@vschsd.org

Figure 6.

IMMIGRANT PARENTS AND ACADEMIC SUCCESS

1

Participation in this study is completely voluntary. If you decide not to participate there will not be any negative consequences. Please be aware that if you decide to participate, you may stop participating at any time and you may decide not to answer any specific question.

By choosing "I agree" I am attesting that I have read and understand the information above, and I freely give my assent to participate. *

I agree

2

How old are you? *

14

15

16

17

18

Other

Figure 7.

IMMIGRANT PARENTS AND ACADEMIC SUCCESS

3

Do you identify as any of the following: Hispanic, Latino? *

- Hispanic
- Latino
- Both
- Neither of these

4

Are your parent(s) or grandparent(s) immigrants? *

Choose the closest relatives to you. If both your parent(s) and grandparent(s) are immigrants, choose the option correlating with your parent or parents.

- Yes, one of my parents is
- Yes, both parents are
- Yes, one grandparent is
- Yes, two or more grandparents are
- No

Figure 8.

5

What country did the immigrant from the previous question emigrate from? *

Enter your answer

6

What countries did the immigrants from the previous question emigrate from? *

Input as: country 1, country 2

Enter your answer

Figure 9.

IMMIGRANT PARENTS AND ACADEMIC SUCCESS

7

97-100	4.0
93-96	4.0
90-92	3.7
87-89	3.3
83-86	3.0
80-82	2.7
77-79	2.3
73-76	2.0
70-72	1.7
67-69	1.3
63-66	1.0
59-62	0.7
55-58	0.3
51-54	0.0

What range does your GPA fall in? (Refer to the scale above if you are unsure.)

- 4.0
- 3.0-3.7
- 2.0-2.7
- 1.0-1.7
- 0.0

8

To what degree do you agree with the following statements?

	Strongly Disagree	Disagree	Agree	Strongly agree
I do well in school.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I get good grades.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My parents place a lot of value on education.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My parents push me to excel academically.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Grades are very important to me.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I want to get into a good college.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My parents want me to get in to a good college.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 10.

9

Do you identify as any of the following: Black, African American, Afro Latino? *

- Black
- African American
- Afro-Latino
- Any combination of these
- None of these

10

To what degree do you agree with the following statements?

	Strongly Disagree	Disagree	Agree	Strongly agree
I feel a sense of belonging around other Latinos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel comfortable in most spaces I enter.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel a strong connection to my cultural background.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

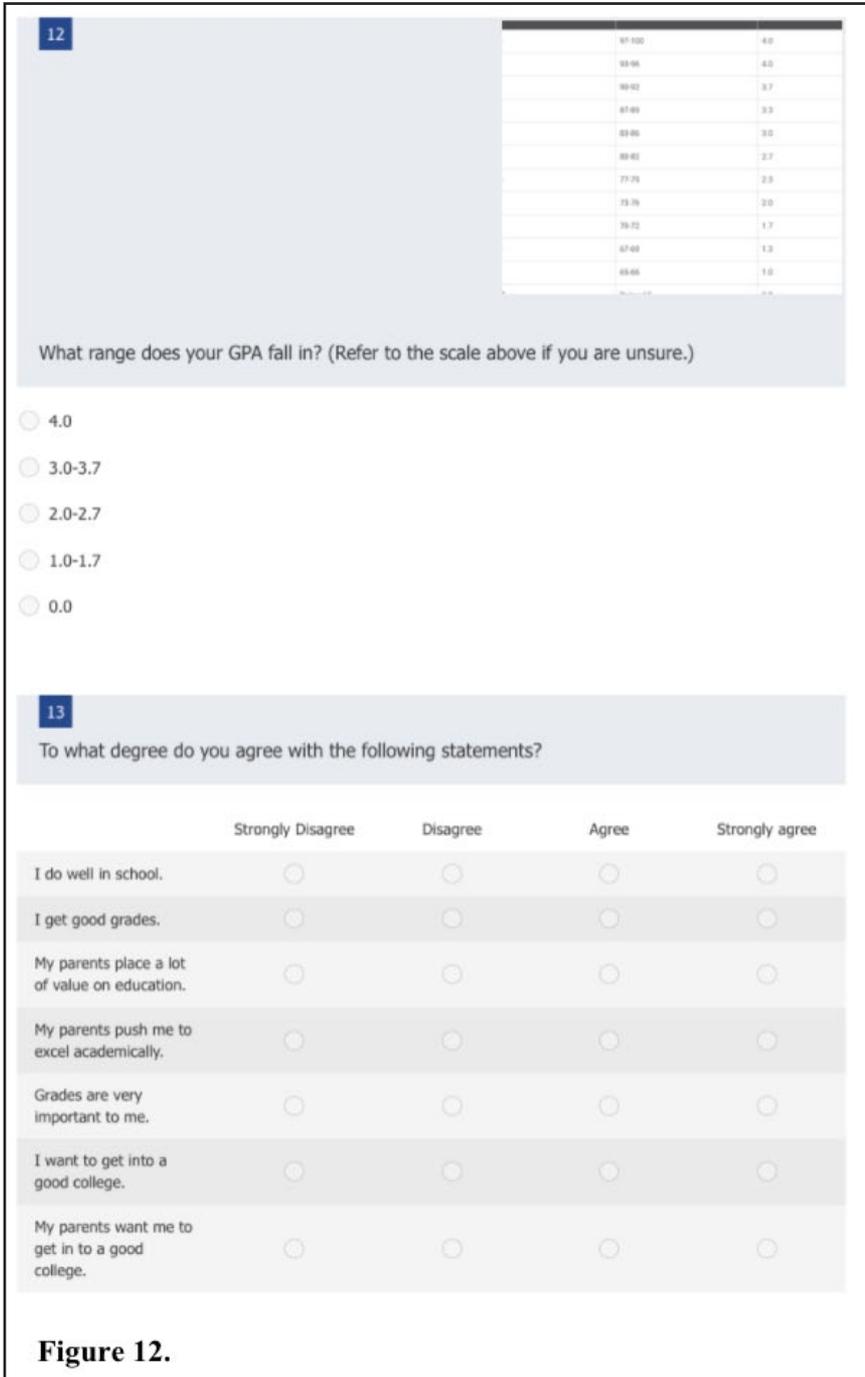
11

To what degree do you agree with the following statements?

	Strongly Disagree	Disagree	Agree	Strongly agree
I feel a sense of belonging around other Latinos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel a sense of belonging around other Black people/African Americans.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel comfortable in most spaces I enter.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel a strong connection to my cultural background.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 11.

IMMIGRANT PARENTS AND ACADEMIC SUCCESS



IMMIGRANT PARENTS AND ACADEMIC SUCCESS

14

Do you identify as any of the following: Black, African American, Afro Latino? *

- Black
- African American
- Afro-Latino
- Any combination of these
- None of these

15

To what degree do you agree with the following statements?

	Strongly Disagree	Disagree	Agree	Strongly agree
I feel a sense of belonging around other Latinos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel comfortable in most spaces I enter.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel a strong connection to my cultural background.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I connect to both parts of my ethnic background almost equally.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 13.

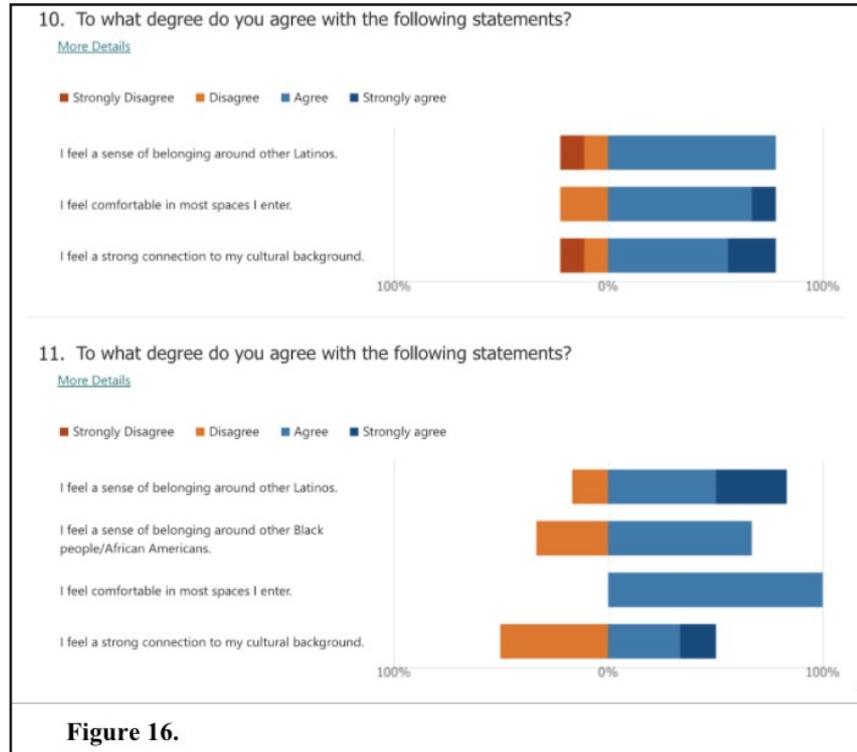
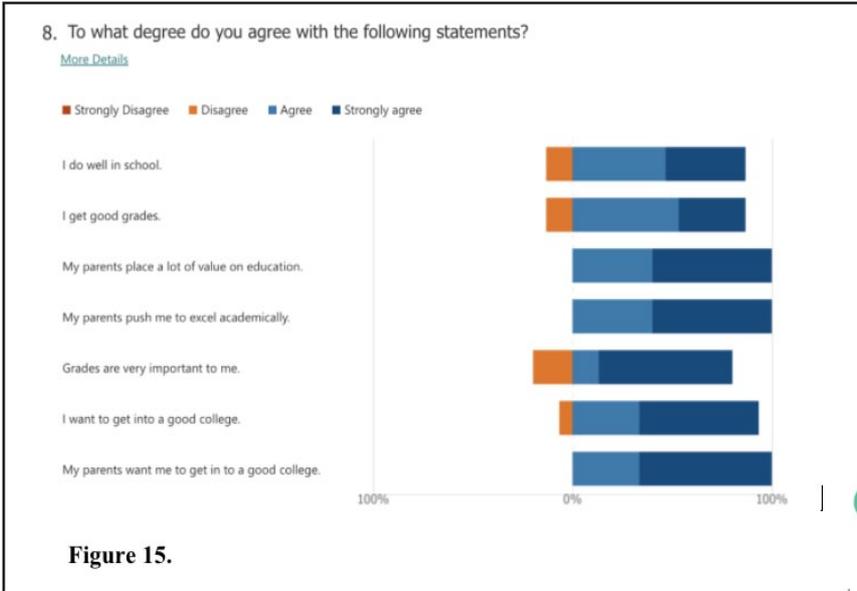
16

To what degree do you agree with the following statements?

	Strongly Disagree	Disagree	Agree	Strongly agree
I feel a sense of belonging around other Latinos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel a sense of belonging around other Black people/African Americans.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel comfortable in most spaces I enter.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel a strong connection to my cultural background.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I connect to both parts of my ethnic background almost equally.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 14.

IMMIGRANT PARENTS AND ACADEMIC SUCCESS



IMMIGRANT PARENTS AND ACADEMIC SUCCESS

13. To what degree do you agree with the following statements?

[More Details](#)

Strongly Disagree Disagree Agree Strongly agree

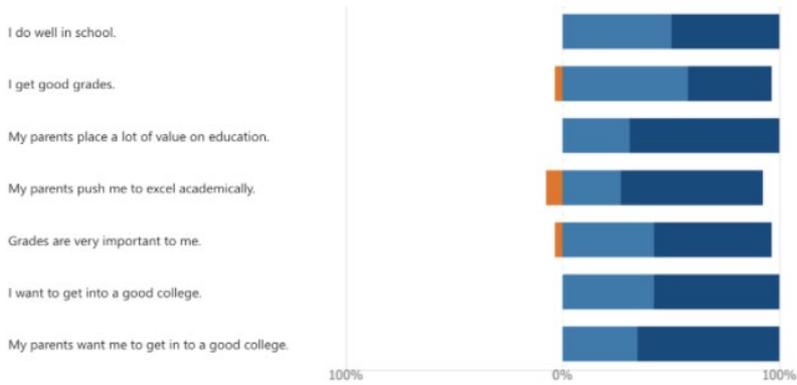
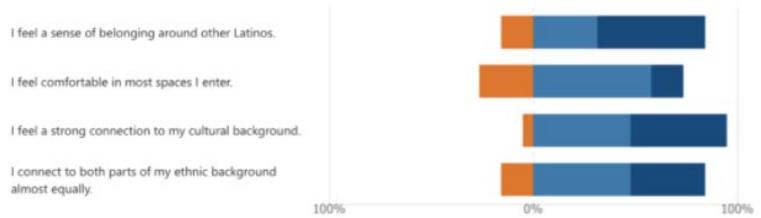


Figure 17.

15. To what degree do you agree with the following statements?

[More Details](#)

Strongly Disagree Disagree Agree Strongly agree



16. To what degree do you agree with the following statements?

[More Details](#)

Strongly Disagree Disagree Agree Strongly agree

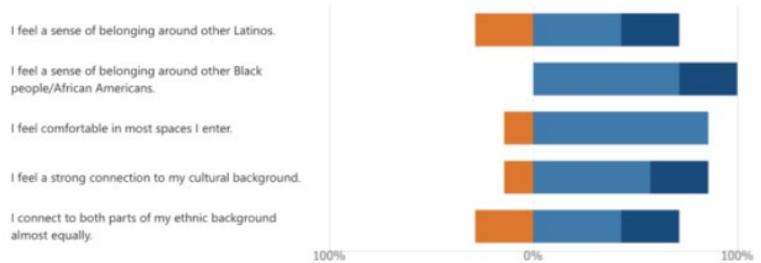


Figure 18.

INQUIRY PROPOSAL FORM – November 2021

1) Research question (with associated project goals if applicable).

To what extent does having immigrant parents affect academic success and motivations for Black Latino students versus non-Black Latino students between the ages of 14 and 18? The goal of this project is to fill the gap in research on Black Latino students and find out if their race, ethnicity, generational status, and cultural identity has any effect on their academic success and motivation.

2) Reasons for choosing the topic of interest and research question/project goal.

There has been a lot of research on a possible “immigrant advantage” where children of immigrants, specifically second-generation youth, have shown an increase in academic motivation and greater academic success than their non-immigrant related counterparts. This has been explored among different races and ethnicities or different generations, but Black Latinos specifically have not been studied. There is a widespread lack of Black Latino representation in research and I intend to fill that gap.

3) Data or information that will have to be collected to answer the research question/address the project goal.

I will conduct a survey to collect data from various Black and non-Black Latino students between the ages of 14 and 18 and ask them questions on their immigrant generational status, their family’s country of origin, their academic success/grades, and their academic motivation. Academic success and motivation will be asked using a scale and academic success will also be measured by asking participants to provide a range for their GPA.

4) Brief list of possible sources of information to discuss during the introduction of the paper. (5 Sources and explain how they will be used)

- a. **Source #1** - “Multiple contexts, multiple methods: a study of academic and cultural identity among children of immigrant parents”. This source will be used to introduce the body of knowledge on the topic and show that many researchers have already showed an academic advantage among children of immigrants and that there are differences in the extent of the advantage among different races and ethnic groups. This source will also be used to justify my use of a survey in my research, since this source surveyed 93 university students, both children of immigrants and children of native-born parents, to compare academic and cultural identity between the two groups.
- b. **Source #2** - “Generational Differences in Academic Achievement Among Immigrant Youths: A Meta-Analytic Review”. This source will be used to expand the knowledge in my field. This source researches the importance of generational status to the immigrant advantage and concluded that first and second-generation immigrants have a greater advantage over third and later generations. It also states that this first and second-generation advantage is more pronounced in Black and Asian immigrants. Since my research includes Black students and the impact of generation, this study helps justify researching that specific area.
- c. **Source #3** - “Doing Latinidad While Black: Afro-Latino Identity and Belonging”. This study examines how Black Latinos reconcile those two identities when society so often separates them. Other studies hypothesize that feelings of belonging and a strong cultural identity lead to greater academic success, so the lack of those feelings, which, according to this study, is commonly seen among Black Latinos, may cause an academic disadvantage. This source studies the hardships many Black Latinos face due to colorism within the Latin community as well as racism from outside the

Figure 19.

community. The recognition of Black Latinos as part of both communities creates a better sense of belonging and cultural identity, which in turn may lead to an increase in the immigrant academic advantage. This source can justify my focus on Black Latinos in my research and help discuss the gap in research on Black Latinos, which this source addresses.

- d. **Source #4** - "The Academic Achievement of Adolescents from Immigrant Families: The Roles of Family Background, Attitudes, and Behavior". This source talks about the importance of both generational status and ethnic background to the immigrant academic advantage. This source agrees with previous sources that second generation immigrants have an advantage over other generations and agrees on the importance of ethnicity and that students from certain ethnic backgrounds may have an advantage over others, similar to what Source #2 concludes. This source will be used to further justify my inclusion of generational status and specific country of origin in my research since this study and others have seen a difference in academic advantage between different groups under these categories. This study will also be used to show that while many researchers have discussed generational and ethnic differences, no one has focused on or even included Black Latinos or used my specific age range of 14-18, as this survey uses sixth, eighth, and tenth graders.
- e. **Source #5** "The (In)Significance of Race and Discrimination among Latino Youth: The Case of Depressive Symptoms". This source can be used to add to the justification for focusing on Black Latino students. It explains how Black Latino students are often at a disadvantage due to the racism and/or colorism they may face inside or outside the Latino community. This study examines how that discrimination can lead to depressive symptoms and other negative emotions, like feeling like an outsider. Since other studies show that these negative emotions can hinder academic motivation and success, it is clear that these variables are correlated, and Black Latino students should be studied further to understand how being Black and Latino affects academic motivation and success.

5) Chosen or developed research method to collect and analyze the above data/information. (alignment justify every decision about the method you will make)

My chosen research method is a survey. Many other studies in the same field or working with similar data and subjects have also used a survey. My survey will first ask participants their age, to make sure they are in the targeted age range of 14-18. I am using this age range because most other studies in the field used either a range of younger middle and high school students, from sixth to tenth grade, or university students from 18-25. My study creates a middle range that has not been studied extensively and is beneficial because many students in this age range are thinking about or applying to college and getting more serious about their studies, so they are the best to study on academic motivation and success. Next, my survey will ask participants whether they, their parents, or their grandparents are immigrants and from what country in order to categorize participants by generational status and ethnic background. Participants can indicate if one parent/grandparent or two parents/grandparents are immigrants and if the one of the options for two immigrant relatives is chosen, they will be able to input two countries. Many Black Latinos are from mixed backgrounds which contributes to their lack of representation in research because they are often forced into one category, splitting their identity to either Latino or Black, not both. Participants will then be asked if they identify as Black/African American/Afro-Latino so I am able to separate results of Black Latinos and non-Black Latinos and compare them. Then, my survey will ask students to choose a range that their GPA fits in and evaluate their feelings on a series of statements about their academic success and academic motivations using a 4-point Likert scale with anchors at Strongly Disagree and Strongly Agree. The Likert scale has been used in data collection for similar studies and will allow responses to be categorized more easily because there is no ambiguity. I am using a 4-point scale without a neutral option because selection of the neutral option would not add any data and could hinder my data collection. Finally, participants will be asked to evaluate their feelings on cultural identity and belonging using the same Likert scale. Some studies in the field have shown a correlation between cultural identity and academic success and other

Figure 20.

studies have hypothesized that Black Latinos may have a weaker cultural identity and sense of belonging due to their racial and ethnic identities often being separated and feelings of exclusion from both communities. Asking participants questions on this can compare the cultural identities of non-Black Latinos to those of Black Latinos as well as see how that affects academic motivation and success. These questions will be asked last so that participants' answers and thoughts that arise from this topic do not affect their answers on academic motivation or success.

6) Equipment or resources needed to collect data or information.

I will need Microsoft forms to create and administer my survey, which I already have access to.

7) Anticipated challenges to implementing the chosen research method (to collect and analyze data or to pursuing research methods appropriate to a paper that supports a performance/exhibit/product). (limitations or wherever you think you will encounter challenges)

Since I will not have access to participants' grades and will have to rely on their self-reported academic success the data may not be as accurate as I would like, but keeping the survey anonymous and asking grade ranges, not specifics, should make participants more comfortable answering truthfully and help keep the results accurate. I also may not have access to a large enough sample size of Black Latino students in my school so I might have to distribute my survey to students in the sample size from different schools.

8) Expected approvals needed and from where (IRB, etc.).

I will need approval from the IRB on my project as well as informed consent forms for my 18-year-old participants and parental consent for my participants who are minors, 14 to 17-year-olds.

Figure 21.

REGENERON STS INSTITUTIONAL REVIEW BOARD (IRB) APPROVAL FORM

(Required for all research involving human participants. (Institutional Form or IRB (IRB) form may be substituted.)

Student's Name: Quinn Kifical Title of Project: Academic Success and Retention of Black Latino Students

Adult Sponsor: _____ Contact Phone/Email: williamk@msdhsd.org

To be completed by Student/Researcher in collaboration with the Adult Sponsor/Designated Supervisor/Qualified Scientist.

1. I have submitted my Research Plan which addresses research methodology, participant recruitment, confidentiality and privacy issues, informed consent procedures and a risk and benefit analysis for the human participants.
2. I have attached any surveys or questionnaires I will be using in my project.
3. I have attached an informed consent that I would use if required by the IRB.
4. Yes No Are you working with a Qualified Scientist?

Name: _____ Degree: _____

Email Address/Phone Number: _____

Experience/Training as it relates to this project: _____

ITEMS IN THIS BOX MUST BE COMPLETED TO BE VALID

To be completed by Institutional Review Board (IRB) after review of the research plan.

Check **one** of the following:

- Research project requires revisions and is NOT approved at this time. IRB will attach document indicating concerns and/or requested revisions.
- Research project is Approved with the following conditions below: (All 5 must be answered)
 1. Risk Level (check one): Minimal Risk More than Minimal Risk
 2. Qualified Scientist (QS) Required Yes No
 3. Written Minor Assent required for minor participants:
 - Yes No Not applicable (No minors in this study)
 4. Written Parental Permission required for minor subjects:
 - Yes No Not applicable (No minors in this study)
 5. Written Informed Consent required for subjects 18 years or older:
 - Yes No Not applicable (No subjects 18 yrs or older in this study)

IRB SIGNATURES (All 3 signatures required) None of these individuals may be the adult sponsor, designated supervisor, qualified scientist or related to (e.g., mother, father of) the student (conflict of interest).

I attest that I have reviewed the student's project and agree with the above IRB determinations.

Medical or Mental Health Professional (a psychologist, psychiatrist, medical doctor, licensed social worker, licensed clinical professional counselor, physician's assistant, or registered nurse)

<u>Kerry Haddock</u> Printed Name	<u>Psy. D.</u> Degree/Professional License
<u>Kerry Haddock</u> Signature	<u>11/9/21</u> Date of Approval

School Administrator <u>Patricia Cotten</u> Printed Name	Educational Leadership A&D Degree
<u>[Signature]</u> Signature	<u>11/9/2021</u> Date of Approval

Educator (not involved with the project) <u>Jeffrey He</u> Printed Name	MA Secondary Science Degree
<u>[Signature]</u> Signature	<u>11/9/21</u> Date of Approval

Average High School Teacher Salaries vs Student Performance On Standardized Tests: A Correlational Analysis of Northern New Jersey High Schools

Justin Sousa

Low teacher salaries have been prevalent since the mid-1900s, and teacher strikes, like the one that occurred in Chicago in 2019, have resulted in attempts to raise these salaries. This study analyzes whether raising teacher salaries could have benefits to other groups besides teachers themselves. Through a correlational analysis, the relationship between average teacher salaries and student performance on standardized tests (measured with SAT scores, math proficiency rates, and ELA proficiency rates) was examined among thirty randomly selected high schools in eight northern New Jersey counties. The findings show that average teacher salaries were positively correlated with all three metrics, but only those with SAT scores and ELA proficiency rates were significant correlations. These results do not suggest that raising teacher salaries will increase these metrics of student performance; they are solely an acknowledgment that such causation should be explored in order to properly inform policy on this issue.

Keywords: teacher salaries, student performance, northern new jersey, correlational analysis, high schools, stratified random sampling

Introduction

The issue of low teacher salaries in the United States, while commonly seen as a strictly contemporary issue, has been prevalent since the mid-1900s. This is evident by teacher strikes that occurred in the early 1950s in cities such as St. Paul and New York mainly due to demands for increased salaries (Schiff, 1953, 133). These strikes have continued to take place in the 21st century. Such examples of teacher strikes in the present-day that resulted from a demand for salary increases include the infamous Chicago teachers' strike that happened in October of 2019 where 25,000 teachers missed work for eleven days (Vlamis, 2019). In February of that same year, strikes and walkouts occurred in California, West Virginia, Arizona, Virginia, and many other states around the country due

to low teacher remunerations (Wolf, 2019).

While there may be many factors that contribute to the current low teacher salaries, the most direct one is the way that teacher compensation systems in the US are organized. In the US, 96% of public school districts pay teachers using a single salary schedule that bases teacher pay off of both their education levels and their teaching experience (Podgursky & Springer, 2011, 167). This leads to those with more years of experience being paid much higher salaries than those who are new entrants to the education field. This can deter new teachers from joining the field as they know that their salaries will most likely be very low, ultimately leading to schools struggling to hire qualified teachers for certain subjects, and lowering the overall quality of education (Podgursky & Springer, 2011, 169). Lowering education quality can also lead to lower academic achievement among students, which

begs the question of whether or not lower teacher salaries lead to lower student academic achievement and performance or vice versa; some studies have already analyzed similar relationships (Jimenez-Castellanos, 2010; Akhtar et al., 2016).

Literature Review

A sequential mixed-method study conducted in a large urban school district in Southern California (Jimenez-Castellanos, 2010, 351) analyzed the allocation of resources between schools of different Title and Program Improvement (PI) statuses (Jimenez-Castellanos, 2010, 356) and whether or not this contributed to an “achievement gap” between white students and students of color. PI schools do not meet state testing requirements, and Title 1 schools obtain financial assistance in order to provide resources to schools where economically disadvantaged students go (Jimenez-Castellanos, 2010, 356). Through conducting interviews with principals and observing schools, this study ultimately found that schools that have Title 1 status generally receive more administrative and discretionary funds and that non-PI (non-Program Improvement) schools, which generally have higher levels of student achievement, also have higher teacher salaries, indicating a possible correlation between the two (Jimenez-Castellanos, 2010, 364).

Likewise, a study done in Pakistan on secondary schools evaluated as one of its sub-hypotheses whether or not teacher salaries were correlated to academic achievement, defined as “a performance judgment test which evaluates that a learner has gained particular information or he has mastered the essential expertise or not” (Akhtar et al., 2016, 391). While the authors had originally hypothesized that there would not exist any correlation between teacher salaries and student academic achievement, they eventually rejected this hypothesis after gathering and evaluating their data. Using Pearson’s product-moment correlation coefficient, the results showed that the correlation between teacher salaries and academic achievement had a value of 0.71, indicating a strong positive correlation between these two variables (Akhtar et al., 2016, 398).

Contrary to the positive correlations that were discovered in these two studies, Springer and Winters (2009) found that increases in teacher base pay did not

have any significant impact on student performance (Podgursky & Springer, 2011, 184). When examining the impact of the School-wide Performance Bonus Program (SPBP), a teacher compensation program in New York City that rewarded teachers if the students in their school met predetermined performance targets set by the NYC Department of Education’s accountability program, the researchers compiled data regarding student outcomes and the class learning environment through a randomized selection process (Podgursky & Springer, 2011, 184). Ultimately, they found that this program had no significant impact on student proficiency and performance after it had been implemented for two years in the city; Goodman and Turner (2011), researchers who also examined the SPBP program, came to similar conclusions (Podgursky & Springer, 2011, 184). Similarly, other studies that analyzed compensation programs in other US cities (Nashville, TN and Chicago, IL) also found that these programs exhibited no overall effect on student test scores (Podgursky & Springer, 2011, 182). However, the authors did note that this lack of an impact only seemed to occur in the United States; the other studies mentioned in this paper that examined compensation systems in countries such as Israel, India, and Kenya, found that these programs had an overall positive effect on student performance (Podgursky & Springer, 2011).

Looking at all of the existing literature on the relationship between teacher salaries and student performance, it is evident that there exists a gap in the research base on this topic, specifically in the United States. While there are some findings in the US regarding this topic (Podgursky & Springer, 2011; Jimenez-Castellanos, 2010), the majority of the studies have taken place in countries such as Pakistan (Akhtar et al., 2016), Israel, India, and Kenya (Podgursky & Springer, 2011). Current findings from the US show that there is either a positive or no significant correlation between teacher salaries and student academic achievement or performance. However, these findings cannot yet be generalized to the entire country due to both the little research that has been done on this topic and the lack of homogeneity in how school districts are run in the US. These studies were also largely conducted in the late 2000s and early 2010s, so their findings may be outdated and the relationship between teacher salaries and student achievement/performance may have

changed since then. This study will attempt to contribute to the nascent research that is currently published on this topic and revise the current findings so that more concrete generalizations can be made about the correlation between teacher salaries and student performance in the United States. While this study does not explore the direct causation of teacher salaries on student performance, these generalizations can still lead to initial considerations about if teacher salaries should be increased because of potential benefits to groups other than teachers themselves, specifically to students and their performance.

This study will examine the correlation between teacher salaries and student performance on both college entrance and state standardized tests in high schools across northern New Jersey. According to the National Center for Education Statistics (NCES), the average teacher salary in New Jersey for the 2018-2019 school year was \$70,212 (National Center for Education Statistics, 2019). Compared with other states in the country, New Jersey, as of the 2018-2019 school year, had the 9th highest average teacher salary (National Center for Education Statistics, 2019). Despite this higher average salary, teacher protests due to issues with pay have still occurred in New Jersey. For example, teachers from Montclair, NJ protested in October of 2019 due to delayed pay raises and the fact that they were still being paid their 2017-2018 salaries (Martin, 2019). These protests indicate that many New Jersey teachers still want their salaries raised, so researching the possible effects of raising teacher salaries, which this study aims to do, is of utmost importance in order to begin to ascertain if raising teacher salaries in New Jersey could be worth the budgetary investment. *The hypothesis for this study is that there will be an overall positive correlation between average high school teacher salaries and student performance on standardized tests.*

Methodology

To test this hypothesis, a correlational analysis was used in order to observe any potential relationship between average high school teacher salaries and student performance. Using such a framework allows for a more quantifiable and more interpretive result to examine any possible correlation between the two variables. It also fits the needs of this study since it analyzes data from a specific area of New Jersey, meaning that collecting data and statistically assessing the level of correlation between the data is more feasible than using and analyzing the results of other studies, all of which do not analyze the New Jersey region specifically. The method implementation includes sampling for thirty northern New Jersey high schools, collecting teacher salary and student performance data, and computing a final Spearman correlation coefficient value for the average teacher salaries against each metric for student performance.

Sampling

Due to time constraints and convenience, only thirty northern New Jersey high schools were selected for the data collection process through a stratified random sampling technique. This sampling method includes dividing the population into different strata (groups) and randomly selecting a sample from each stratum (Taherdoost, 2020, 21). However, the random sample taken from each stratum relative to the sample size has to be proportional to the population of each stratum relative to the total population being drawn from, meaning that strata with more high schools also had more high schools in the sample than strata with fewer high schools. Compared with other non-randomized sampling methods, stratified random sampling is the best fit for the purposes of this study as it ensures that every stratum being drawn from is adequately represented (Taherdoost, 2005, 21), meaning that the final results are more generalizable to the population being analyzed.

In this study, the strata used for the sampling were the eight counties in northern New Jersey: Bergen County, Essex County, Hudson County, Morris County, Passaic County, Sussex County, Union County, and Warren County. A sample size of thirty high schools

AVERAGE HIGH SCHOOL TEACHER SALARIES VS STUDENT PERFORMANCE

was randomly chosen from a population of 418 high schools. The number of high schools being analyzed per county is listed in parentheses next to the county name:

Bergen (7); Essex (6); Hudson (3); Morris (4); Passaic (4), Sussex (1); Union (4); Warren (1)

The stratified random sampling was conducted on Microsoft Excel where all high schools were sorted into their strata (by county), and the '=RAND()' function, which randomly denotes each high school from each county an associated value from zero to one, was

used. For each stratum, the schools were then sorted from largest to smallest associated value, and the schools with the highest associated values were chosen for each stratum. However, some of the schools with the highest associated values did not have the available data vital for this study; in these instances, the succeeding high schools in the assortment of schools from highest associated value to lowest associated value whose data was available were chosen instead. The thirty high schools that were ultimately chosen are listed in the table below sorted by county.

Bergen	Essex	Hudson	Morris	Passaic	Sussex	Union	Warren
Emerson Jr Sr High School	Cedar Grove High School	Dr. Ronald E McNair High School	Hanover Park High School	Passaic County Technical Institute	Wallkill Valley Regional High School	David Brearley Middle/High School	Warren County Vocational Technical School
Bergen County Academies	East Orange Stem Academy High School	Bayonne High School	West Morris Mendham High School	West Milford High School		Academy for Information Technology	
Northern Valley Regional High School at Demarest	Columbia High School	Harrison High School	Dover High School	Clifton High School		New Providence High School	
Ramapo High School	Essex County Newark Tech		Parsippany High School	Hawthorne High School		Jonathan Dayton High School	
Rutherford High School	Bard Early College High School						
North Arlington High School	Millburn High School						
Fort Lee High School							

Data Collection

The independent variable for this research study was the average teacher salaries for all thirty high schools being analyzed. However, at least in New Jersey, the average teacher salary per school is not publicly available data. In order to get an estimate of the average teacher salary at each of the schools, the teacher salaries had to be individually obtained and then the average (arithmetic mean) of the salaries had to be calculated. The final procedure for obtaining the average teacher salaries included:

Visiting a particular school's website and finding the staff directory.

Reviewing the directory and collecting each individual teacher's salary through [NJ Records](#), which contains the salaries of teachers from New Jersey for the 2018-2019 school year. The categories of teachers that were included in the collection of salaries were math, social studies, English/language arts, science, business organization, health, physical education, art, music, world language, and special education. There were also some miscellaneous categories of teachers, especially at vocational schools, including carpentry, automotive, dance, and other subjects that fell outside of traditional academic areas. While these types of teachers do not teach the subjects that are tested on the standardized tests being analyzed, they were still included for the purposes of calculating the average salary of all teachers within a particular high school.

Adding the salaries together and then dividing that sum by the number of salaries collected from that particular school in order to ultimately obtain the average teacher salary rounded to two decimal places.

Note that these average salaries are most likely not exact as in some instances, the teachers' names were in the staff directory but did not show up for that particular high school on the NJ Records website. Therefore, those particular salaries were not included in the average salary for that high school's average salary calculation.

The dependent variable for this research study was student performance, which was broken down into three metrics: average SAT scores (out of 1600), math proficiency rates (%), and English language arts (ELA) proficiency rates (%). Average SAT scores were calculated by adding the average scores for both the reading/writing and math sections for each school. The

math proficiency rate is the percentage of students that met or exceeded expectations on the math portions of New Jersey statewide assessments (NJDOE). The ELA proficiency rate is the percentage of students that met or exceeded expectations on the English/language arts portions of New Jersey statewide assessments (NJDOE). The data for these metrics were obtained from the official New Jersey government website under the [NJ School Performance Reports](#) for the 2018-2019 school year, the same year that the teacher salary data was obtained. The average teacher salaries and student performance data for each high school were placed in a Microsoft Excel spreadsheet (see Appendix).

Average ACT scores, the alternative college admissions standardized test to the SAT, were not included in this analysis as the SAT is the widely more popular test in New Jersey: 82% of high school students from the New Jersey class of 2019 took the SAT (College Board) while only 25% of high school students from the New Jersey class of 2019 took the ACT (ACT Inc.). Using average SAT scores allows for a larger and more diversified sample of students to be included in the average score of that test; this is not the case with average ACT scores in New Jersey high schools.

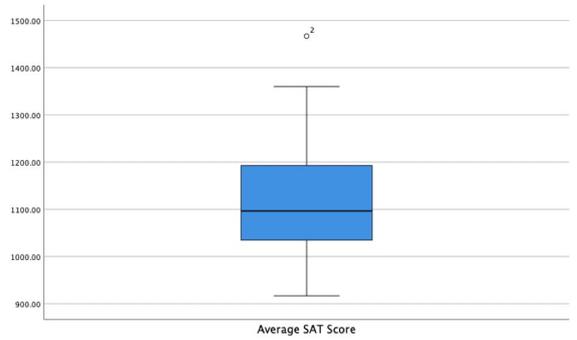
Correlational Analysis

Using the collected teacher salary and student performance data, a correlational analysis was conducted in order to reveal any potential relationship between the two variables. Spearman's rho correlation coefficient (r_s) was used for this analysis as opposed to Pearson's product-moment correlation coefficient due to the presence of outliers in the data. These outliers are shown in the four box plots, with the circles and asterisks indicating statistical outliers in the data sets. With the exception of ELA proficiency rates (Box Plot 4), all of the variables, both independent and dependent, being analyzed have at least one outlier in their set (Box Plot 1; Box Plot 2; Box Plot 3). Outliers in a data set have a significant influence on the value of Pearson's correlation coefficient and may skew it in a particular direction; therefore, Spearman's coefficient was used as it is relatively resistant against outliers (Schober et al., 2018, 1765).

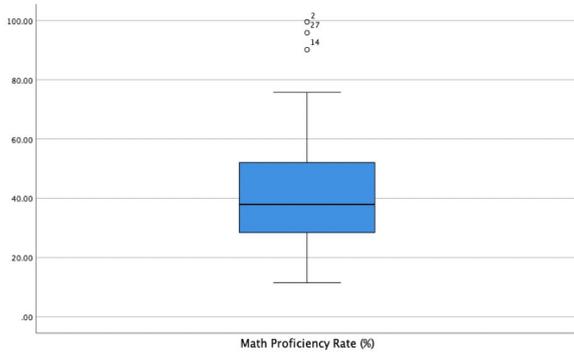
AVERAGE HIGH SCHOOL TEACHER SALARIES VS STUDENT PERFORMANCE



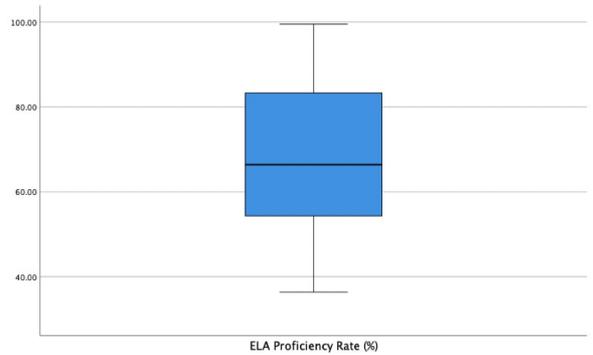
Box Plot 1 Spread and Distribution of Average Teacher Salaries with Associated Outliers



Box Plot 2 Spread and Distribution of Average SAT Scores with Associated Outliers



Box Plot 3 Spread and Distribution of Math Proficiency Rates with Associated Outliers



Box Plot 4 Spread and Distribution of ELA Proficiency Rates with Associated Outliers

The final value of Spearman's correlation coefficient is a number between -1 and +1 (Prematunga, 2012, 196). In the context of this study, a value of -1 means that there is a perfect negative correlation between average teacher salaries and student performance (as teacher salaries increase, student performance decreases by the same percentage), a value of +1 means that there is a perfect positive correlation between the two variables (as teacher salaries increase, student performance increases by the same percentage), and a value of 0 means that there exists no correlation at all (Prematunga, 2012, 196). For this research study, three correlation coefficients were computed to find the relationship between average teacher salaries and student performance. The three different correlations include average teacher salaries and average SAT scores, average teacher salaries and math proficiency rates, and average teacher salaries and ELA proficiency rates. To test these correlations for significance, a two-tailed p-value was calculated for each correlation coefficient with thirty pairs in the sample (N=30) and the degrees of freedom equal to twenty-eight (N-2=28). Calculating a two-tailed p-value means the expectation was that the correlation could go in either direction. The significance level was set at $\alpha=0.05$, and both a null and an alternate hypothesis, listed below, were established:

Null hypothesis: there is no statistically significant correlation between average high school teacher salaries and student performance on standardized tests.

Alternative hypothesis: there is a statistically significant correlation between average high school teacher salaries and student performance on standardized tests.

If the two-tailed p-value for a correlation was less than the significance level, then the null hypothesis was able to be rejected and the alternative hypothesis was accepted. However, if the two-tailed p-value for a correlation was greater than the significance level, then the null hypothesis was not able to be rejected and that correlation would not be statistically significant. All three correlation coefficients and their associated two-tailed p-values were calculated using IBM's SPSS Statistics software, which has shown to provide valid tests for the significance of correlation coefficients (Obilor & Amadi, 2018, 23).

Results

Average Teacher Salaries vs Average SAT Scores

Table 1 shows the value for Spearman's Rho correlation coefficient for the relationship between average teacher salaries and average SAT scores and the associated two-tailed p-value. A mild positive and statistically significant correlation was found between average teacher salaries and average SAT scores, $r_s(28) = .378$, $p = .039$ (Table 1). This mild relationship is also apparent in Figure 1, which shows the scatter plot for the two variables and the associated trend line, which is in an upward, positive direction: as average teacher salaries increase, average SAT scores generally increase as well. Since this correlation is statistically significant, the null hypothesis is able to be rejected and the alternative hypothesis is accepted, meaning that there is some significant positive correlation between average teacher salaries and average SAT scores for the thirty schools analyzed.

		Average Teacher Salaries	Average SAT Score
Spearman's rho	Average Teacher Salaries	1.000	.378*
	Correlation Coefficient		
	Sig. (2-tailed)		.039
	N	30	30
Average SAT Score	Correlation Coefficient	.378*	1.000
	Sig. (2-tailed)	.039	
	N	30	30

*. Correlation is significant at the 0.05 level (2-tailed).

Table 1 Correlation Results for Average Teacher Salaries vs Average SAT Scores

Average Teacher Salaries vs Math Proficiency Rates

Table 2 shows the value for Spearman's Rho correlation coefficient for the relationship between average teacher salaries and math proficiency rates and the associated two-tailed p-value. A mild positive but not statistically significant correlation was found between average teacher salaries and math proficiency rates, $r_s(28) = .306$, $p = .100$ (Table 2). This mild relationship is exemplified in Figure 2, which shows the scatter plot for the two variables and the associated trend line, which is in an upward, positive direction:

AVERAGE HIGH SCHOOL TEACHER SALARIES VS STUDENT PERFORMANCE

		Average Teacher Salaries	Math Proficiency Rate
Spearman's rho	Average Teacher Salaries	1.000	.306
	Correlation Coefficient		
	Sig. (2-tailed)		.100
	N	30	30
Math Proficiency Rate	Correlation Coefficient	.306	1.000
	Sig. (2-tailed)	.100	
	N	30	30

Table 2 Correlation Results for Average Teacher Salaries vs Math Proficiency Rates

as average teacher salaries increase, math proficiency rates generally increase as well. Since this correlation is not statistically significant, the null hypothesis is unable to be rejected, meaning that there is no notable correlation between average teacher salaries and math proficiency rates for the thirty schools analyzed.

Average Teacher Salaries vs ELA Proficiency Rates

Table 3 shows the value for Spearman's Rho correlation coefficient for the relationship between average teacher salaries and English language arts proficiency rates and the associated two-tailed p-value. A mild

		Average Teacher Salaries	ELA Proficiency Rate
Spearman's rho	Average Teacher Salaries	1.000	.395*
	Correlation Coefficient		
	Sig. (2-tailed)		.031
	N	30	30
ELA Proficiency Rate	Correlation Coefficient	.395*	1.000
	Sig. (2-tailed)	.031	
	N	30	30

*. Correlation is significant at the 0.05 level (2-tailed).

Table 3 Correlation Results for Average Teacher Salaries vs ELA Proficiency Rates

positive and statistically significant correlation was found between average teacher salaries and ELA proficiency rates, $r_s(28) = .395$, $p = .031$ (Table 3). This mild relationship is also shown in Figure 3, which shows the scatter plot for the two variables and the associated trend line, which is in an upward, positive direction: as average teacher salaries increase, ELA proficiency rates generally increase as well. Since this correlation is statistically significant, the null hypothesis is able to be rejected and the alternative hypothesis is accepted, meaning that there is some notable positive correlation between average teacher salaries and ELA proficiency rates for the thirty schools analyzed.

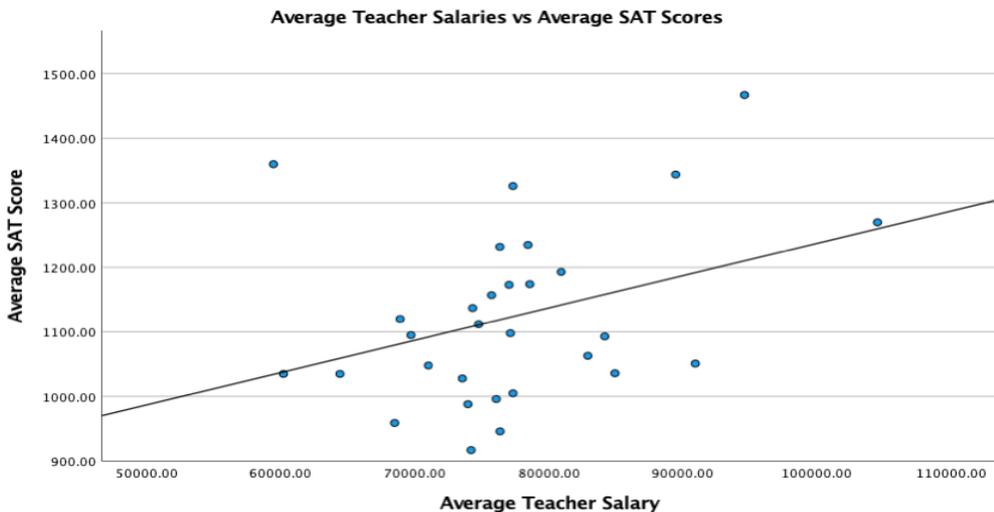


Figure 1 Scatter Plot and Trend Line for Average Teacher Salaries vs Average SAT Scores

AVERAGE HIGH SCHOOL TEACHER SALARIES VS STUDENT PERFORMANCE

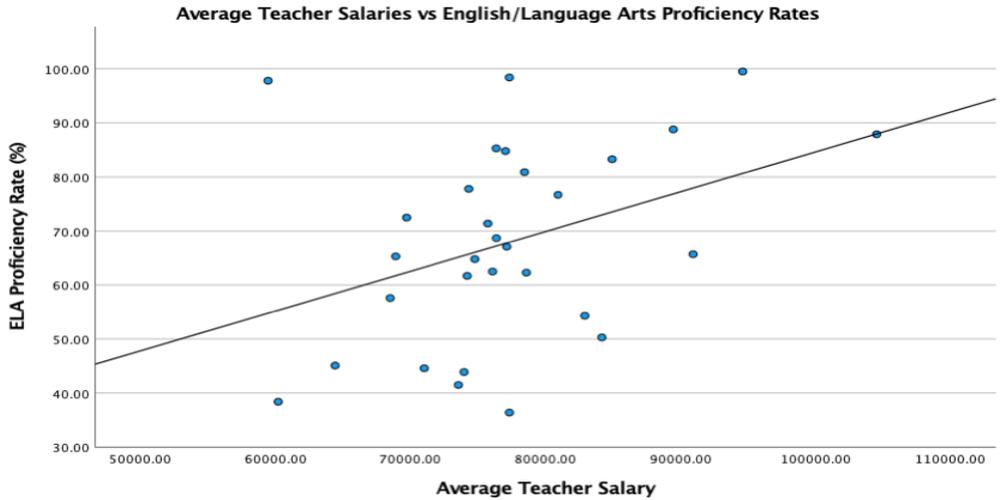


Figure 2 Scatter Plot and Trend Line for Average Teacher Salaries vs Math Proficiency Rates

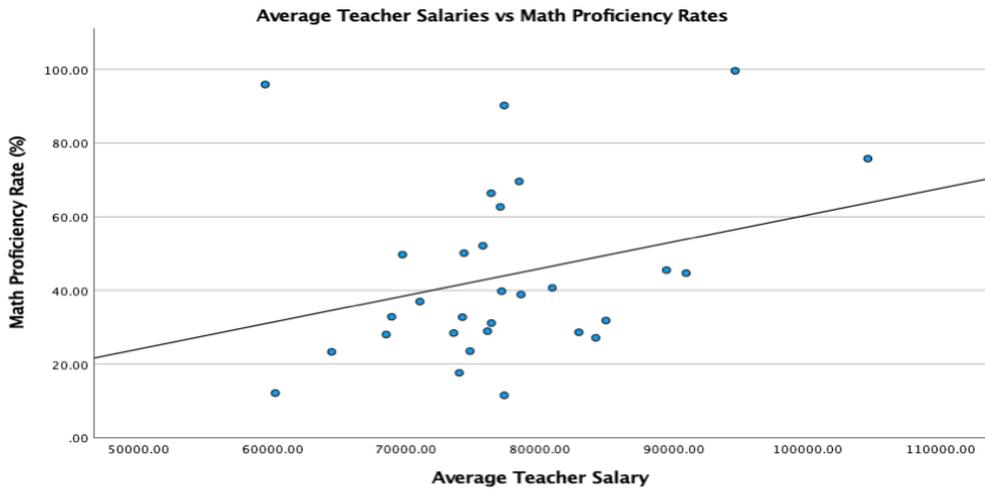


Figure 3 Scatter Plot and Trend Line for Average Teacher Salaries vs ELA Proficiency Rates

Discussion

The data shows that there exists an overall mild positive correlation between average high school teacher salaries and student performance on stan-

dardized tests for the region analyzed, with all of the correlation coefficients being between $r_s(28) = 0.300$ and $r_s(28) = 0.400$. These findings generally confirm the initial hypothesis that there would be a positive correlation between average teacher salaries and stu-

AVERAGE HIGH SCHOOL TEACHER SALARIES VS STUDENT PERFORMANCE

dent performance on standardized tests. However, the only statistically significant correlations at the 0.05 level occurred between average teacher salaries and average SAT scores ($p=0.039$) (Table 1), and average teacher salaries and ELA proficiency rates ($p=0.031$) (Table 3).

The statistical significance of these results indicates that there is a notable positive correlation between both average teacher salaries and average SAT scores and average teacher salaries and ELA proficiency rates among the thirty high schools analyzed in this study. Furthermore, because of the stratified random sampling method that was utilized to select the thirty high schools, this statistically significant correlation is generalizable to all the high schools in the eight northern New Jersey counties where the schools in the sample came from. However, it is important to note that this statistical significance can be considered weak since the two-tailed p -value for both correlations was just under the significance level of $\alpha=0.05$; there is still a 3.9% probability that the observed correlation between average teacher salaries and average SAT scores occurred by chance, and a 3.1% probability that the observed correlation between average teacher salaries and ELA proficiency rates occurred by chance. Although these probabilities may seem small, they are nonetheless important to consider as they could impact the existence of these correlations to the region being analyzed.

The relationship between average teacher salaries and math proficiency rates was found to be $r_s(28) = 0.306$ (Table 2), indicating a mild positive correlation. However, unlike the other two analyzed, this correlation was not found to be statistically significant at the 0.05 level. Based on the two-tailed p -value of 0.100 (Table 2) associated with this relationship, there is a 10% probability that the observed correlation between average teacher salaries and math proficiency rates occurred by chance. Therefore, it cannot be said that any true correlation between average teacher salaries and math proficiency rates exists in these thirty schools and in the eight northern New Jersey counties where these schools are located.

It is important to note that correlation does not equal causation, meaning that just because two of the relationships were found to be statistically significant, does not necessarily mean that the independent variable (average teacher salaries) is what caused changes

to the dependent variables (average SAT score and ELA proficiency rates). Nonetheless, if such causation were to exist, there are two factors that could potentially be contributors to the statistically significant positive correlations found between average teacher salaries and the two dependent variables: increases in teacher satisfaction and increases in teacher quality.

Higher teacher salaries could have increased teacher satisfaction and morale, which further increased student performance. Results from other studies have shown that low teacher salaries are the prime contributors to low teacher satisfaction (Miraj et al., 2018) and that increasing their salaries is a way to ultimately increase their overall job satisfaction (Fuming & Jiliang, 2007). These studies have also examined correlations between teacher satisfaction and student achievement, finding that better teacher job satisfaction can positively affect teacher performance and ultimately student achievement (Miraj et al., 2018; Akhtar et al., 2016). The significant correlations that were observed in this study may have occurred due to an indirect effect of teacher salaries on student performance through teacher satisfaction. However, because the studies that have come to these conclusions were conducted outside of the United States (Pakistan and China), those results may not be applicable as potential explanations for the observed correlations in this study.

It is also possible that increases in teacher salaries increase student performance indirectly by raising teacher quality. Findings from a study conducted in US public schools using teacher-level data from the Schools and Staffing Survey (SASS) found that, almost without exception, there is a positive relationship between teacher salaries and teacher quality (Figlio, 1996), suggesting that raising teachers' wages generally will attract higher-quality teachers (Figlio, 1996; Hanushek, 2020). Raising teacher quality through increasing salaries could also mean that students are able to learn and absorb class material more efficiently, which would in turn raise student performance. Because this study was conducted within the US, its findings regarding teacher quality are more generalizable to this study than the findings from studies that evaluated teacher satisfaction (Miraj et al., 2018; Akhtar et al., 2016).

Conclusion

Limitations

In assessing the validity of these findings, it is important to state the limitations that may have had an effect on their reliability. The most apparent limitation to these findings is that the correlation results cannot be used to identify any causation of raising teacher salaries on student performance; the only thing that can be said with certainty is that for the thirty high schools analyzed, higher teacher salaries generally correlated with better student performance. In regards to the data collection, there were numerous instances when a teacher who was listed on the school's staff directory would not have salary records for the 2018-2019 school year. This could potentially mean that the average teacher salaries that were calculated for the schools could have been either inflated or deflated from the true average. These results may also be slightly outdated since they are from the 2018-2019 school year, which was the only year in which the teacher salary data could be obtained, and may not reflect current relationships between average teacher salaries and student performance in northern New Jersey.

Connection to Field

The findings of this research study display similarities with those of other studies. For example, Jimenez-Castellanos examined the relationship between school achievement and educational resource allocation among public schools in a California school district, and he found that schools with higher achievement generally had more funds allocated to them (Jimenez-Castellanos, 2010, 364). Akhtar et al. examined the causes of teacher satisfaction and how it related to performance among secondary school students in Pakistan, and they found that higher teacher satisfaction was associated with better achievement among students. Since teacher salaries were a big contributor to high or low teacher satisfaction, they also concluded that higher teacher salaries were associated with better student achievement. The findings of this study

also challenge those of other studies. One such example is the study discussed by Podgursky & Springer mentioned previously. This study analyzed the School-wide Performance Bonus Program in New York City that would provide teachers with extra compensation if students in their school met NYC performance targets, but they found that there was no significant relationship between compensation and performance (Podgursky & Springer, 2011, 184). These results are contradictory to the positive correlation that was generally observed between teacher salaries and student performance in this study.

While this study's structure and findings have similarities with those of others, none of these other studies examine direct relationships between teacher salaries and student performance on standardized tests in the United States. The findings of this study contribute to this gap as this direct relationship was examined in Northern New Jersey schools, but more research is still needed from other schools in other states in order for this gap to continue to be addressed.

Implications

There exists ever-growing pressure from teachers and teacher unions across the United States for teacher salaries to be raised. However, elected government officials and officials on school boards will most likely not want to raise teacher salaries, as this introduces higher labor costs, unless there is some external benefit in doing so. The findings of this study exhibit a possible external benefit as they imply that increasing teacher salaries could have slight positive impacts on student performance on both state and college-required standardized tests. This could potentially motivate municipal and state governments, who typically regulate teacher salaries in the US, to raise them as if they truly do positively affect student performance, then raising salaries could make their town's or state's education system look superior to others in the country.

The results of this study could provide teachers and teacher unions in the United States with extra evidence for the fact that raising their salaries will be highly beneficial not only to them but to their students as well. Having such results could ultimately lead to a more convincing argument in favor of raising teacher salaries, but as explained previously, the causation of

changes in teacher salaries to changes in student performance cannot be confirmed by the findings of this study. Parents of high school students may also be interested in the positive correlation observed in this study as it could potentially positively impact their child's quality of education and academic standing if teacher salaries were to be raised. This could also cause parents to consider moving to school districts where teachers are paid higher salaries or to support teacher strikes that aim to raise these salaries.

Future Directions

The limitations that are currently present in this study should be minimized for future studies in order to maximize the validity of the results. This study examined the statistical correlation between teacher salaries and metrics of student performance, and the results of this study shed light on a potential causation of raising teacher salaries and improvements in student performance in high schools. However, the causation of teacher salaries on student performance was not directly analyzed, meaning that the correlational results of this study, while implying a possible causation, do not verify that such causation exists. Future research should attempt to directly examine whether or not raising teacher salaries has direct causes on student performance as the results of these studies will provide further clarity as to whether raising teacher salaries will have positive or negative effects on student performance. Along with this, external factors that could affect student performance such as household income or expenditure per student could also be analyzed to see whether these factors or teacher salaries are what lead to changes in student performance. Future research could also analyze whether or not raising teacher salaries indirectly affects student performance by impacting other variables such as teacher job satisfaction or teacher quality (Figlio, 1996; Miraj et al., 2018; Akhtar et al., 2016). Lastly, future research may also want to look into both middle and elementary schools along with high schools so that the findings can be generalizable to a wider group of both teachers and students.

References

- ACT Inc. (2019). The Condition of College & Career Readiness 2019 New Jersey Key Findings [PDF]. 12.
- Akhtar, T., Shah, R. U., Ghazi, S. R., & Khalil, Y. K. (2016, October). Morale as a Predictor of Secondary School Teachers' Performance: A Study of the Schools of Khyber Pakhtunkhwa. *Dialogue (Pakistan)*, 13.
- College Board. (2019). 2019 New Jersey SAT Suite of Assessments Annual Report [PDF]. 18.
- Estimated average annual salary of teachers in public elementary and secondary schools, by state: Selected years, 1969-70 through 2018-19.* (2019). National Center for Education Statistics. Retrieved April 20, 2022, from https://nces.ed.gov/programs/digest/d19/tables/dt19_211.60.asp
- Figlio, D. N. (1997). Teacher salaries and teacher quality. *Economics Letters*, 55(2), 5. [https://doi.org/10.1016/S0165-1765\(97\)00070-0](https://doi.org/10.1016/S0165-1765(97)00070-0)
- Fuming, X., & Jiliang, S. (2007). Research on Job Satisfaction of Elementary and High School Teachers and Strategies to Increase Job Satisfaction. *Chinese Education and Society*, 40(5), 11. <https://www.tandfonline.com/doi/abs/10.2753/CED1061-1932400509>
- Hanushek, E. (2020, January 20). It's Not "for the Children". *Hoover Institution*, 4.
- Jimenez-Castellanos, O. (2010). Relationship Between Educational Resources and School Achievement: A Mixed Method Intra-District Analysis. *Urban Review*, 42(4), 22. JSTOR. 10.1007/s11256-010-0166-6
- Martin, J. (2019, October 15). Montclair teachers rally over errors in pay and delayed raises. *NorthJersey*. <https://www.northjersey.com/story/news/essex/montclair/2019/10/15/montclair-nj-teachers-plan-rally-after-series-paycheck-problems/3985822002/>
- Miraj, S., Reba, A., & Ud Din, J. (2018, August). A Comparative Study Regarding Teachers' Morale among Public and Private Schools at Secondary Level in Peshawar. *Bulletin of Education and Research*, 40(2), 14. <https://eric.ed.gov/?id=EJ1209723>
- New Jersey Department of Education - Home Page.* NJ.gov. Retrieved February 15, 2022, from <https://rc.doe.state.nj.us/>
- NJ Records.* Bergen Record. Retrieved February 15, 2022, from <https://content-static.northjersey.com/Data/caspio/bundle/NJTeachersPay.html>
- Obilor, E. I., & Amadi, E. C. (2018). Test for Significance of Pearson's Correlation Coefficient (r). *International Journal of Innovative Mathematics, Statistics & Energy Policies*, 6(1), 13. https://www.researchgate.net/publication/323522779_Test_for_Significance_of_Pearson's_Correlation_Coefficient
- Podgursky, M., & Springer, M. (2011, March). Teacher Compensation Systems in the United States K-12 Public School System. *National Tax Journal*, 64(1), 28. JSTOR. <https://www.journals.uchicago.edu/doi/10.17310/ntj.2011.1.07>
- Prematunga, R. K. (2012). Correlational analysis. *Australian College of Critical Care Nurses*, 25(3), 5. <https://pubmed.ncbi.nlm.nih.gov/22464607/>
- Schiff, A. (n.d.). Teachers' Strikes in the United States. *Phi Delta Kappa International*, 34(4), 3. JSTOR.
- Schober, P., Boer, C., & Schwarte, L. A. (2018, May). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, 126(5), 6. https://journals.lww.com/anesthesia-analgesia/Fulltext/2018/05000/Correlation_Coefficients__Appropriate_Use_and.50.aspx
- Taherdoost, H. (2016). Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research. *International Journal of Academic Research in Management*, 5(2), 10. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3205035
- Vlamis, K. (2019, October 31). Why US teachers have been walking out of schools nationwide. *BBC*. <https://www.bbc.com/news/world-us-canada-50233474>
- Wolf, Z. B. (2019, February 23). Why teacher strikes are touching every part of America. *CNN*. <https://www.cnn.com/2019/02/23/politics/teacher-strikes-politics/index.html>

AVERAGE HIGH SCHOOL TEACHER SALARIES VS STUDENT PERFORMANCE

Appendix

Average Teacher Salary and Student Performance Data for all 30 High Schools Sorted by County

BERGEN				
Schools	Avg Salary	Avg SAT Score	Math Proficiency Rate	Reading Proficiency Rate
Emerson Jr Sr High School	\$69,714.97	1095	49.70%	72.50%
Bergen County Academies	\$94,614.64	1467	99.60%	99.50%
Northern Valley Regional High School at Demarest	\$104,525.84	1270	75.80%	87.90%
Ramapo High School	\$77,043.39	1173	62.70%	84.80%
Rutherford High School	\$77,141.78	1098	39.80%	67.10%
North Arlington High School	\$64,422.94	1035	23.30%	45.10%
Fort Lee High School	\$80,937.97	1193	40.70%	76.70%
ESSEX				
Schools	Avg Salary	Avg SAT Score	Math Proficiency Rate	Reading Proficiency Rate
Cedar Grove High School	\$68,905.27	1120	32.80%	65.30%
East Orange Stem Academy High School	\$84,952.84	1036	31.80%	83.30%
Columbia High School	\$78,586.30	1174	38.90%	62.30%
Essex County Newark Tech	\$76,371.15	946	31.10%	68.70%
Bard Early College High School	\$74,213.19	917	32.70%	61.70%
Millburn High School	\$89,476.20	1344	45.50%	88.80%
HUDSON				
Schools	Avg Salary	Avg SAT Score	Math Proficiency Rate	Reading Proficiency Rate
Dr Ronald E McNair High School	\$77,330	1326	90.20%	98.40%
Bayonne High School	\$76,088.34	996	28.90%	62.50%
Harrison High School	\$77,334.50	1005	11.50%	36.40%
MORRIS				
Schools	Avg Salary	Avg SAT Score	Math Proficiency Rate	Reading Proficiency Rate
Hanover Park High School	\$74,330.68	1137	50.10%	77.80%
West Morris Mendham High School	\$76,356.31	1232	66.40%	85.30%
Dover High School	\$68,497.04	959	28%	57.60%
Parsippany High School	\$75,737.96	1157	52.10%	71.40%
PASSAIC				
Schools	Avg Salary	Avg SAT Score	Math Proficiency Rate	Reading Proficiency Rate
Passaic County Technical Institute	\$90,949.03	1051	44.70%	65.70%
West Milford High School	\$82,933.34	1063	28.60%	54.30%
Clifton High School	\$73,562.31	1028	28.40%	41.50%
Hawthorne High School	\$71,013.61	1048	37%	44.60%
SUSSEX				
Schools	Avg Salary	Avg SAT Score	Math Proficiency Rate	Reading Proficiency Rate
Walkill Valley Regional High School	\$84,191.55	1093	27.10%	50.30%
UNION				
Schools	Avg Salary	Avg SAT Score	Math Proficiency Rate	Reading Proficiency Rate
David Brearley Middle/High School	\$73,979	988	17.60%	43.90%
Academy for Information Technology	\$59,457	1360	95.90%	97.80%
New Providence High School	\$78,451.62	1235	69.60%	80.90%
Jonathan Dayton High School	\$74,776.59	1112	23.50%	64.80%
WARREN				
Schools	Avg Salary	Avg SAT Score	Math Proficiency Rate	Reading Proficiency Rate
Warren County Vocational Technical School	\$60,204.05	1035	12.10%	38.40%

Recurrent Neural Networks and Web Crawlers

Dhiraj Kanneganti

Web crawling techniques in conjunction with Recurrent Neural Networks (RNNs) have been applied to several areas in the field of data mining on the Internet, but how they would best be applied to searching for open datasets has not yet been studied. Open data are data generated by the community and serve as effective alternatives to more centralized data collection, such as the US Census. Since the individuals and small organizations that collect open data often lack the infrastructure to make it widely available, open data portals serve as key access points to centralize open data. Unfortunately, due to lack of funding, open data portals struggle to efficiently scrape large amounts of open data from the internet. The purpose of this experiment is to bridge the gap between open data sources and open data portals by creating an algorithm that can quickly find open data on the internet. Through the use of an RNN and focused web crawler, a search algorithm was developed that could scrape 1000 web pages per minute and identify open datasets at an 85% accuracy, both metrics suggesting that the algorithm is a significant improvement over existing open data collection methods. For future research into this field of study, this work suggests that the application of automated open data collection and the implications of the proliferation of open data portals be studied.

Introduction

With the amount of information on the internet rapidly increasing over the past decade, the need for computer programs capable of filtering through online data, also known as web scraping algorithms, has significantly grown [1,2]. Web scraping algorithms have been implemented in order to handle large amounts of information in several data-dependent fields such as data storage, search engines, and medical databases [3]. In particular, one type of web scraping algorithm, known as the web crawler, has grown in popularity in recent years. The web crawler is a type of web scraping algorithm that functions by accessing web pages through hyperlinks inputted by the user to extract Hyper Text Markup Language (HTML) elements¹ from

the web page. The web crawler will then access all the links on the web page and will keep repeating the cycle until the user terminates the program or the web crawler encounters a “dead end,” which is a web page with no further links [2]. Web crawlers have experienced this surge in popularity over other types of web scraping algorithms due to their simplicity and ability to adapt to a wide variety of situations [3]. While there are dozens of web crawling techniques available, the most commonly used is the distributed web crawler, which is similar to that of the basic web crawler blueprint, except that the distributed web crawler is designed to extract large amounts of information from many different styles of web pages [4]. However, little research has been conducted on the other types of web crawling techniques, which has led to many questions

1. HTML elements refers to the semantic categories used in HTML code to group text together. Notable examples are “title,” “header,” etc.

about the effectiveness of other types of web crawling techniques that were not built to be a “one size fits all,” like the distributed web crawler.

Much of the interest in web crawling has been fueled by potential applications in the data scraping market, which has led to a disproportionately heavy focus on distributed web crawling techniques over other forms of web crawlers [5]. However, there has been a recent spike in the usage of other web crawling techniques, specifically focused crawlers, which are web crawlers designed to search for and extract information from the internet [6]. Focused crawlers have a distinct advantage over distributed crawlers when it comes to finding specific information on the web; however, focused crawlers have taken the backseat to distributed crawlers since, unlike distributed crawlers, focused crawlers cannot sort through large amounts of information, which is the primary application of web crawling in the status quo [5,6]. Recent advances in data processing of focused crawlers have led to an increase in their viability in the data scraping market. Yet, the lack of research on focused crawlers has led to many struggles to find the most effective applications of this powerful technology. It is hoped that the results of this research will help pave the way for more insights into focused crawlers and their potential.

Literature Review

Thanks to the development of data processing technologies, focused web crawlers are well on their way to revolutionizing data retrieval on the open web. Focused web crawlers have been implemented in a variety of fields from hospital database administration to cybersecurity [7]. One prominent example of the use of focused crawlers was the creation of a web crawling bot by engineers at the Bharati Vidyapeeth College of Engineering capable of tracking price changes of a specific product across thousands of different e-commerce sites simultaneously [8]. The ability to extract and maintain access to that much information at once is practically unheard of in the current data

market and is just one example of the many potential uses of focused web crawlers. And while e-commerce is a field which has seen a drastic rise in the use of focused crawlers, text-based web page filtration has seen an even more extreme rise. Text-based filtering is when focused crawlers are trained to look for certain features in the text of certain HTML elements of web pages [10]. Focused crawlers play a key role in this process. In a text-based filtering project conducted by Gunawan et al., researchers of computer science at the University of Sumatra, found that “focused web crawler[s] ... were critical to improve performance of ... the automatic harvesting of articles from online publications” [10]. Not only are focused crawlers key to text-based classification in the status quo, but Dong et al., researchers of web design at Curtin University of Technology, also predict that “the plethora of data on the internet will force ... increased usage of focused web crawlers in the future” [11].

Text-based filtering does not solely comprise the focused crawler; it also requires the use of a Recurrent Neural Network (RNN), which is a type of machine learning algorithm² that creates a network of information that classifies input data by inputting it into layers of nodes, depicted as vertical columns of circles in Figure 1, which are connected by continuous series of loops, depicted as the lines in between each layer in Figure 1, known as neural networks [10, 11]. RNNs can be used in conjunction with focused crawlers in text-based filtration due to their ability to classify text through a process known as tokenization. RNNs use tokenization to convert textual data, which would consist of the HTML elements extracted from web pages using the focused crawler, into numerical data referred to as “tokens” [12]. These tokens are inputted into the RNN and are classified into categories based on underlying characteristics identified by the neural network. Due to the high classification accuracy of tokenization, RNNs have become the industry standard of machine learning classifiers to be used in conjunction with focused classification in text-based classification. Zulqarnain et al, researchers of artificial intelligence at the University of Tun Hussein, conducted an experiment on the text classification accu-

2. Machine learning is a subset of computer programs that can find trends in data that are difficult to find [13]

racy of several machine learning models and found that “RNNs ... have been achieving outstanding results in various areas of ... text classification” [13].

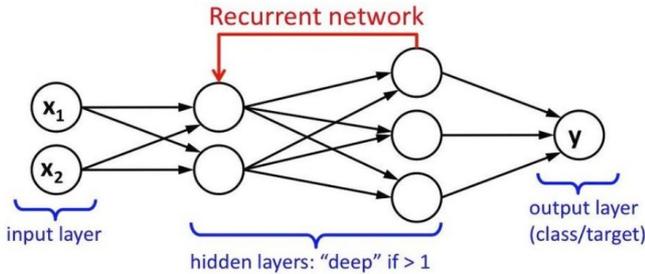


Figure 1 | Neural Network Neural networks are a series of layers that have checkpoints called “nodes.” The data will activate a node if it meets certain criteria set by the neural network, which can then classify data based on the path of activated nodes. In this illustration, the circles represent nodes and the vertical columns of nodes represent the layers [14].

Many studies have combined the usage of RNNs and focused crawlers in order to make text-based web page classification systems. Buber and Diri, researchers of computer science at Yildiz Technical University, created “an information retrieval application that... makes use of an RNN ... based architecture in conjunction with... a focused web crawler... which achieved a web page classification accuracy of 85%” [15]. The synergy between RNNs and focused crawlers has led to successful applications created in both the scholarly and corporate fields.

The research conducted in this project is valuable for a multitude of reasons. Not only will the results of this experiment give important insight into the potential uses of focused crawlers, but they will also generate a tool that could be used to extract open data. The goal of this experiment is to create a program that uses Recurrent Neural Networks in conjunction with Web Crawling techniques to crawl through the internet and identify open data to be scraped.

Open data, also referred to as community-generated datasets, are collected from the internet independently from large institutions[16]. Currently, there is

a strong need in the status quo for community-generated datasets. The current source of data for many businesses and individuals is large government datasets like the US Census. However, the Census is often-

times unreliable due to data collection occurring once every ten years [17]. Furthermore, a lack of government funding in less populated areas means that the Census often produces inaccurate data about rural communities [18]. However, there is a solution to the lack of reliable data in the form of open data portals. Open data portals are websites that keep data that is generated by the community. Open data gives rural populations an opportunity to collect their own data,

which, in theory, would be more accurate than that collected by an infrequent, centralized survey. Yet, there is often a lack of communication between open data portals and those who make community-generated datasets. The current expansion process used by open data portals is to manually scour the internet for these community-generated datasets [19]. This is because community-generated datasets are created by a wide variety of individuals, so it would be incredibly difficult to contact each individual and ask for their data. However, the process of combing through the internet, while more effective than directly asking the creator(s) of the dataset, is also very time consuming and has stunted the growth of open data portals around the world.

Even though text-based classifiers created to find open data on the internet would be ideal for alleviating this issue, the lack of profitability of open data portals has led to a gap in the research in this field due to a lack of financial incentive. As Zuiderwijk & Janssen, researchers at the Delft University point out, there is “little research on current methods open data portals could use to expand their databases” [16]. This leads to the research question: How can Recurrent Neural Networks be used in conjunction with Web Crawling techniques to identify and scrape open datasets? Thus, by creating a program which can automatically search the internet for open data, this gap can be alleviated, and it could spark the proliferation of open data portals around the world [16].

Methodology

The purpose of this experiment is to create an application capable of making community-generated datasets more accessible to open data portals. When building this program, there was an emphasis placed on efficiency and usability. In order to prioritize these two aspects, an Engineering Design methodology was developed to create and test an application capable of crawling through the internet in order to extract community-generated data.

A. Development Environment & Software Packages

To first go about constructing such an application, one must consider the software that will be used to construct the program. Although there are many possible coding languages that could be utilized in this experiment, Python version 3.6 was selected due to its flexibility and the availability of plenty of up-to-date machine learning libraries [20]. Python 3.6 is one of the latest versions of the language and is compatible with the web scraping software used in this experiment. Older versions of Python often have severe issues when running web scraping software, and were avoided in this project to prevent unnecessary bugs and/or glitches. PyCharm was selected as the primary development environment for this project as PyCharm is one of the premier integrated development environments³ (IDEs) for Python. And while there are several other IDEs that were considered, such as Google CoLab and Sublime Text Editor, Pycharm was chosen due to its compatibility with the wide variety of software packages⁴ needed in this experiment. Shetty et al, esteemed researchers of machine learning at the Anjuman Institute of Technology who have had significant experience with several Python IDEs, explain:

“Pycharm allows for complex... manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with [other] programs... making it a useful tool in ...

software experimentation” [21].

PyCharm enables the researcher to have flexibility when creating the application and the ability to approach this issue from multiple angles, both of which are valuable tools in this experiment.

Two main software packages were installed in order to create the application, Scrapy and TensorFlow. Scrapy is a web scraping framework that comes with pre-built web crawlers called “spiders” that can be modified to fit the user’s needs. Scrapy is not the only web scraping package available in Python 3.6, but it was used in this project due to its “powerful engine capable of controlling large amounts of data... and its ability to create stable focused crawlers” [22]. Other web crawling packages such as BeautifulSoup and Selenium often struggle when scraping from a variety of different web pages, but Scrapy’s data scraping engine is perfectly suited for doing such. TensorFlow was also used because it is the only software package available that has all of the tools required by the researcher to construct the RNN [22]. Other minor packages that were installed for machine learning include Keras, Pandas, NumPy, and Sklearn.

B. Web Crawler Construction

The web crawler was constructed in PyCharm using the Scrapy framework. The “broad crawl” spider was selected and installed from the Scrapy library as it is the most suitable to create the desired focused crawler. The web crawler was then adapted to scrape HTML elements of the previously selected web pages such as the title, paragraphs, and headers. These three HTML tags were chosen in particular since they are the most indicative of a webpage’s genre and purpose [9]. For the RNN to determine whether or not a web page is a community-generated dataset, it is important that it has access to the most important information on the web page, which most often lies within these three categories.

3. An IDE is a software that can be used as an environment to code computer programs.

4. A software package is a pre-made assortment of code that can be used to make new coding projects. Due to Python’s intertwined development with machine learning, many of Python packages relate to AI.

C. RNN Construction

The first step in the construction of the recurrent neural network is to gather web pages that can be used to test the web crawler and train the RNN text classifier. In this experiment, 300 web pages were collected after evaluation of the amount of text, size, and genre of the web page, a similar selection framework to that used for Zulqarnain et al.'s experiment. One hundred of the web pages can be considered community-generated datasets. This 33% split between non datasets and datasets was chosen because it is considered the industry standard in most RNN test classification projects [15]. It has been found that the most optimal way to train an RNN text classifier is to have one third of the training data contain the specific attribute that is being classified [9,10,12]. Training data is used to construct the machine learning model, while testing data can be used to check the accuracy and refine the trained model. There was also an effort to select web pages from a variety of different genres: business, entertainment, environment, demographics, and culture. This would ensure that the RNN would have a wide variety of text to train on, which would boost its accuracy and robustness [15]. Each web page's genre would not be included in the data analyzed by the RNN; however, the genre was used for analysis of the RNN's accuracy, which will be later discussed.

After the web crawler is able to scrape the desired text from the web pages, the RNN text classifier must be built. This was done through the use of TensorFlow and other previously mentioned machine learning packages. The tokenization process was performed through the Natural Language ToolKit (NLTK). The NLTK is one of the most prominent software libraries used in tokenization for RNNs because it can be applied to a wide variety of RNNs and the implementation process of the NLTK is very simple and quick. After the tokenization and RNN have been constructed, the text classifier must be trained on the datasets that were selected. Since the RNN's accuracy has to

be tested later on using new web pages it has not seen before, the web pages must be split into training and testing. The training group will contain about 66% of the web pages and the testing group will consist of the remaining 33%. The groups were randomly chosen and the training dataset, which consists of text from the titles, paragraphs, and headers of the web pages in the training group, was automatically inputted into the tokenization process and RNN. The final step in the construction of the RNN is the creation of the output layer. This layer will have two nodes: open dataset and not an open dataset. The text classifier's analysis of the scraped text will determine which of these two nodes will be activated, outputting the classification of the web page.

After training the RNN and constructing the web crawler, the final step before evaluation is to connect the two components. This can be done by having the web crawler automatically dump scraped text into a vector array⁵ which the RNN will access and classify, both of which will occur in real time.

D. CSV File Intermediary

Originally, the scraped text was immediately entered into the RNN, however, this led to several issues with accuracy since much of the unstored text would be lost. This lack of text drastically reduced the RNN accuracy, as it did not have sufficient information to construct the classification model. To better store information, the scraped text would be sent to a Comma Separated Values (CSV) file, where it could be stored without data loss and easily accessed by the RNN. Storing data in a CSV file meant that a NumPy function had to be used to process the CSV data to turn it into tokens. The use of this converter led to several latency issues and required several safeguards to be coded in to ensure that glitches with the CSV would not break the program.

5. A vector array is a two-dimensional storage unit that can be used to hold data from multiple variables. In the context of this experiment, the vector array will hold four columns of data: header, paragraphs, titles, and 1 or 0 to indicate if the web page is an open dataset.

E. Evaluation

In order to evaluate the effectiveness of the program, criteria that measure speed and accuracy such as web pages scraped per minute (WSPM) and RNN text classification accuracy of web pages were considered.

WSPM was used to measure the usability of the web crawling algorithm. If the WSPM is considered too slow or subpar, no open data portal will be incentivized to use this program since it is not an improvement over manual collection of community-generated datasets [16]. It is crucial that the speed of this program is a considerable improvement over current collection methods, otherwise the long-term project goal of increasing the proliferation of open data portals will not be met.

Classification accuracy is also held in high importance because speed of text extraction is irrelevant if the program cannot consistently identify community-generated datasets. The goal of this project is to have an RNN accuracy of at least 80%. Buber and Diri claim that RNN text classifiers that have at least an 80% accuracy can be deemed “viable tools in most industries” [15]. RNN Accuracy will be measured by comparing the RNN’s classification of a web page to the web page’s “ground truth.” Each web page used in this experiment was manually evaluated by the researcher and determined to be an open dataset or non-open dataset. This manual classification is a web page’s “ground truth.” The RNN will attempt to determine which of these two categories a web page falls into after analyzing the scraped text from that web page. The RNN’s classification for each web page will then be compared to the “ground truth.” If the RNN’s classification is the same as the “ground truth,” the classification is considered to be successful; if it differs, then the classification will be unsuccessful. The number of successes will be divided by the total number of attempts to determine the classification accuracy.

In addition to accuracy, loss will also be used to measure the effectiveness of the RNN. Loss is a self-inflicted penalty used by machine learning models to determine their efficacy [15]. A higher loss indicates a

weaker performance. Loss is a reliable indicator of the model’s performance and will be used to evaluate the RNN in conjunction with accuracy.

F. Optimization

After the accuracy and loss are obtained, the RNN will then be optimized using epochs. An epoch is an optimization the RNN makes as it re-analyzes data to create an improved model. Epochs allow the algorithm to make corrections to faulty predictions and find new trends in the data. The RNN was programmed to re-adjust its internal parameters after every epoch in order to decrease loss as much as possible. In this experiment, 10 epochs were performed to optimize the RNN. This specific number of epochs was selected as too many epochs would result in the RNN “over-learning” which the process of the algorithm becoming too accustomed to the training dataset and unable to classify any new data, and too few epochs would not give the algorithm enough opportunities to adjust itself.

Results

After the algorithm was executed in a PyCharm environment, the loss and accuracy were automatically measured using internal functions pre-built into the Keras software package. Figure 2 and Figure 3 indicate the accuracy and loss of the RNN when classifying web pages in Training and Testing Datasets, respectively. The accuracy and loss are graphed over the number of epochs. As such, in both datasets, the classification accuracy of the RNN increases as epochs increase, while the loss decreases. These values are also available in the Appendix.

In order to more closely examine the performance of the model and put the RNN's accuracy and loss into context, confusion matrices were generated for both the Training and Testing datasets. Figures 4 and 5 are confusion matrices that visually display Predicted Positives, Predicted Negatives, True Positives, and True Negatives. In this study, a Predicted Positive is a web page that the RNN classifies as an open dataset. A Predicted Negative is a web page that the RNN classifies as not an open dataset. A True Positive is a web page that is an open dataset, and a True Negative is a web page that is not an open dataset. Of the 300 webpages used to construct and train the RNN, 262 were classified correctly, bringing the classification accuracy to 87.33%. However, when just analyzing the results of the testing dataset, the model had a classification accuracy of 85%. This is the metric that will be used for evaluation, since the training dataset predictions may be slightly biased since they were used to construct and test the uncompleted versions of the

model, while the testing dataset was only used to test the final version. Therefore, when analyzing the final version of the RNN classifier, the testing dataset accuracy is more representative.

Furthermore, the classification accuracy was also further analyzed using the genres of the web pages. The percentage of correctly classified web pages in each genre was plotted in Figure 6.

Metrics to evaluate the performance of the web crawler were also collected, mainly the WSPM. The WSPM was automatically measured using functions from the ScraPy framework and came out to be 1000 web pages per minute.

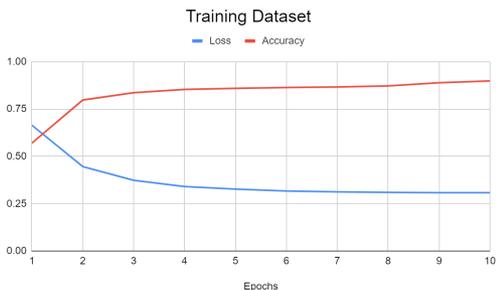


Figure 2 | Accuracy and Loss of RNN classifying Training Dataset over 10 Epochs



Figure 3 | Accuracy and Loss of RNN classifying Testing Dataset over 10 Epochs

		Training Dataset		
		True Positive	True Negative	Total
Predicted	Positive	60	6	66
	Negative	7	117	124
Total		67	123	200

Figure 4 | Confusion Matrix of Training Dataset Classification

		Testing Dataset		
		True Positive	True Negative	Total
Predicted	Positive	27	9	36
	Negative	6	58	64
Total		33	67	100

Figure 5 | Confusion Matrix of Testing Dataset Classification

Discussion

The results of this experiment suggest there is important discussion to be had about the use of RNNs to identify and extract open datasets from the internet. The discussion has been split into the following sections: Web Crawling, RNN Performance, Classification, and Web Page Genres.

A. Web Crawling

One of the primary components of this experiment was to evaluate the ability of a focused crawler to scrape text from open datasets. As previously mentioned, the WSPM of the web crawler was measured to be 1000, which is significantly higher than existing manual open dataset collection methods, meeting one of the research objectives. However, when compared to other crawlers, the classification algorithm is somewhat slow. Botify, one of the leading companies in the web crawling space has produced crawlers that can scrape text at 12000 WSPM [23]. This drastic difference in speeds is most likely because the crawler in this experiment must wait for the RNN to classify the web page before moving to the next page, while the

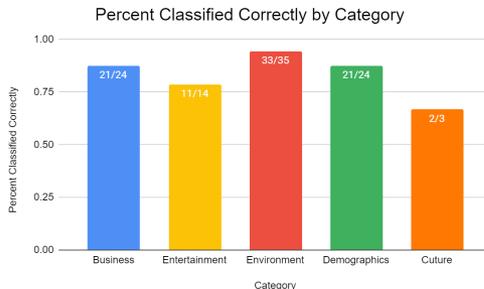


Figure 6 | Classification Accuracy of Different Open Dataset Genres At the top of each bar, the number of correctly classified datasets is divided by the number of datasets within the genre.

Botify crawler is not attached to any external classification algorithm.

While the crawler was efficient at scraping text from most web pages, it did run into several issues with more complicated web pages. The crawler functions by looking for HTML elements within the web page; if they are present then the crawler will extract the text and send it to the RNN, if not, the crawler will move onto the next page. Embedded HTML elements are elements inside of another element. The crawler aimed to scrape the paragraphs, titles, and headers of

each web page, but it was observed after preliminary testing that these target HTML elements that were embedded inside non-target elements would be completely skipped. For example, the text in a header embedded in a footer would not be extracted. This was not a severe issue as the amount of text left unextracted had a minimal impact on the RNN classification accuracy. The real issue was with non-target elements embedded inside target elements. In these cases, the crawler would instantly shutdown and lead to several code-breaking errors. While this occurred merely four times during the hundreds of tests, it is indicative of the primary weakness of the focused crawler: inflexibility.

The distinction between distributed and focused crawlers is that distributed crawlers are much more open-ended and adaptable than their counterparts [4, 21]. Adaptability comes at a cost, as distributed crawlers have a significantly longer run time than focused crawlers [4]. This trade-off between time consumption and flexibility was the primary reason for the use of focused crawlers in this experiment, as it was hypothesized that the rigidity of focused crawlers would not affect their ability to find open datasets. There is no need for adaptability when the crawler is looking for only one type of web page. However, it was apparent when the focused crawler was struggling to scrape text from complex pages with embedded elements that there were drawbacks to using the focused crawler. The straightforwardness of the focused crawler, while increasing time efficiency and helping meet one of the aims of this research, led to several errors when the crawler could not immediately find what it was looking for. This presents a unique scenario, as the crawler could be modified to scrape all the text from the web page, but that would take significantly more time, mostly removing the advantage of using focused crawlers. Through the implementation of focused crawlers in complex scraping tasks, this experiment has brought new insights into the tradeoff between time consumption and web scraping efficiency of focused crawlers.

B. RNN Performance

After epoch optimization, the RNN had a web page classification accuracy of 85% on the testing dataset,

meeting the research goal of above 80% accuracy. This means that the RNN can be considered a viable tool that can be used within professional environments, including open data portals [15]. The loss of the RNN was also very low, at only .32 for the testing dataset, further indicating that the RNN had high efficacy. Even though the accuracy of the RNN was satisfactory, there were some issues with its performance. RNNs usually classify text with no time constraints; however, in this experiment, the RNN had to classify a web page before the crawler could move to the next, meaning that the latency issues caused by the CSV converter were a significant issue. The implementation of a CSV file to act as a real time medium gave useful information to the potential data loss and lag with real-time text classification.

C. Classification

Interesting trends emerged when analyzing the classification patterns of the RNN. In both the Training and Testing datasets, the accuracy of the RNN when classifying non-open datasets was higher than when classifying open datasets. The classification accuracy for True Negatives in the Training dataset was 95%, while it was 89% for True Positives. In the Testing dataset, the classification accuracy was 86%, but only 81% for True Positives. This difference between True Positives and True Negatives may have been caused by non-open data web pages that contained datasets. When composing the 300 web pages for this experiment, several online datasets that are not open datasets, such as census data, were selected. This was done to ensure the RNN would not create a model that determined if a web page was an open dataset solely based on the presence of numbers or tables.

One recommendation the researcher makes for future work is that an RNN trained to classify open datasets is compared with a standard algorithm that determines whether or not a web page is a dataset by looking for the presence of numerical data, and the absence of keywords such as “government,” “census,” etc. The lack of an RNN would drastically decrease the time needed to execute the program, and if the classification accuracy of the standard algorithm is similar, then it could provide new insights into the application of RNNs.

D. Web Page Genres

The classification accuracy of the RNN was also measured across the different genres of web pages in the Training dataset. Across the five genres selected, Environment web pages were the most likely to be correctly classified at 88%. This was most likely because environmental open datasets are easily distinguishable due to their frequent use of data tables. Similar arguments can be made about Business and Demographics web pages which both had the second highest accuracy at 87.5%.

E. Limitations

While the results of this experiment are promising for the proliferation of open data portals, the results come with a caveat. There were two main limitations to this experiment: the RNN could only classify datasets in English and the composition of the web pages chosen may have influenced the classification accuracy. Even though the NLTK software package can be easily modified to be compatible with dozens of languages, this RNN could only classify English web pages [13]. This was due to the fact that the English-speaking researcher had to verify the web pages the RNN trained on, meaning that only English web pages were selected for this experiment. Further research should be conducted to examine if the classification accuracy of the RNN changes drastically as the language of the web pages changes. Secondly, the uneven distribution of genres may have led to alterations in the RNN classification model. Since the RNN was largely successful across all types of genres and this distribution is similar to what the RNN would experience in a non-controlled environment, this should not have too large of an impact on the results [15]. The researcher suggests that experimentation should be conducted, similar to that of Biber and Diri, that analyzes the effect of different genres in the Training set on classification accuracy.

Conclusion

The purpose of this experiment was to create a real time text classification system using an RNN and web crawler to identify open data sets on the internet. The algorithm's drastic improvements in speed and accuracy over existing methods to find open data suggest that this purpose was filled. Buber and Diri's framework of evaluating RNNs argues that an 80% classification accuracy is critical to be a useful tool, a metric that the RNN of this experiment comfortably meets [15]. These overall positive results validate several previous experiments conducted in the fields of web crawlers and RNN text classification.

The speed and general efficiency of the focused crawler provide further evidence to substantiate Dong et al.'s argument that focused crawlers will see a drastic increase in usage in the future [10]. Currently, outside of the focused crawler, there are no other tools in the market that are capable of filtering through that amount of text so quickly. The ability of the focused crawler to swiftly parse through HTML elements is unmatched and was key to meeting the experimental goal of improving upon existing open data collection techniques.

As previously mentioned, the primary shortcoming of the algorithm was the focused web crawler's inability to effectively scrape data from more complicated web pages, which reaffirms the results of Shkapenyuk & Suel who find that distributed crawlers are more suited to scraping data from more complicated web pages [4]. Yet, it contradicts the work of Khan & Sharma who claim that focused crawlers can be applied to a wide variety of situations [3]. The disagreement between studies and the results of this experiment defines a niche for focused web crawlers that entails a distinct usage pattern, only retrieving data from simplistic web pages, but a usage pattern that has usefulness across a wide range of industries. After analyzing the effectiveness of the algorithm, this experiment creates a new understanding of the strengths of focused web crawling algorithms with their high scraping speed, as well as the potential issues focused crawlers may have with more complicated web pages. The results of the RNN also paint a new understanding of their high accuracy with text classification of open datasets, a novel application of such text classifiers, but also the difficult implement-

tation and high maintenance required to ensure the RNNs work properly.

While the experimental goals of this research have been met, further study needs to be done on the actual implementation of this algorithm into open data portals. Studies such as Pisani et al. & Zuiderwijk & Janssen give valuable data about the impacts of the proliferation of open data portals, but existing methods to collect open data have prevented this proliferation from occurring in the first place. In order to better evaluate the new understanding generated from this experiment, as well as garner more insight into the nature of open data and the effects of open data portals on society at large, the implementation of this algorithm in the future is crucial.

References

- [1] A. Holst, "Total Data Volume Worldwide 2010-2025," Statista, 07-Jun-2021. [Online]. Available: <https://www.statista.com/statistics/871513/worldwide-data-created/>.
- [2] T. Karthikeyan, K. Sekaran, D. Ranjith, V. Vinoth Kumar, and J. M. Balajee, "Personalized content extraction and text classification using effective web scraping techniques," *International Journal of Web Portals*, vol. 11, no. 2, pp. 41–52, Jul. 2019. Available: https://www.researchgate.net/publication/336733112_Personalized_Content_Extraction_and_Text_Classification_Using_Effective_Web_Scraping_Techniques
- [3] S. Sirisuriya, "A Comparative Study on Web Scraping," *International Research Conference KDU*, vol. 8, no. 1, pp. 135–140, Nov. 2015. Available: <http://ir.kdu.ac.lk/bitstream/handle/345/1051/com-059.pdf?sequence=1&isAllowed=y>
- [4] V. Shkapenyuk and T. Suel, "Design and implementation of a high-performance distributed web Crawler," *International Conference on Data Engineering*, vol. 18, no. 1, pp. 1–12, 2012. Available: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.13.4762&rep=rep1&type=pdf>
- [5] L. Yu, Y. Li, Q. Zeng, Y. Sun, Y. Bian, and W. He, "Summary of web crawler technology research," *Journal of Physics: Conference Series*, vol. 1449, no. 1, pp. 1–7, Jan. 2020. Available: <https://iopscience.iop.org/article/10.1088/1742-6596/1449/1/012036/pdf>
- [6] T. V. Udupure, R. D. Kale, and R. C. Dharmik, "Study of web crawler and its different types," *IOSR Journal of Computer Engineering*, vol. 16, no. 1, pp. 01–05, Feb. 2014.
- [7] M. A. Khan and D. K. Sharma, "Self-adaptive ontology-based focused crawling: A literature survey," 2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), vol. 5, no. 1, pp. 595–601, Sep. 2016. Available: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7785024>
- [8] A. Ambre, P. Gaikwad, K. Pawar, and V. Patil, "Web and Android application for comparison of e-commerce products," *International Journal of Advanced Engineering, Management and Science*, vol. 5, no. 4, pp. 266–268, Apr. 2019.
- [9] F. Brenner, F. Platzer, and M. Steinebach, "Discovery of single-vendor marketplace operators in the Tor-Network," *The 16th International Conference on Availability, Reliability and Security*, 2021.
- [10] D. Gunawan, A. Amalia, and A. Najwan, "Improving data collection on article clustering by using distributed focused crawler," *Data Science: Journal of Computing and Applied Informatics*, vol. 1, no. 1, pp. 1–12, Jan. 2017. Available: https://www.researchgate.net/publication/330187644_Improving_Data_Collection_on_Article_Clustering_by_Using_Distributed_Focused_Crawler
- [11] H. Dong, F. K. Hussain, and E. Chang, "State of the art in semantic focused crawlers," *Computational Science and Its Applications – ICCSA 2009*, pp. 910–924, Dec. 2009. Available: https://www.researchgate.net/publication/44241179_State_of_the_Art_in_Semantic_Focused_Crawlers
- [12] L. Yu, Y. Li, Q. Zeng, Y. Sun, Y. Bian, and W. He, "Summary of web crawler technology research," *Journal of Physics: Conference Series*, vol. 1449, no. 1, pp. 1–7, Jun. 2020.
- [13] M. Zulqarnain, R. Ghazali, Y. M. Hassim, and M. Rehan, "A comparative review on Deep Learning Models for text classification," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 1, pp. 325–335, Jul. 2020. Available: https://www.researchgate.net/publication/340601688_A_comparative_review_on_deep_learning_models_for_text_classification
- [14] V. Mishra, M. Agarwal, and N. Puri, "Comprehensive and comparative analysis of neural network," *International Journal of Computer Applications*, vol. 2, no. 8, pp. 126–137, 2018.
- [15] E. Suber and B. Diri, "Web page classification using RNN," *Procedia Computer Science*, vol. 154, no. 1, pp. 62–72, Jan. 2019. Available: https://www.researchgate.net/publication/334474452_Web_Page_Classification_Using_RNN

[16] A. Zuiderwijk and M. Janssen, "The negative effects of open government data - investigating the Dark Side of Open Data," Proceedings of the 15th Annual International Conference on Digital Government Research - dg.o '14, vol. 3, no. 2, pp. 147–152, 2014.

[17] S. Noble, "Does census data still have value almost ten years later?," OCSI, 17-Dec-2020. [Online]. Available: <https://ocsi.uk/2020/11/23/does-census-data-still-have-value-almost-ten-years-later/>.

[18] N. A. Wardrop, W. C. Jochem, T. J. Bird, H. R. Chamberlain, D. Clarke, D. Kerr, L. Bengtsson, S. Juran, V. Seaman, and A. J. Tatem, "Spatially disaggregated population estimates in the absence of national population and Housing Census Data," Proceedings of the National Academy of Sciences, vol. 115, no. 14, pp. 3529–3537, Feb. 2018. Available: <https://www.pnas.org/content/115/14/3529>

[19] J. Umbrich, S. Neumaier, and A. Polleres, "Quality Assessment and evolution of Open Data Portals," 2015 3rd International Conference on Future Internet of Things and Cloud, vol. 7, no. 3, pp. 1–8, Feb. 2015. Available: <https://aic.ai.wu.ac.at/~polleres/publications/umbr-et-al-2015OBD.pdf>

[20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, and B. Thirion, "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research , vol. 12, pp. 2825–2830, 2011. Available: <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?ref=https://githubhelp.com>

[21] A. R. Shetty, F. B. Ahmed, and V. M. Naik, "CKD Prediction Using Data Mining Technique As SVM And KNN With Pycharm," International Research Journal of Engineering and Technology, vol. 6, no. 5, pp. 4399–4405, May 2019.

[22] Y. Fan, "Design and implementation of distributed crawler system based on Scrapy," IOP Conference Series: Earth and Environmental Science, vol. 108, p. 042086, 2018. Available: <https://iopscience.iop.org/article/10.1088/1755-1315/108/4/042086/pdf>

[23] A. Bouard, "Crawl speed: How many pages/second? 7 points to take into account," *Botify*, 11-May-2020. [Online]. Available: <https://www.botify.com/blog/crawler-impact-performance>. [Accessed: 23-Jun-2022].

Appendix

Appendix A | Loss and Accuracy of Training Dataset

Training Dataset		
Epoch	Loss	Accuracy
1	0.6647	0.5696
2	0.4459	0.7974
3	0.3735	0.8358
4	0.3404	0.8536
5	0.3268	0.8588
6	0.3167	0.8637
7	0.3126	0.8661
8	0.3094	0.8719
9	0.3081	0.8887
10	0.3079	0.8984

Appendix B | Loss and Accuracy of Testing Dataset

Testing Dataset		
Epoch	Loss	Accuracy
1	0.5139	0.7703
2	0.4192	0.8425
3	0.3722	0.8297
4	0.3396	0.8469
5	0.332	0.8604
6	0.331	0.8479
7	0.3355	0.8641
8	0.3239	0.8578
9	0.3438	0.849
10	0.3233	0.8557

Assessing the Threat: Antibiotic Resistant Bacteria Near Pittsburgh Hospitals

Brendon Frankel

The threat of antibiotic resistant bacteria is not only present, but increasing rapidly. Surveillance data that provides information about the location and caliber of antibiotic resistant bacteria is proving valuable when it comes to arranging appropriate medical responses. This study aims to address the antibiotic resistance surveillance data gap within the city of Pittsburgh and its hospital system. Six soil bacteria samples were collected at varying distances from a hospital in urban and suburban levels of anthropogenic activity and tested against ampicillin, ciprofloxacin, and erythromycin individually. The Traditional Agar Plating method was used to determine the resistance ratio of each sample by comparing colony forming units on control plates to antibiotic treated plates. Trends identifying the relationship between levels of anthropogenic activity/ distance from a hospital and antibiotic resistance ratios were present with Pearson correlation coefficients of $\sim-.92$ and $\sim-.62$ for suburban and urban anthropogenic activity respectively. Since these correlations are statistically insignificant due to the low sample size, they are unable to prove the diffusion of antibiotic resistance from hospitals to other areas. The resistance ratios did, however, identify the possibility of antibiotic resistance genetic exchange between the soil reservoir and pathogens circulating within the hospital, imposing health risks to the subject hospitals' patients. Aside from this genetic exchange dynamic, the implications of this study are minimal. However, the study acts as a framework for future research that entails a larger sample size to statistically prove the connection between the distance from a hospital and antibiotic resistance in bacteria.

Keywords: Antibiotic Resistance, Antibiotic resistant bacteria, nosocomial ARB, horizontal gene transfer of ARB, ARB in soil

Introduction

Prior to the discovery of Penicillin, in the early 1900s, a common abrasion had the potential to become life-threatening, and the odds of surviving bacteria-related infections were concerningly low (1, 2). After Penicillin gained clinical traction in 1940, millions of lives were saved, along with the credited transformation into the modern age of medicine. However, this illusory victory against infectious bacteria was short-lived. By 1950, the integrity of antibiotic treatment began to struggle as the inevitability of bacterial

evolution began to supply resistance against treatment (1). This evolution provides the conflict between advancements in antibiotic development and evolution of bacteria to resist treatment that continues to this day, with antibiotic resistant bacteria causing 2.8 million infections, 35,000 deaths, and \$4.6 billion in treatment costs every year (3, 4). Examples of pathogens that have become a particularly infamous threat regarding antibiotic resistance include *Clostridium difficile*, Carbapenem-resistant Enterobacteriaceae, and Methicillin-resistant *Staphylococcus aureus*. This increase of antibiotic resistant infection has to do with,

ANTIBIOTIC RESISTANT BACTERIA NEAR PITTSBURGH HOSPITALS

among other factors, excessive and often unnecessary use of antibiotics in hospitals as well as proximity to human environmental influence (3, 5). This evolved resistance is then accelerated by horizontal gene transfer. Evidently, this issue is both dangerous and rapidly evolving, increasing the need for and importance of surveillance data. Contributing any information regarding the risk of transmission, population-level resistance, and population-level antibiotic use is considered extremely valuable for the surveillance effort (6). Addressing this issue and the surveillance data necessity in the Pittsburgh area prompts the research hypothesis: Bacteria will become increasingly resistant to antibiotic exposure with closer proximity to hospitals in the Pittsburgh area in regions of both high and low human activity.

Literature Review

Key Definitions

Anthropogenic activity- Anthropogenic effects, processes, objects, or materials are those that are derived from human activities, as opposed to those occurring in natural environments without human influences (7).

Antibiotic overuse- The misuse of antibiotics combined with the unavailability of novel drugs, creating reduced financial inducements (8).

Antibiotic Resistant Bacteria- Bacteria demonstrating the ability to resist the effects of antibiotics (9).

Broad spectrum antibiotics- A broad-spectrum antibiotic targets both Gram-positive and Gram-negative bacteria e.g., ampicillin (10).

Gram negative bacteria- surrounded by a thin peptidoglycan cell wall, which itself is surrounded by an outer membrane containing lipopolysaccharide (11).

Gram positive bacteria- lack an outer membrane but are surrounded by layers of peptidoglycan many times thicker than is found in the Gram-negatives (11).

Gram Variable Bacteria- Bacteria that cannot be classified into either gram positive or gram negative (12).

Horizontal gene transfer- the movement of genetic information between organisms (except for those

from parent to offspring), a process that includes the spread of antibiotic resistance genes among bacteria, fueling pathogen evolution (13).

Narrow spectrum antibiotics- Effective against specific families of bacteria ex. erythromycin and ciprofloxacin (10).

Nosocomial- Originating within a hospital.

Horizontal Gene Transfer

In the context of antibiotic resistance, when a selective pressure (an antibiotic) is presented, genetic information that favors resistance can quickly (relative to the standard vertical evolutionary process) be spread throughout a bacterial population through horizontal gene transfer (HGT) (13). HGT is a significant factor that contributes to the rapid spread of antibiotic resistance, and can be the source of an outbreak due to the transfer of resistance genes to previously treatable pathogens (14). The importance of HGT in prokaryotic evolution is so extreme that 81% (+/- %15) of bacterial genomes have been shown to have originated through this evolutionary strategy (15). Combining the information that HGT is profoundly common and accelerates the rate of antibiotic resistance gene spread, a default explanation for antibiotic resistance in a certain area most likely includes the concept of horizontal gene transfer.

Nosocomial Antibiotic Resistant Bacteria

Hospitals rely heavily on antibiotics, with 79% of hospitalizations leading to antibiotic prescriptions and a 1.29:1 antibiotic prescription to hospitalization ratio (16). Due to this overuse of antibiotics, hospitals provide a selective pressure and promote antibiotic resistance with the consequent event of horizontal gene transfer. In a sense, hospitals serve as a “breeding ground” for the development and spread of antibiotic resistant bacteria (17). Control techniques such as frequent hand-washing, alcohol-based hand rubs, and glove use have been enacted to diminish the cases of antibiotic resistant bacteria in hospitals in response (18, 19). In addition to hygiene efforts, computerized antibiotic prescription monitors have been implemented to document and restrict the use of antibiotics. Unfortunately, even with these control efforts, nosocomial antibiotic resistant bacteria are becoming

ANTIBIOTIC RESISTANT BACTERIA NEAR PITTSBURGH HOSPITALS

more frequent. Nosocomial antibiotic resistance is increasing and responsible for one in four infections with long-term hospital visits (20, 21). These nosocomial pathogens also have the potential to contaminate the surrounding environment, including the soil (22). The threat of antibiotic resistant bacteria contaminating the surrounding soil lies in the possibility of antibiotic resistance genes exchanging (through HGT) between the soil and clinical pathogens, effectively making the soil a reservoir for nosocomial antibiotic resistance (23). Considering the high level of use of antibiotics as well as the potential of bacteria in soil acting as resistance reservoirs near hospitals, it is critical that monitoring efforts are directed towards this setting/interaction in particular (24, 23).

Anthropogenic Activity and Antibiotic Resistant Bacteria

In an abundance of studies, anthropogenic activity has been shown to be directly related to the presence of antibiotic resistant bacteria in the soil environment. Excluding variables such as soil composition, it has been demonstrated that areas with relatively high anthropogenic activity had an increased presence of antibiotic resistant bacteria within the soil (25). Conversely, samples from polar regions, where there is little anthropogenic activity, had a lower presence of antibiotic resistant bacteria (26). However, samples from minimally active regions were found to not be entirely free of antibiotic resistance genetic sequences (27). Even with this seemingly simple relationship between anthropogenic activity and antibiotic resistance, the mechanisms that cause this trend vary greatly and are not entirely understood (26). Acknowledging that anthropogenic activity may variate the amount of antibiotic resistant bacteria within the environment, surveillance data that tests this relationship may prove useful.

Antibiotic Types and Intent of Research

While bacteria are a relatively simple form of life, there are complex variations that differentiate the bacterial community into three main categories. These categories incorporate gram-positive, gram-negative, and gram-variable bacteria. The differences of these categories that are of concern is the construction of

the cell membrane/wall. Specific antibiotics (broad spectrum and narrow spectrum; *see Key Definitions section*) are needed for targeted treatment depending on the variation of the bacterial membrane. Using both broad spectrum and narrow spectrum antibiotics in a bacterial resistance study increases the implications of the results as multiple bacteria categories are being tested and multiple antibiotics are represented. Erythromycin was used as the gram-positive narrow spectrum antibiotic for the study, as it is established for treating gram-positive infections (28). Erythromycin functions by disrupting protein production to prevent further growth of bacteria (29). In addition to erythromycin, the gram-negative narrow spectrum antibiotic being tested is ciprofloxacin. Ciprofloxacin usually proves effective against gram-negative bacteria, but is still prone to resistance (30). Ciprofloxacin works, at the basic level, by inhibiting bacterial DNA replication, thus preventing reproduction (31). Lastly, the broad-spectrum antibiotic in the study is ampicillin (32). Ampicillin inhibits cell wall synthesis by binding to enzymes responsible for the formation of the cell structure (33). These antibiotics in particular have been determined to be representative of common treatments, which creates valuable data when testing for resistance against them.

Assumptions

It is assumed that various abiotic factors such as seasonal moisture composition of soil that may have impacts on antibiotic resistance (34) will be constant in each of the soil samples collected. Additionally, environmental factors that may have an effect on antibiotic resistance including: pH, temperature, and other chemical variables (26), are assumed to be constant throughout the soil samples collected. The study is grounded in the assumption that the antibiotics work as intended and are present in a consistent effective dosage in each of the sample plates.

Justification

Recapping the points that formulate the hypothesis, hospitals are still heavily involved in the “breeding” of antibiotic resistant pathogens, even with the control efforts mentioned previously. The incomplete control efforts combined with the overuse of antibiot-

ANTIBIOTIC RESISTANT BACTERIA NEAR PITTSBURGH HOSPITALS

ics have contributed to the development of nosocomial antibiotic resistant bacteria and the potential spread into the surrounding soil environment. This dynamic prompts the need for surveillance data (17, 18, 23). Adding another layer of inquiry to expand the surveillance data, anthropogenic activity is directly related to the amount of antibiotic resistant bacteria in the soil (excluding antibiotic resistance genes) (5, 25, 26, 27). The decision to include erythromycin, ciprofloxacin, and ampicillin serves to test the relationships between the vast array of bacterial species and multiple categories of commonly used antibiotics (10, 11). Considering the factors of anthropogenic activity and hospital roles in antibiotic resistant bacteria development, the hypothesis incorporating these subjects into variables becomes as follows: bacteria will become increasingly resistant to ampicillin, ciprofloxacin, and erythromycin in soil samples with proximity to Pittsburgh hospitals. Additionally, bacteria in these samples will demonstrate more resistance in an urban level of anthropogenic activity relative to a suburban level.

Methods

Sample Selection

Following Institutional Review Board approval, samples were collected at the chosen intervals (1mile, .5 mile, and on site of the hospital) in relation to a hospital located near urban downtown Pittsburgh (high anthropogenic activity) and in relation to a hospital in a suburban township (low anthropogenic activity) adding up to six samples in total. The low sample size is due to the constraints of equipment/bench space and the purpose of framing larger studies in the future. For consistent volume and depth of the samples, the rim of identical sample jars was pressed into the soil to create a circumference of the sampling space and the jars were filled with the same volume. Holding the circumference and volume constant for each sample ensures the same depth of soil was acquired. Another defining aspect of the study involves prevention of confounding variables that may have an effect on antibiotic resistance, such as water treatment plants, farms, and heavy metals (35). The samples were derived from locations with no known exposure to these

confounding elements. To eliminate horizontal gene transfer of antibiotic resistance genes between the collected samples, the shovel was sterilized with bleach before and after collection of each sample. Since seasonal moisture and temperature can have an effect on abundance of antibiotic resistance in a given sample, samples were collected on the same day in areas that portrayed similar moisture and temperature status (26, 35).

Procedure Selection

Among the wide array of procedure outlines that evaluate antibiotic resistance in bacteria, a traditional agar plating procedure has been determined to be the most beneficial in terms of features for the purposes of this study. This procedure involves the inoculation of bacterial solution on a petri dish, followed by counting the colony forming units (CFU) on each dish to derive quantitative data. Traditional agar plating encompasses attainable materials with low technical difficulty and is highly standardized. Traditional agar plating can assess non-isolated bacterial colonies (such as a diluted soil sample) and is able to determine the relative resistance of a sample compared with others, thus, fulfilling the intention to compare samples with one another (25, 26). However, this procedural outline requires optimizing the concentration of each sample to ensure a countable amount of CFU. Therefore, the serial dilution assay must be conducted for each sample to achieve the experimental standard colony forming unit count of 30-300 CFU. Additionally, the use of generalized Luria Broth (LB) media will limit the amount of culturable species within the sample (bacteria requiring specialized agar will not survive). Even with these limitations, the traditional agar plating method is superior for the purposes of this study compared with other methods. For example, traditional agar plating has been chosen above flow cytometry due to enormous upfront costs and equipment limitations inherent in the use of flow cytometry.

Procedure Methods

A 100-1000 μ L micropipette was utilized to ensure that a constant amount of bacterial solution was transferred to the petri dish and to eliminate other mea-

ANTIBIOTIC RESISTANT BACTERIA NEAR PITTSBURGH HOSPITALS

surement errors. Standard LB agar was used to provide a generalized growth medium that accounted for the growth of as many types of bacteria as possible. To prevent contamination, sterile pipette tips, dilution vials, an alcohol lamp, and scale trays were used along with sterile technique.

Concentration optimization. Via the serial dilution assay, each sample was diluted to an optimized concentration to yield the statistically reliable CFU count of 30-300 CFU (abiding by traditional agar plating protocol). In reference to the serial dilution assay procedure, one gram of soil from a given sample was weighed out using a digital scale and added to a sterile centrifuge tube containing 9 mL of distilled water. This effectively dilutes the original gram of soil to 10^{-1} . After the mixture has been stirred by inverting the tube continuously, 0.1 mL of this diluted concentration was added to an LB broth petri dish using the micropipette. To dilute the sample to 10^{-2} , 1 mL of the 10^{-1} dilution mixture was pipetted to an additional tube containing 9 mL of distilled water. After mixing the 10^{-2} tube, it was added to a petri dish as well. This process

was repeated for additional dilutions (10^{-3} , 10^{-4} , etc.). The plates containing the various dilutions were placed in a 37°C incubator for 24 hours (incubation protocol states 24- 48 hours). It is necessary to mention that the plates were oriented agar-side-up to prevent condensation and contamination. After incubation, the CFU present on each plate were counted and noted. The dilution with a CFU count of between 30-300 was considered statistically reliable and the optimized dilution for that given sample.

Preparation of control and experimental groups. To test the resistance ratio of the optimized solutions, the various antibiotics must be added. Four petri dishes were used for each sample consisting of an ampicillin plate, a ciprofloxacin plate, an erythromycin plate, and a control plate (no antibiotic). To ensure the doses of each antibiotic are both effective and quantifiable, effective concentrations of antibiotic/ LB broth solution suggested by the manufacturers were used. The effective concentrations are 100 $\mu\text{g}/\text{mL}$ for ampicillin, 10 $\mu\text{g}/\text{mL}$ for ciprofloxacin, and 250 $\mu\text{g}/\text{mL}$ for erythromycin. These concentrations were achieved by add-

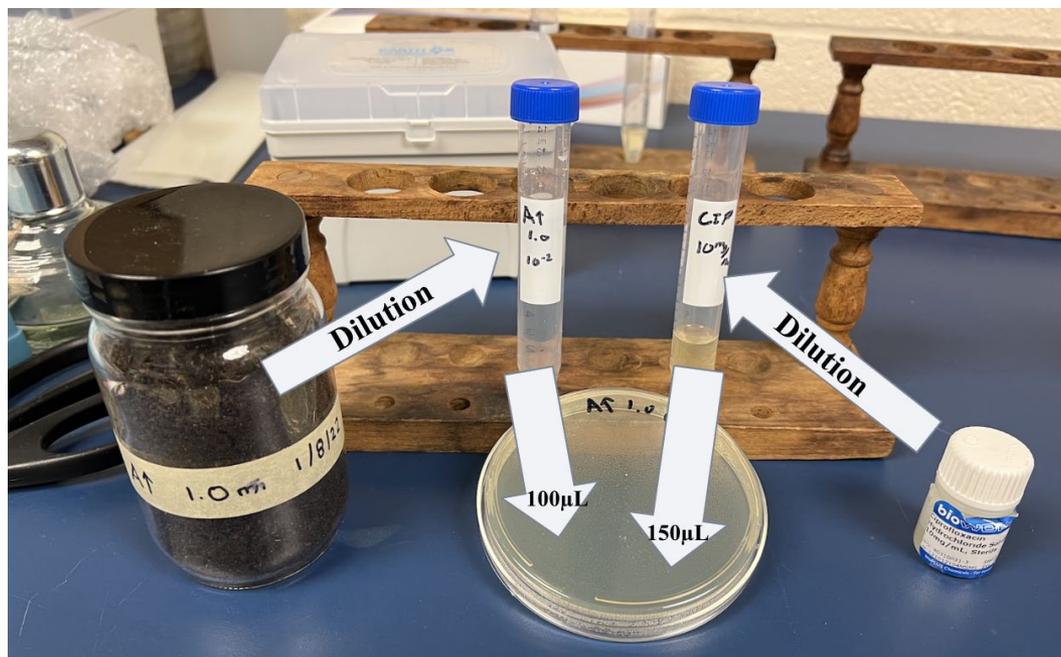


Fig 1 Demonstration of the materials used with high anthropogenic activity soil, diluted to 10^{-2} , 1 mile away from the hospital treated with 10 $\mu\text{g}/\text{mL}$ diluted ciprofloxacin.

ANTIBIOTIC RESISTANT BACTERIA NEAR PITTSBURGH HOSPITALS

ing 15 μL of 10 mg/ mL ampicillin solution to 1485 μL of LB broth, 3 μL of 10 mg/ mL ciprofloxacin solution to 2997 μL of LB broth, and 7.5 μL of 50 mg/ mL erythromycin solution to 1492.5 μL of LB broth. These solutions created the recommended effective concentrations while providing enough solution to spread across all six experimental plates for each antibiotic. Using the 100-1000 μL micropipette, 150 μL of each effectively concentrated antibiotic was transferred to the corresponding plate. A sterile glass spreader was used to evenly distribute the effective solution to ensure a constant amount of exposure regardless of the location of the bacteria upon the agar surface. After the antibiotic solution was allowed to absorb into the agar for 30 minutes, 100 μL of the optimized bacterial solutions from the serial dilution assay were added and spread onto the corresponding plates. After this addition, a plate consisted of its evenly spread out assigned antibiotic (except for the control plate) and the optimized bacterial solution for that particular soil sample (Fig. 1). The plates were placed lid side down overnight.

Data Analysis

After incubation of the experimental plates, the CFU on each plate was recorded on the CFU count section of a data sheet. The code written into the spreadsheet calculated the resistance ratio for each antibiotic (CFU of experimental plate / CFU of control plate) and the resistance ratio for the entire sample at a given location (average resistance ratio of the three experimental plates for that location). The resistance ratios for each sample in a given category of anthropogenic activity were then averaged to find the cumulative resistance ratio for that level of anthropogenic activity.

After identifying the resistance ratios for each sample and anthropogenic activity level, statistical tests need to be performed to determine the significance of the differences between the resistance ratios. Since the independent variable (distance from the hospital) and dependent variable (resistance ratio) are continuous, the Pearson correlation coefficient (r-value) was calculated to identify correlation between the resistance ratios and distances from the hospital. An r-value can range from $|1|$, which signifies a strong correlation, to zero, which signifies no correlation. The Pearson

correlation coefficient can also be used to derive a p-value. Following standard experimental alpha values, a p-value below 0.05 is considered significant while a p-value greater than 0.05 is considered insignificant. In terms of statistically comparing the anthropogenic activity average resistance ratios, a two-tailed, two-sample t-test was used. The type of t-test was chosen based on the necessity to compare the differences of separate sets of data.

The methods used align with the research question as they allow a resistance ratio for each sample to be calculated through the practice of traditional agar plating and counting CFU. Taking the resistance ratios into account, a Pearson correlation coefficient and consequent p-value are able to establish significance of the difference in data between the samples. Due to the precision measurements of the micropipette, constant conditions provided to the samples, and relation to a control group configured for each sample, the experiment suggests that any observed differences in resistance ratios are due to antibiotic resistance in bacteria alone. Statistically significant evidence can be used to contribute data to the surveillance of antibiotic resistant bacteria and confirm previous studies that correlate anthropogenic activity and distance from an origin of antibiotic resistance to antibiotic resistant bacteria present within a sample. The evidence will also provide insight to the general effectiveness of each antibiotic.

Results

The optimal concentration for each bacterial solution was successfully determined and countable CFU were present on every petri dish within at least one of the experimental runs. Resistance ratios were calculated based on the CFU counts, and trends were identified. The suburban anthropogenic activity samples displayed a strong inverse relationship between resistance ratio and distance from the hospital, while the urban anthropogenic activity samples presented a more casual inverse relationship. Both sets of data were determined to be statistically insignificant, thus the null hypothesis failed to be rejected. The difference in average resistance ratios between the levels of anthropogenic activity was minimal, and ultimately determined to be statistically insignificant.

Suburban Anthropogenic Activity CFU

The CFU counts of the suburban anthropogenic activity samples ranging from onsite of the hospital to 1 mile away were derived (Fig. 2). The effective bacterial solution concentration was 10^{-2} for these three samples to display between 30-300 CFU on the control plate. The CFU displayed with antibiotic treatment is noted below the control group.

	1 mi from hospital	.5 mi from hospital	On site of hospital	
	CFU (Control)	236 CFU (Control)	214 CFU (Control)	286
Suburban Anthropogenic Activity	CFU (AMP)	107 CFU (AMP)	141 CFU (AMP)	200
	CFU (CIP)	108 CFU (CIP)	194 CFU (CIP)	189
	CFU (ERY)	1 CFU (ERY)	23 CFU (ERY)	122

Fig 2 Results of CFU counts of control and experimental plates for the suburban anthropogenic activity samples relative to distance from the hospital. (AMP- Ampicillin, CIP- Ciprofloxacin, ERY- Erythromycin).

	1.0 Mile from Hospital	1 0.5 Miles from Hospital	0.5 On site of Hospital	0		
	Resistance (AMP):	0.45338983	Resistance (AMP):	0.6588785047	Resistance (AMP):	0.6993006993
Suburban Anthropogenic Activity	Resistance (CIP):	0.45762711	Resistance (CIP):	0.9065420561	Resistance (CIP):	0.6608391608
	Resistance (ERY):	0.00423728	Resistance (ERY):	0.1074766355	Resistance (ERY):	0.4265734266
Cumulative Mean Resistance	Sample Resistance:	0.30508474	Sample Resistance:	0.5576323988	Sample Resistance:	0.5955710956
0.48609608						
			Pearson Correlation Coefficient:	-0.919819657		P- Value: .257

Fig 3 Resistance ratios derived from CFU counts of control and experimental plates for suburban anthropogenic activity samples relative to distance from the hospital.

Suburban Anthropogenic Activity Resistance Ratios

The positive trend in average resistance ratios (.30→.55→.60) between the samples signifies an inverse relationship between distance from a hospital and antibiotic resistance (Fig. 3). The Pearson correlation coefficient results in $\sim -.920$, a relatively strong inverse correlation. However, upon calculation of the 0.25 p-value, which is derived from the Pearson correlation coefficient, this suburban anthropogenic activity trend has been determined not to be statistically significant.

The sample resistance represents the average resistance of the antibiotic treated plates. Aside from the resistance trend, the resistance ratio of the soil on site of the hospital is the greatest when compared with the other samples in the suburban level of anthropogenic activity.

Urban Anthropogenic Activity CFU/ Sample Resistance Ratios

The process of analyzing the urban anthropogenic activity sample data will mirror that of the suburban anthropogenic activity sample data. Each of the bacterial solutions were diluted to 10^{-2} to display between

ANTIBIOTIC RESISTANT BACTERIA NEAR PITTSBURGH HOSPITALS

	1 mi from hospital	.5 mi from hospital	On site of hospital
	CFU (Control)	156	278
Urban Anthropogenic Activity	CFU (AMP)	99	128
	CFU (CIP)	90	175
	CFU (ERY)	38	37
		CFU (Control)	197
		CFU (AMP)	122
		CFU (CIP)	179
		CFU (ERY)	64

Fig 4 Results of CFU counts of control and experimental plates for the urban anthropogenic activity samples relative to distance from the hospital.

	1.0 Mile from Hospital	1 0.5 Miles from Hospital	0.5 On site of Hospital
	Resistance (AMP): 0.63461538	Resistance (AMP): 0.4604316547	Resistance (AMP): 0.6192893401
Urban Anthropogenic Activity	Resistance (CIP): 0.576923071	Resistance (CIP): 0.6294964029	Resistance (CIP): 0.9086294416
	Resistance (ERY): 0.24358974	Resistance (ERY): 0.1330935252	Resistance (ERY): 0.3248730964
Cumulative Mean Resistance	Sample Resistance: 0.48504273	Sample Resistance: 0.4076738609	Sample Resistance: 0.6175972927
0.5034379629			
		Pearson Correlation Coefficient:	P- value: .570
			-0.624292703

Fig 5 Resistance ratios derived from CFU counts of control and experimental plate for urban anthropogenic activity relative to distance from the hospital.

30 and 300 CFU. The CFU counts of each urban anthropogenic control plate and experimental plate (Fig. 4) were used to calculate the resistance ratios (Fig. 5).

Urban Anthropogenic Activity Sample Resistance Ratios

Referencing the change in resistance ratio (.49→.62) (Fig. 5) without regard to significance testing, there seems to be an unclear relationship between resistance ratio and distance from the hospital. This increase is due to the resistance ratio first decreasing, then sharply increasing.

Beyond the surface level trend observations, a Pearson correlation coefficient was calculated to be ~0.62, which signifies a relatively casual inverse relationship

between resistance ratio and distance from the hospital. A p-value of .570 was derived from the Pearson correlation coefficient, which nullifies the significance of this trend. Independent from the trend, the resistance ratio derived from the soil on site of the hospital is the greatest relative to the other samples.

Comparison of Anthropogenic Resistance Averages

The sample resistance ratios, based on what category of anthropogenic activity they are included in, were averaged and compared (Fig. 6).

As shown in Figure 6, the urban anthropogenic activity mean resistance ratio is only ~.14% greater than the suburban activity. A two-tailed, independent sam-

Urban Anthropogenic Activity	Suburban Anthropogenic Activity
Cumulative Mean Resistance	Cumulative Mean Resistance
0.5034379629	0.48609608
T Test: t(4)=.87234, p=.432	

Fig 6 The average resistance ratio respective to the level of anthropogenic activity. A t-score and p-value was calculated using the sample resistance ratios.

ANTIBIOTIC RESISTANT BACTERIA NEAR PITTSBURGH HOSPITALS

ple T-test with 4 degrees of freedom was calculated to be $\sim .872$. A p-value of .432 was derived from this t-score, which nullifies the significance in the difference between the anthropogenic averages.

Analysis

Upon analysis of the data points and trends, connections to previous antibiotic resistance trend studies have been made. The data does not adequately support the conclusions of these studies entailing the notion of hospitals contributing to increased antibiotic resistance in the environment within the hospital's influence. However, with the presence of antibiotic resistance discovered in the environment surrounding the subject hospitals, the following conclusions that infer related health risks due to the presence of antibiotic resistant bacteria have been identified although not necessarily proven. The findings regarding differences in antibiotic resistance related to anthropogenic activity do not refute or support the existing data in the field due to statistical insignificance of the observed trends.

The observation that the samples closest to the hospital have a relatively high percentage of antibiotic resistant bacteria, along with the casual (although statistically insignificant) trend of an increasing antibiotic resistance ratio within proximity to a hospital, confirm that the soil environments near a hospital have a prominent presence of antibiotic resistant bacteria. This finding is in line with previous conclusions, stating that proximity to antibiotic use in practice is linked to the "breeding" and contamination of antibiotic resistant bacteria in the subject hospital areas (17, 22). As previously established, there is a possibility that antibiotic resistant bacteria contaminating the soil can act as a reservoir for nosocomial antibiotic resistance and exchange with the circulating pathogens in the hospital (23). Considering this, it is reasonable to expect that this same reservoir exchange dynamic is happening in the subject hospitals. This inference can further imply that the control techniques enacted by the hospital are not potent enough to reduce the level of antibiotic resistance to a level that is consistent with other environments. Overall, these research findings lead to a failure to reject the portion of the null hypothesis that states the presence of an insignificant

trend between proximity to a hospital and antibiotic resistance. However, the observed antibiotic resistance from the onsite samples supports that the subject hospitals, with their current control techniques or lack thereof, are exposed to the same antibiotic resistance exchange and associated health risks demonstrated in the referenced studies due to a considerable percentage of bacteria in the surrounding environment demonstrating resistance.

The minimal 0.14% greater resistance ratio in the urban anthropogenic activity samples is consistent with the expected direct relationship between anthropogenic levels and antibiotic resistance. However, the t-test has proven that this difference is statistically insignificant. Considering the insignificance of the difference between the anthropogenic activity resistance ratios, it is difficult to draw a valid conclusion from this data. The conclusions of related studies describing a positive relationship between the level of anthropogenic activity and the antibiotic resistance present in an environmental sample are relevant for comparison (25, 26). The data discovered, without statistical input, lightly supports the conclusions of these studies. However, due to statistical insignificance, the data can neither refute nor support these existing studies in the field. The portion of the null hypothesis stating that increased anthropogenic activity is significantly related to increased average antibiotic resistance ratio has failed to be rejected.

Limitations

As expected, many experimental limitations exist when working with bacteria (contamination/miscounts). Aside from that, the incomplete knowledge of the conditions of the samples, as well as the design of the experiment, add to the limitations that need to be addressed. Starting with the most common limitations characteristic of the traditional agar plating method, there is a possibility of miscounting CFU and/or contamination from either the instruments used or contaminated air in the lab space. Attempts to avoid contamination included using sterile procedures, using sterile or sanitized instruments, and creating antibiotic and bacterial solutions next to an alcohol lamp. To minimize the possibility of a mis-

count, each sample was counted twice and a marker was used to label CFUs that had already been counted. Additionally, the status of the sample sites prior to collection was unknown. This means that antibiotic resistance factors could potentially enter the sampled environment through vectors from a confounding external source, leading to a possible misrepresentation of the antibiotic resistance stemming from the origin that is the subject of study. Lastly, the scope of the study is inadequate to supply confidence in the trends, due to only six samples being tested. This repurposes the study into a “proof of concept” that can frame future research with a larger sample size.

eliminate the associated confounding variables. Most notably, an increased sample size is necessary to increase the chances of statistically significant trends and the overall merit of future studies.

Conclusions/ Future Research

The hypothesis connoted that bacteria within the soil samples will become increasingly resistant to the select antibiotics with proximity to Pittsburgh hospitals as well as exposure to urban anthropogenic activity (relative to suburban activity). The findings did not support the significance of these trends, largely due to insufficient sample size. However, the study did identify, from prominent concentrations of antibiotic resistance in the soil surrounding the hospital, the potential for antibiotic resistant bacteria to exchange resistance properties with pathogens circulating in the subject hospitals. This can introduce health risks as nosocomial infections have an increased chance to demonstrate antibiotic resistance due to this exchange via horizontal gene transfer. Additionally, the findings provide an idea of the sheer amount of antibiotic resistance existing in the soil environment alone, with ~60% and ~62% of bacteria demonstrating resistance in the suburban and urban hospitals, respectively.

This data addresses the unknown antibiotic resistance status as well as the presence and potential threat of antibiotic resistant bacteria in the city of Pittsburgh and surrounding areas. Improvement upon the limitations of this study have the potential to yield statistically significant trends and, consequently, more precise data. Conducting the experiment in a pressurized lab environment with continuous filtering and access to an autoclave could eliminate the concern surrounding sample contamination. Avoiding the sampling of environments that are suspected to be in the influence of more than the intended origin of resistance would

References

1. Ventola CL. 2015. The antibiotic resistance crisis: Part 1: Causes and threats. *Pharmacy and Therapeutics: a peer-reviewed journal for formulary management.* 40(4): 277.
2. 2021. How antibiotic resistance became an urgent threat. George Washington University, Milken Institute of Public Health. Online resource at <https://onlinepublichealth.gwu.edu/resources/how-antibiotic-resistance-became-an-urgent-threat/>
3. Ayukekbong J, Ntemgwa M, Atabe AN. 2017. Table 1 from the threat of antimicrobial resistance in developing countries: Causes and control strategies. *Antimicrobial Resistance and Infection Control.* 6. DOI:10.1186/s13756-017-0208-x
4. 2021. 2019 antibiotic resistance threats report. Centers for Disease Control and Prevention. Online resource at <https://www.cdc.gov/drugresistance/biggest-threats.html>
5. Kim S, Aga D. 2007. Potential ecological and human health impacts of antibiotics and antibiotic-resistant bacteria from wastewater treatment plants. *Journal of Toxicology and Environmental Health Part B.* 10(8): 599-73.
6. Huijbers PMC, Flach C-F, Larsson DGJ. 2019. A conceptual framework for the environmental surveillance of antibiotics and antibiotic resistance. *Environment International.* 130(2019): 104880.
7. Syvitski J, Waters C, Day J, Milliman J, Summerhayes C, Steffen W, Zalasiewicz J, Cearreta A, Gałuszka A, Hajdas I, Head M, Leinfelder R, McNeill J, Poirier C, Rose N, Shotyk W, Wagemich M, Williams M. 2020. Extraordinary human energy consumption and resultant geological impacts beginning around 1950 CE initiated the proposed Anthropocene Epoch. *Communications Earth and Environment.* 1. DOI:10.1038/s43247-020-00029-y
8. Aslam B, Wang W, Arshad MI, Khurshid M, Muzammil S, Rasool MH, Nisar MA, Alvi RF, Aslam MA, Qamar MU, Salamat MKF, Baloch Z. 2018. Antibiotic resistance: A rundown of a global crisis. *Infection and drug resistance.* 11: 1645-58. DOI:10.2147/IDR.S173867
9. 2021. About antibiotic resistance. Centers for Disease Control and Prevention. Online resource at <https://www.cdc.gov/drugresistance/about.html>
10. Acar J. 1997. Broad- and narrow-spectrum antibiotics: An unhelpful categorization. *Clinical Microbiology and Infection.* 3(4): 395-6. DOI:10.1111/j.1469-0691.1997.tb00274.x
11. Silhavy TJ, Kahne D, Walker S. 2010. The bacterial cell envelope. *Cold Spring Harbor perspectives in biology.* 2(5): a000414. DOI:10.1101/cshperspect.a000414
12. Patolia SK. 2020. Gram stain: Reference range, interpretation, collection and panels.. Medscape. Online resource at <https://emedicine.medscape.com/article/2093371-overview>
13. Burmeister AR. 2015. Horizontal gene transfer. *Evolution, medicine, and public health.* 2015(1): 193-4.
14. Lermينياux NA, Cameron ADS. 2007. Horizontal transfer of antibiotic resistance genes in clinical environments. *Canadian journal of microbiology.* 65(1): 34-44.
15. Luis B. 2009. Horizontal gene transfer in evolution: Facts and challenges. *Proceedings of the Royal Society B: Biological Sciences.* 277(1683): 819-27.
16. Amaha ND, Berhe YH, Kaushik A. 2018. Assessment of inpatient antibiotic use in Halibet national referral hospital using who indicators: A retrospective study. *BMC research notes.* 11(1): 904. DOI:10.1186/s13104-018-4000-7
17. Struelens MJ. 1998. The epidemiology of antimicrobial resistance in hospital acquired infections: Problems and possible solutions. *The BMJ.* 317(7159): 652-4. DOI:10.1136/bmj.317.7159.652
18. Weinstein RA. 2001. Controlling antimicrobial resistance in hospitals: Infection Control and use of antibiotics. *Emerging infectious diseases.* 7(2): 188-92. DOI:10.3201/eid0702.010206
19. Larson E. 1999. Skin hygiene and infection prevention: More of the same or different approaches? *Clinical infectious diseases.* 29(5): 1287-94. DOI:10.1086/313468
20. Smith DL, Dushoff J, Perencevich EN, Harris AD, Levin SA. 2004. Persistent colonization and the spread of antibiotic resistance in nosocomial pathogens: Resistance is a regional problem. *Proceedings of the National Academy of Sciences of the United States of America.* 101(10): 3709-14.
21. 2016. Superbugs Threaten Hospital patients. Centers for Disease Control and Prevention. Online resource at <https://www.cdc.gov/media/releases/2016/p0303-superbugs.html>
22. Polianciuc SI, Gurzău AE, Kiss B, Ștefan MG, Loghin F. 2020. Antibiotics in the environment: Causes and consequences. *Medicine and pharmacy reports.* 93(3): 231-40. DOI: 10.15386/mpr-1742
23. Lee JH, Park KS, Jeon JH, Lee SH. 2018. Antibiotic resistance in soil. *The Lancet: Infectious Diseases.* 18(12): 1306-7. DOI: 10.1016/S1473-3099(18)30675-3
24. Tenover F. 2004. Novel and emerging mechanisms of antimicrobial resistance in nosocomial pathogens. 91(3B): 76S-81S. DOI: 10.1016/0002-9343(91)90347-z
25. Scott LC, Wilson MJ, Esser SM, Lee NL, Wheeler ME, Aubee A, Aw TG. 2021. Assessing visitor use impact on antibiotic resistant bacteria and antibiotic resistance

ANTIBIOTIC RESISTANT BACTERIA NEAR PITTSBURGH HOSPITALS

- genes in soil and water environments of Rocky Mountain National Park. *Science of the Total Environment*. 785: 147122. DOI: 10.1016/j.scitotenv.2021.147122
26. Scott LC, Lee N, Aw TG. 2020. Antibiotic resistance in minimally human-impacted environments. *International Journal of Environmental Research and Public Health*. 17(11): 3939. DOI: 10.3390/ijerph17113939
27. Wang F, Stedtfeld R, Kim O, Chai B, Yang L, Stedtfeld T, Hong S, Kim D, Lim H, Hashsham S, Tiedje J, Sul W. 2016. Influence of soil characteristics and proximity to Antarctic research stations on abundance of antibiotic resistance genes in soils. *Environmental Science and Technology*. 50(23): 12621-9. DOI: 10.1021/acs.est.6b02863
28. Pagan FS. 1981. Antibiotics for gram-positive organisms. *British journal of hospital medicine*. 25(1): 24-7.
29. 2021. Erythromycin: Compound Summary. National Center for Biotechnology Information. U.S. National Library of Medicine. Online resource at <https://pubchem.ncbi.nlm.nih.gov/compound/erythromycin>
30. Exner M, Bhattacharya S, Christiansen B, Gebel J, Goroncy-Bermes P, Hartemann P, Heeg P, Ilschner C, Kramer A, Larson E, Merkens W, Mielke M, Oltmanns P, Ross B, Rotter M, Schmithausen RM, Sonntag HG, Trautmann M. 2017. Antibiotic resistance: What is so special about multidrug-resistant gram-negative bacteria? *GMS hygiene and infection control*. 12: doc5. DOI: 10.3205/dgkh000290
31. Thai T, Salisbury BH, Zito PM. 2021. Ciprofloxacin. Treasure Island, Florida: StatPearls Publishing.
32. Soares GMS, Figueiredo LC, Faveri M, Cortelli SC, Duarte PM, Feres M. 2012. Mechanisms of action of systemic antibiotics used in periodontal treatment and mechanisms of bacterial resistance to these drugs. *Journal of applied oral science*. 20(3): 295-309. DOI: 10.1590/s1678-77572012000300002
33. Rafailidis PI, Ioannidou EN, Falagas ME. 2018. Ampicillin/sulbactam: Current status in severe bacterial infections. *Drugs*. 67(13): 1829-49. DOI: 10.2165/00003495-200767130-00003
34. Lima-Bittencourt CI, Cursino L, Gonçalves-Dornelas H, Pontes D, Nardi R, Callisto M, Chartone-Souza E, Nascimento A. 2007. Multiple antimicrobial resistance in Enterobacteriaceae isolates from pristine freshwater. *Genetic and Molecular Research*. 6(3): 510-21.
35. Cycoń M, Mrozik A, Piotroska-Seget Z. 2019. Antibiotics in the soil environment-degradation and their impact on microbial activity and Diversity. *Frontiers in microbiology*. 10: 338. DOI: 10.3389/fmicb.2019.00338

Application of Data Envelopment Analysis in a Small-Scale Transportation Network

Alexandra McWatters

Further optimizing small-scale bus networks has proven to be an obstacle in linear programming over the previous decade. Data Envelopment Analysis (DEA) is a mathematical approach designed to organize a system's inputs and outputs and determine the efficiency of its decision-making units. Existing studies show that DEA is a logical approach for network planning in larger, highly-urbanized areas, yet there remains a gap in its effectiveness in smaller, more suburban settings. To explore this inquiry, five DEA models were designed and applied to 19 bus routes operating through Delaware, USA, and their results were statistically analyzed. Findings suggest that an individual can implement DEA in virtually any environment, regardless of size and population dispersion, as this study yielded results ranging from 28% to 100% efficiency, consistent with similar studies; therefore supporting the technique's future applicability and success.

Keywords: bus network, data envelopment analysis, excel solver, linear programming, optimization model, transportation efficiency

Application of Data Envelopment Analysis in a Small-Scale Transportation Network

Network optimization is developing into a vital component of the transportation industry. As identified and reported by Lao and Liu (2009), the application of recent findings has become an integral and crucial component (p. 247) of local systems. Due to this operation being largely comprised of strategic planning techniques, Deng and Yan (2019) point out that required optimization methods can be complex, yet vital to transit efficiency (p. 2).

Further, examining the environment in which a bus network operates is imperative to developing an optimal mathematical model. Zhang et al. (2015) present

a strong argument on how the regional characteristics of an area significantly contribute to traffic dynamics and affect the performance of the chosen model. In many studies consistent with the article authored by Lao and Liu (2009), experts identify yet discount attributes such as route network, population, and travel patterns of an area (p. 247) when computing and analyzing results.

All previously published works do not simultaneously account for both spatial and mathematical elements and have solely been tested on a large scale. Thus, there exists a gap in the knowledge regarding the effectiveness of specific models in unexplored settings. As evidence of geographic differences, Elliott et al. (2013) recognize the fact that some high-population areas require riders to transfer between multiple lines, whereas others do not. Utilizing a method that acknowledges all relevant criteria is an essential step

toward most efficiently networking a small system of bus lines.

Literature Review

Data Envelopment Analysis

Omidi et al. (2020) describe Data Envelopment Analysis (DEA) as a computational approach for managing a system's decision-making units (DMUs), or input and output values. As Deng and Yan (2019) explain, DMUs are resources that produce the final goods or services within a model. To determine the extent of a unit's efficiency, Tavana et al. (2013) write that one must identify if "other DMUs can produce more outputs using an equal or lesser amount of inputs" (p. 502). If this is the case, the system is not in its optimal state. These units account for multiple performance measures and are the best choice for more intricate efficiency problems.

Deng and Yan (2019) recognize that DEA can be used in a wide variety of applications, ranging from education and production to banking. However, Toloo and Nalchigar (2008) reveal that in recent years, the transportation industry too began relying on DEA principles to plan effective networks due to their successful applications in existing studies (p. 598). For example, Lao and Liu (2009) note that the combination of the data mapping system, Geographic Information Systems (GIS), and DEA principles has generated approaches "used to integrate data input, spatial analysis, and visualization" (p. 254). Findings from studies like these have led to improvements in the specialization of DEA.

Crucial Components

Indicators are a significant component of Data Envelopment Analysis. Deng and Yan (2019) specify that these facets account for the complexity and generate a systematic outlook on commuting networks (p. 2) by defining specific criteria on which the program should focus. Quality examples of input indicators include operation time and number of stops on a route, whereas a common output indicator is the mean ridership, or number of passengers, on a bus. As summarized by Lao and Liu (2009), these constituents

"[reflect] the overall objective" (p. 250) of the system, therefore reinforcing the necessity for carefully selected indicators.

Radial and non-radial characteristics bear distinct disparities from one another, and it is imperative to account for these differences. Tone and Tsutsui (2010) delineate that radial measures are concerned with parallel reduction of input values (p. 1), suggesting that an upsurge in the weight of an input indicator would result in a proportionate rise in output. However, non-radial measures do not consider a correlation between input and output indicators. In transit applications, Hahn et al. (2011) attest to the fact that experts should implement both types in conjunction, as the inputs and outputs do not necessarily change proportionally.

Common Models

A description of common DEA models is essential as they differ slightly in regards to procedure. The Charnes, Cooper, Rhodes (CCR) model is a type of radial DEA application that can predict the final input and output values of a DMU (Chin Tsai et al. 2009, p. 338) using past data and trends. Known as the most efficacious and favored DEA model, CCR is concerned with the equivalent reduction of variables. Tone and Tsutsui's study uses labor, materials, and capital as inputs; however, because these indicators do not transform proportionally, the CCR model would be impractical and fruitless. Toloo and Nalchigar (2008) deduce that the necessity for additional standards has resulted in many extensions, notably BCC (p. 598).

Likewise, the Banker, Charnes, Cooper (BCC) model is a radial DEA application. However, the distinction is that the CCR model assumes a constant return to scale, whereas the BCC model presents a variable return to scale alternative, to which Gofran et al. (2021) give attention. Deng and Yan (2019) explain that variable return to scale implies "efficiency may increase or decrease with a change of size in input or output" (p. 3). Therefore, this suggests, that BCC is applicable for bus lines with more significant productivity variability than CCR (Lao and Liu 2009).

The third and final classification of DEA standards is Slack-Based Measures (SBM). This model "accounts for the non-radial slacks which are not considered in the radial models" (Tone and Tsutsui 2010, p. 4).

Engineers use slacks, the disproportionate improvements needed for a variable to become efficient, in applications where it is necessary to allow independent change (Tavana et al. 2013). Due to the loss of the SBM model's proportionality in indicators, teams are developing new radial/non-radial combination models (Deng and Yan 2019). By understanding the advantages and disadvantages of each major primary type, one can pose a rational justification for the chosen methods.

Multi-Stage Applications

When referencing complex multi-stage mathematical models, Tavana et al. (2013) conclude that it is crucial to comprehend the notion of intermediate outputs and inputs, indicators introduced between stages of a model. Radial applications overlook the inclusion of intermediate developments that necessitate different solutions. These indicators represent interior and linking activities within DMUs (Deng and Yan 2019, p. 8) and consequently produce distinct results. For example, Hahn et al. (2011) chose travel speed and vehicle type as intermediate outputs. In contrast, Deng and Yan (2019) use employment coverage, bus connectivity, and residential coverage. This strategy has received an abundance of transit-related attention over recent years (Hahn et al. 2011) because it can assist in tackling formerly foreign components of public transportation (p. 1120).

Finally, the affinity index gauges the similarities between measured categories. Tone and Tsutsui (2010) summarize that this variable "represents the diversity or the scattering of the observed database" (p. 23). In regards to real-world application, a multi-stage networking problem was successfully solved by combining radial and non-radial characteristics using the affinity index in a case study located in Nanjing, China (2019). However, readers cannot use this research to suggest new results due to discrepancies in tested settings.

Analysis of Efficiency

The applied software analyzes the conclusive effectiveness of sampled DMUs, interpreted using an efficiency score calculated by dividing the weighted sum of outputs by inputs. This value is formatted as a

decimal between zero and one. Ghofran et al. (2021) express that higher score can be associated with a more efficient unit. However, it is imperative to concede that even DMUs with high operational efficiency scores may lack spatial awareness (Lao and Liu 2009). Likewise, routes that are geographically aware may lack operational effectiveness.

Proposed Research

Therefore, this research desires to answer the question: "To what extent can Data Envelopment Analysis be used to further optimize a small-scale bus network in Delaware, USA?" The objective is to enhance the prior investigation on optimization approaches and resolve if the DEA strategy applied to metropolitan locations is feasible for use in a dissimilar setting. A plan must establish a comprehensive viewpoint for the most valid deduction. When crafting a hypothesis for this exact question, evidence supports both anticipated success and failure.

However, an analysis of scholarly publications suggests that the model has a low chance of effectiveness in a new setting due to differences in user intention between locations. In urban China, Liu and Lawell (2015) discover that public passenger load has a major effect on the purchase of civilian vehicles (p. 28), consequently resulting in user dependency on public transit as their primary means of transportation. In Northern Delaware, however, this is not the case. A contrasting factor of this sort may cause failure to implement DEA in new settings. Regardless of the result, testing this process using public transportation data is crucial for advancing knowledge and ensuring a bridge between mathematical and spatial analyses.

Methods

This study sought to dissect the effectiveness of Data Envelopment Analysis, specifically a CCR model, on a small-scale transportation network in Delaware, USA, via a quantitative approach. This alternative was most appropriate as the study's primary measurement instrument was an online software devised to output a series of numerical values corresponding to the chosen inputs. Once the system received this information, efficiency scores were displayed on a scale from zero to

one; the most optimal routes have scores of one (Ghofran et al, 2021, p. 102). These data points illustrated the overall efficiency of particular transportation lines as part of a whole. This data was then utilized for the preponderance of the interpretation of results, making a quantitative approach the most suitable.

When addressing research methods, a traditional case study was implemented to investigate the efficiency of an array of bus lines. This specific course was most fitting because it concentrated on a relatively slim sample of measured units. In this instance, the investigation examined 19 individual bus lines on a designated index of applicable criteria. This method allowed for a theory to be tested within two specific constraints: area and time. The identified bus lines, labeled as urban or combo, were monitored in a specified area of Northern Delaware between December 2021 and early March 2022. Variables were computed and analyzed using Excel Solver, a downloadable software extension to Microsoft Excel.

The original research that this study drew inspiration from was tested in Nanjing, China, using a case study. Deng and Yan (2019) sample 410 individual bus lines in a complicated urban bus network. Hence, it could be inferred that a case study on a comparably more diminutive scale, with less than 20 bus lines spanning a smaller area, would be the most proper choice. Deng and Yan obtain adequate and interpretable results using their chosen method, and theoretically, the same should have been achieved by a case study in this particular setting. Their methods equip “comparisons between descriptive statistics of the top and bottom performers” (para 1). In addition, inefficiency results in a lack of identification of bottom performers for interpretation or selection of forthcoming models. This academic support proves that a quantitative case study was optimal for obtaining the necessary results.

The study appointed 19 bus routes in the New Castle County Delaware Authority for Regional Transit (DART) network. These “urban” or “combo” paths had to be considered “fixed,” meaning they operated on a set schedule using predetermined stops. An urban roadway is characteristic of a town or city, while a combo route contains both metropolitan and suburban attributes. Specifically, lines 2, 4, 5, 6, 8, 9, 10, 11, 13, 14, 15, 18, 25, 28, 31, 33, 35, 40, and 51 were employed for data compilation.

Internal, external, and measurement validity were ensured to harvest accurate results amidst data collection. Regarding internal validity, all data points were received verbatim from a representative at the Delaware Department of Transportation or their authorized website. Additionally, the exact measurement procedure was preserved throughout the data collection and modeling phase without introducing additional variables or constraints. This consistency validates the reliability of the study’s cause-and-effect relationship. External validity was also demonstrated. The employed steps could be replicated in any environment with access to Microsoft Excel and ride data. Finally, measurement validity refers to the desktop computer in which data was processed, as it was the primary mechanism involved with computations. Criterion validity was preserved as the conclusive results accurately gauge route efficiency according to industry standards on a zero to one scale. Content validity was also present as there was a linear association between variables, meaning the input values are representative of what was being calculated. Through all of these actions, it can be extrapolated that the measures taken in this study are proper.

The data implemented in the five DEA models came from various origins; specifically, eleven variables were required to calculate each bus line’s total efficiency. These included: the number of stops, route length, on-time arrival rate, number of buses, daily operation time, frequency, overlapping, route directness, connectivity, operation speed, and annual average daily ridership for all 19 routes.

- **Number of Stops**

This value was obtained directly from the Department of Transportation.

- **Route length**

Using GIS, the total distance from end to end of a path could be calculated. This was measured in miles.

- **On-time arrival rate**

This value was obtained directly from the Department of Transportation. This percentile was then converted to a less-the-better format. The most efficient routes had scores close to zero, while less efficient routes had higher scores.

- **Number of buses**

This value was obtained directly from the Department of Transportation.

- **Daily Operation Time**

This value was measured in hours. It was derived from the difference between the departure time of the first bus and the arrival time of the last bus each day. This information was collected from public schedules for each route.

- **Frequency**

Frequency was determined using the number of buses and operation time for each route. Specifically, the total number of buses for a line was divided by the operation time in hours (vehicles/hour).

- **Overlapping**

Overlapping is a quality exhibited when a connection between two stops on a route can be covered by another route. The higher the overlap, the less demand for the route. This value was calculated with GIS using the coordinates obtained from the Department of Transportation.

- **Route Directness**

Route directness equals the route length in one direction divided by the Euclidean distance between the endpoints.

- **Connectivity**

Each route's connectivity score could be calculated using GIS. The preliminary score equals "one" for every stop within 300 meters of another bus route. The final bus connectivity score equals the preliminary connectivity scores divided by the number of stops on each route.

- **Operation Speed**

This value was obtained directly from the Department of Transportation.

- **Annual Average Daily Ridership**

This value was obtained directly from the Department of Transportation.

By employing these variables, five DEA models were developed and executed. Each route was given a score from zero to one for each set of parameters, according to the sum of all decision-making units. The following input/output combinations were utilized:

- **Model 1**

Number of Stops & Route Length (Inputs) / On-time Arrival Rate (Output)

- **Model 2**

Number of Buses & Daily Operation Time (Inputs)

/ Frequency (Output)

- **Model 3**

Bus Overlapping & Route Directness (Inputs) / Connectivity (Output)

- **Model 4**

Number of Stops & Bus Overlapping (Inputs) / Operation Speed (Output)

- **Model 5**

Frequency & Connectivity (Inputs) / Annual Average Daily Ridership (Output)

A wide variety of infographics were generated to display the findings for straightforward comprehension. Though conclusions can be drawn from the numerals themselves, graphs and tables were also designed, as the general audience may have difficulty comprehending this concept.

Results

By concentrating on the extent to which Data Envelopment Analysis can further optimize a small-scale transportation network in Delaware, USA, efficiency scores were obtained on a scale from zero to one. Values equal to one indicated the most efficient routes, whereas lower scores indicated less than optimal routes. Five separate two-input/one-output models, each calculating efficiency concerning distinct factors, were calculated and averaged to obtain one comprehensive score.

To obtain conclusive results, efficiency scores were first calculated for every individual model. Using Excel Solver, each DEA application assigned the most optimal route a perfect score of 100% efficiency (1.0). The remaining 18 routes were assigned a score below this benchmark. For Model 1, results ranged from 0.1575 to 1.000. Model 2 yielded results from 0.0867 to 1.000. Model 3 produced generally lower results between 0.0323 and 1.000. Model 4 ranged from 0.0603 to 1.000. Lastly, Model 5 measured efficiency as low as 0.1359 and as high as 1.000.

DATA ENVELOPMENT ANALYSIS IN A SMALL-SCALE TRANSPORTATION NETWORK

Table 1
Efficiency Scores for Each Model When Applied to All 19 Routes

Route (#)	Model 1	Model 2	Model 3	Model 4	Model 5	Total	% Efficient
Concord Pike (2)	0.5466	0.7000	0.1338	0.0960	0.7517	2.2281	64.56 %
W. 4th Street / Governor Printz Blvd. (4)	0.3503	0.7048	0.0580	0.0836	0.6539	1.8506	53.62%
Maryland Avenue / Christiana Mall (5)	0.3923	0.6933	0.2827	0.9796	0.7928	3.1407	91.00 %
Kirkwood Highway (6)	0.3596	0.6857	0.0672	0.1185	0.8884	2.1194	61.41 %
8th Street / 9th Street (8)	1.0000	0.8000	0.0878	0.0603	0.1447	2.0926	60.63 %
Boxwood Rd. / Broom St / Vandever Ave. (9)	0.5678	0.7111	0.1265	0.0988	0.5663	2.0705	59.99 %
Wilmington / University Plaza / Newark (10)	0.2108	0.7333	0.0323	0.0976	0.2632	1.3372	38.74 %
Washington St. / Arden (11)	0.4932	0.7333	0.1212	0.1306	0.4633	1.9416	56.25 %
Philadelphia Pike / DuPont Hwy (13)	0.1575	0.6667	0.0426	0.1188	1.000	1.9856	57.53 %
Baylor / DHSS / Miller Rd. Shopping Center (14)	0.3480	0.7333	0.0589	0.0687	0.3314	1.5403	44.63 %
New Castle Avenue / Basin Rd. / Christiana Mall (15)	0.2038	0.7333	0.1545	0.5522	0.7722	2.4160	70.00 %
Pike Creek / Foulk Rd. (18)	0.1733	0.9067	0.0701	0.1357	0.1791	1.4649	42.44 %
Miller Rd. / Wilmington / DuPont Highway (25)	0.2470	0.6667	0.1415	0.2352	0.4038	1.6942	49.09 %
A.I. DuPont / Wilm. DMV / Probation & Parole / Riveredge Industrial (28)	0.2080	0.9778	0.0770	0.0751	0.1654	1.5033	43.56 %
Market St. / Philadelphia Pike (31)	0.3567	0.9333	0.0877	0.0773	0.2680	1.7230	49.92 %
Christ. Mall / Newark (33)	0.2374	0.7000	0.1485	0.3229	0.7375	2.1463	62.19 %
Brandywine Town Center / Shipley Road (35)	0.2156	1.000	0.0895	0.1525	0.1359	1.5935	46.17 %
Glasgow / US Routes 13 & 40 / Wilmington (40)	0.3415	0.6933	1.000	1.000	0.4166	3.4514	100.00 %
New Castle Ave. / DE 273 / Christiana Mall (51)	0.2208	0.0867	0.0669	0.1116	0.4922	0.9782	28.34 %

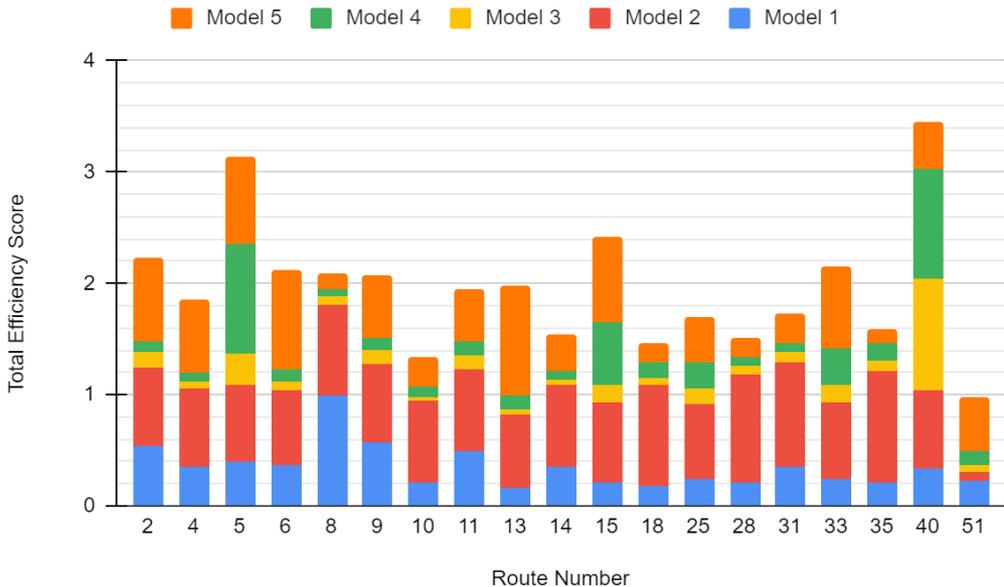
Once these values were obtained (Table 1), they could be analyzed to determine the overall efficiency of each route. A definitive score out of five was calculated by taking the sum of each proportionally weighted output. Additionally, a total score between zero and one was computed by averaging the values, with one route obtaining a flawless score and all others receiving lower. These analyses determined that Glasgow / Route 40 was the most efficient, while New Castle Avenue / DE 273 was the least efficient.

By examining this bar graph (Figure 1), it could be determined which models were prevalent in calculating the efficiency of each route. This was crucial as the exact numerical value does not account for routes highly optimized in terms of frequency but less so in terms of connectivity, for example. That score was used as a general, summarized value. By breaking down these individual scores, each route’s more intricate strengths and weaknesses can be seen, therefore assisting in identifying areas for improvement.

Additionally, it was worth noting that some models yielded more diverse efficiency scores, whereas others were clumped towards the bottom of the specified range. This impacted the average final efficiency score. By taking this fact into account, it can be determined that some models were more influential to final efficiency than others. For example, Model 3 produced generally lower results, grouped around the minimum of 03% efficiency, yet was still required to assign a perfect score for a single route. Annual Average Daily Ridership had the highest impact on overall efficiency, whereas connectivity impacted efficiency the least.

Overall, this study’s results show a diverse range of route efficiency scores in on-time arrival rate, frequency, connectivity, operation speed, and annual average daily ridership. In this way, it was suggested that DEA could be successfully implemented in a small-scale transportation network. Regarding overall efficiency, Glasgow / Route 40, classified as “combo,” was deemed the most optimal route.

Figure 1
Individual Model Weight on Determining Overall Efficiency for All 19 Routes



Discussion

By exploring the extent to which one can implement Data Envelopment Analysis on a small-scale transportation network in Delaware, USA, the researcher drew relevant conclusions. This research gap necessitated the use of a quantitative case study between December 2021 and February 2022. This was the most optimal choice of method because it permitted an in-depth analysis of a designated system under specified time and location constraints. Essentially, by using a series of models consisting of different input/output combinations, researchers obtained scores on a zero to one scale for each route. These specific models measured efficiency in regards to on-time arrival rate, frequency,

connectivity, operation speed, and annual average daily ridership. The software generated a more complete result by calculating the sum of each route's various efficiency scores. Route 40, Glasgow / US Routes 13 & 40 / Wilmington, was the most efficient, with a 100% efficiency score. On the other hand, route 51, New Castle Ave. / DE 273 / Christiana Mall only yielded a 28.34% efficiency score, therefore performing least optimally. The remaining 18 subjects fell between these two values, supporting the idea of a broad final data envelope, characteristic of a successful DEA application.

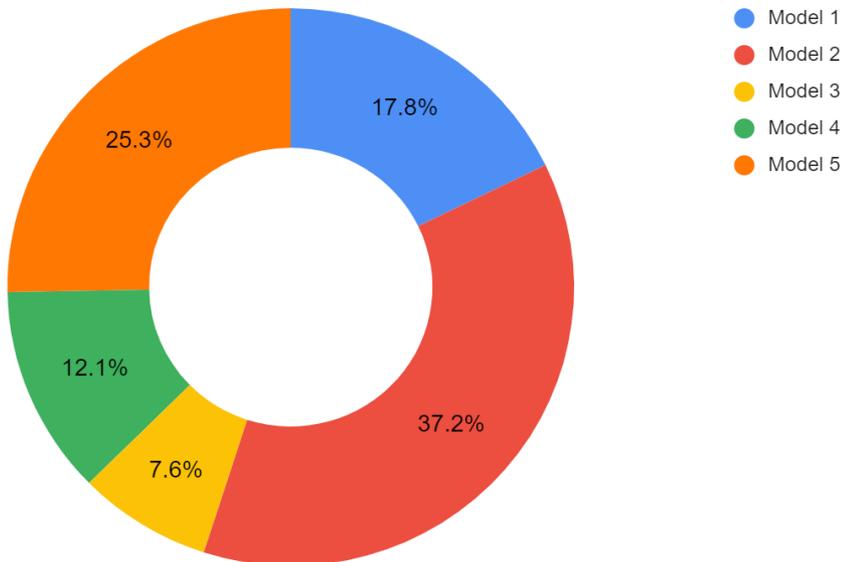
Significance of Results

These results indicate that Data Envelopment Analysis can be a

basis for optimization measures and meaningful conclusions in the transportation industry. In this scenario, findings directly correlate to successful application in a small-scale network. Though DEA has been deemed influential in transportation systems with many routes, it was unknown whether this technique would be valuable with smaller sample sizes.

For example, in Nanjing, China, experts analyze hundreds of routes, each with similar characteristics: urban and large-scale. In other non-transportation-related DEA applications, numerous identifiers are utilized for DMUs. However, this study is distinctive because only 19 routes were manipulated, so it was unclear whether or not the results would be comprehensive enough for in-

Figure 2
Most Influential Models on Overall Average Efficiency Scores



dustry standards. However, these conclusions suggest that DEA is successful in both broader and narrower applications.

This realization is significant because it supports further research on the topic and suggests the accuracy of future results. With improved efficiency in suburban areas, numerous stakeholders will profit. Transportation planners, businesses, and public-transit users can benefit from this growing industry from an economic, environmental, and social standpoint.

Alternative Explanations

Though this research suggests that DEA could be valid in any small-scale application, there may be two alternative explanations driving this result. The researcher compiled the values utilized in this case study in 2022 amidst the COVID-19 pandemic. As a result, there may have been deviations in transportation trends from what experts define as “typical.” However, the thoroughness of the data from 2019, considered “pre-pandemic,” was not adequate for this purpose. As a result, recent data was used. Thus, there are potential drawbacks regarding the results’ exactitude. These findings are correct, however, to the state of transportation as it stands in the present day. Additionally, this study intended to determine if one could apply DEA to a suburban environment. Regardless of how factual the final numerical values are, it can still be suggested that the models were effective based on the resulting envelope and range of final efficiency scores.

Secondly, some may express concern with the accuracy of results due to complications with numerical rounding or the vast number of manipulations made to the data. This is not an issue, as the researcher maintained at least four significant in every calculation, consistent with the industry standard. In addition, precautionary measures were taken to prevent mathematical errors. Though alternative explanations could be presented, one can conclude that the results were reasonably consistent with what was being measured.

Implications and Limitations

Though the research answered the inquiries posed, limitations arose that prevented all aspects of the study from being flawless. The computer only created and applied five DEA models to determine a final efficiency score. Though this set is more thorough than a single two-input/one-output model, additional components would enhance the mathematical processes’ overall precision and result in a more comprehensive resolution. This limit on the number of models was due to accessibility to information and time constraints.

Additional factors undoubtedly influence the overall efficiency of a bus route; however, that data was not feasible to obtain because of lack of access to information. For example, harsh weather conditions, extreme traffic, and reckless drivers impact the speed at which a vehicle travels between two points. Though this study did not account for all factors, the researcher used the most essential DMUs, ensuring an adequate, generalized conclusion. The same is valid with time constraints. Though a longer data collection window may have allowed more developed models or analyzed routes, that was not viable. If the researcher extended this study, more time would be allotted to derive thorough results and eliminate all identified alternative solutions. A researcher can confidently infer that DEA will achieve success with other smaller samples but cannot confirm this with complete certainty.

Suggestions for Further Research

There remains an opportunity to research this specific subject further. The research team would conduct supplemental investigation to both clear up previous uncertainties and elaborate on existing knowledge. Specifically, it is critical to remodel this study with a broader range of input/output identifiers, including, but not limited to, traffic density and weather conditions. This specification will clarify if DEA is applicable in more refined hypothetical situations rather than broad ones. In addition, further study the relationship between DEA success and choice of route settings, including rural areas, may be beneficial. A gap remains in this section of public transport. Further studies could draw a more precise conclusion between DEA’s overall success and sample size.

Conclusion

This study addressed the narrow previous application of DEA in the transportation industry by conducting a case study using a small sample size of bus routes. These findings have introduced new knowledge regarding how mathematical modeling techniques can be further employed to optimize various network scenarios. Experts in the domain can use these conclusions to defend the accuracy of their efficiency calculations within the body of literature. Specifically, these conclusions give professionals reason to believe that DEA can be an effective, practical measure in various geographic areas, regardless of sample size, as the final data envelope is extensive and consistent with those of similar works.

References

- Chin-Tsai, Shih-Nan, & Chia-Ho. (2009). A forecasting model of data envelopment analysis -- Integrate GM(1,1) with CCR model. *Journal of Grey System*. <https://search.ebscohost.com/login.aspx?direct=true&db=asn&AN=47572173&site=ehost-live>
- Deng, Y., & Yan, Y. (2019). Evaluating route and frequency design of bus lines based on data envelopment analysis with network epsilon-based measures. *Journal of Advanced Transportation*. <https://www.hindawi.com/journals/jat/2019/5024253/>
- Elliott, J., Jayachandran, H., Kumar, P., & Metzger, K. (2013). Campus shuttle: Design of a college campus parking and transportation system. *IEEE Systems and Information Engineering Design Symposium*. https://catsr.vse.gmu.edu/SYST490/495_2013_ParkTransSystem/Design_of_a_College_Campus_Parking_and_Transportation_System.pdf
- Ghofran, A., Sanei, M., Tohidi, G., & Bevrani, H. (2021). Applying MCDEA models to rank decision making units with stochastic data. *International Journal of Industrial Mathematics*. <https://search.ebscohost.com/login.aspx?direct=true&db=asn&AN=148693796&site=ehost-live>
- Hahn, J. S., Kim, H. R., & Kho, S. Y. (2011). Analysis of the efficiency of Seoul arterial bus routes and its determinant factors. *KSCE Journal of Civil Engineering*. https://www.researchgate.net/publication/225864310_Analysis_of_the_efficiency_of_Seoul_Arterial_Bus_routes_and_its_determinant_factors
- Lao, Y. & Liu, L. (2009). Performance evaluation of bus lines with data envelopment analysis and geographic information systems. *Computers, Environment and Urban Systems*. https://www.researchgate.net/publication/223498892_Performance_evaluation_of_bus_lines_with_data_envelopment_analysis_and_geographic_information_systems
- .com/login.aspx?direct=true&db=asn&AN=47572173&site=ehost-live
- Liu, Q., & Lawell, C. L. (2015). The effects of public transportation and the built environment on the number of civilian vehicles in China. *Shandong University of Finance and Economics*. http://clinlawell.dyson.cornell.edu/China_transport_built_envir_paper.pdf
- Omidi, M. R., Omidi, N., & Sharabiani, A. M. (2020). Evaluation of decision-making units in reducing traffic accidents using data envelopment analysis. *Health in Emergencies & Disasters Quarterly*. <https://doi.org/10.32598/hdq.5.2.222.7>
- Tavana, M., Mirzagoltabar, H., Mirhedayatian, S., Saen R., & Azadi M. (2013). A new network epsilon-based DEA model for supply chain performance evaluation. *Computers & Industrial Engineering*. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.669.1625&rep=ep1&type=pdf>
- Toloo, M., & Nalchigar, S. (2008). A new integrated DEA model for finding most BCC-efficient DMU. *Applied Mathematical Modeling*. <https://www.sciencedirect.com/science/article/pii/S0307904X08000395>
- Tone, K., & Tsutsui, M. (2010). An epsilon-based measure of efficiency in DEA—a third pole of technical efficiency. *European Journal of Operational Research*. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.614.8699&rep=ep1&type=pdf>
- Zhang, H., Zhao, P., Wang, Y., Yao, X., & Zhuge, C. (2015). Evaluation of bus networks in China: From topology and transfer perspectives. *Discrete Dynamics in Nature and Society*. <https://www.hindawi.com/journals/ddns/2015/328320/>

Investigation on the Effect of Infill Orientation on the Flexural Properties in FDM Parts

Alexander Barry

Parts made with Fused Deposition Modeling (FDM) are anisotropic, meaning that their properties are dependent on their orientation and the direction of forces. This study aimed to investigate how the infill orientation in FDM parts influences their flexural properties. Samples were created with either a 0° or 45° infill orientation and evaluated using a 3-Point Bending test. Samples with a 0° infill orientation had a brittle failure mode, lower coefficient of variation, higher yield strength, and higher flexural modulus, while samples with a 45° infill orientation had a ductile failure mode and a higher maximum force value. Samples with a 45° infill orientation had a higher flexural strain at failure, but samples of both infill orientations had similar flexural strength. This suggests that while both orientations result in similar flexural strengths, the 0° infill orientation is largely superior to the 45° infill orientation in terms of flexural properties.

Keywords: Fused deposition modeling (FDM), anisotropy, infill orientation, flexural properties

Additive Manufacturing (AM) is rapidly taking over many manufacturing industries, such as aerospace and robotics, due to its ability to create complex objects in a timely and cost-efficient manner. There are multiple types of AM, the main ones being Fused Deposition Modeling (FDM), stereolithography, and selective laser sintering [1]. FDM, also known as 3D printing, involves placing melted filament down in a path that cools and hardens, building the part layer by layer. FDM is cost-effective and among the most widespread AM processes [1], [2]. The reason for this cost-effectiveness is FDM's three main infill parameters: pattern, density, and orientation [3]. The infill pattern is a preset layout that determines the internal design of the part [4], [5], [6], [7]. The infill density is the area of the internal section for an FDM part that will be filled [3]. The infill orientation controls the angle at which the pattern is printed [5]. Due to the lack

of uniformity, FDM parts are anisotropic, meaning they vary in strength in different orientations [5]. Although anisotropy is a principal quality of FDM parts, there has been no clear conclusion on how changing the infill orientation affects the flexural properties created when printing the filament.

I. Literature Review

A. Effect of Printing Parameters

Previous studies have investigated various printing parameters' effects on FDM parts. Some parameters directly affect how the filament is deposited, such as printing speed and temperature [8]-[10]. Higher printing speeds increase dimensional inaccuracies and weld lines, and in turn decrease structural integri-

ty and strength [8], [9]. Higher printing temperatures lead to fewer dimensional inaccuracies and improved strength, due to the filament fusing better when more melted [8]. Additionally, other parameters, such as layer thickness and air gaps, control the characteristics of the print. Increasing layer thickness and air gap size decreases the precision of the part, and air gaps can lead to structural weaknesses [5], [8], [11].

B. Effect of Infill Parameters

Similarly, the three primary infill parameters have a large effect on the part's mechanical properties. Studies have found that higher infill densities generally result in higher mechanical strength and increased brittleness, but also raise manufacturing costs with increased print times and total filament usage [3]-[5], [8], [12]-[16]. While the infill pattern attribute has a negligible impact on cost, many sources have affirmed that it has a significant impact on a part's mechanical properties [4], [6], [12], [13], [17]. Although Gonabadi et al. [12] concluded that the infill pattern has little effect on tensile strength or Young's modulus, numerous other studies have challenged that conclusion, generally finding that the hexagonal and grid patterns are stiffer and have higher tensile strengths while triangular patterns have higher Young's moduli [4], [13], [17], [18]. Also, provided that the 0° orientation aligns with the direction of applied force, the mechanical properties generally decrease the more the infill orientation deviates away from 0° [17], [19]. Many studies have called attention to this anisotropy in all parts manufactured with FDM [5], [12], [15], [17], [19], [20].

C. Anisotropy in FDM Parts

While FDM parts' anisotropic nature has been identified, there remains a need for additional research on anisotropy. Casavola et al. [19] had shown up to a 50% and 60% decrease in maximum stress and energy absorption capability, respectively, between a filament orientation aligned vs not aligned with the load direction. Those results reveal that the anisotropy heavily affects the properties of FDM parts, and not understanding those anisotropic behaviors can cause items made with this process to be unusable. Despite its significance, there is still an absence of comprehensive

data about the effect of infill orientation on FDM parts.

D. Research Gap

The goal of this research study was to answer the question: How does changing the infill orientation affect the flexural properties in parts made with FDM? This study measured the flexural properties of FDM parts while only changing the infill orientation to further investigate the effect of orientation. Furthermore, it builds upon the more general knowledge of the anisotropy in FDM parts by providing specific insight into how orientation affects the various flexural properties of those parts.

II. Material and Methods

A. Research Approach and Methods

This study used the mixed methods approach to investigate the effect of infill orientation on the flexural properties in FDM parts. Quantitative data were utilized to determine the mechanical properties of the samples. Qualitative data were utilized to record observations and failure modes. The quantitative and qualitative data collected were able to reliably validate the conclusions made in the study at hand.

The experimental research method was used in this study to determine the effect that the infill orientation of FDM parts had on their flexural properties. This method was used to identify the changes to the flexural properties caused by modifying the infill orientation, which provided enough evidence to thoroughly answer the research question.

B. Subject Selection

This study used the Acrylonitrile Butadiene Styrene (ABS) polymer and the grid infill pattern. ABS was chosen because it is one of the most common materials used in FDM and was chosen for many similar studies including [9], [11], and [19]. The grid infill was chosen due to it being one of the most common infill patterns and having no significant structural abnormalities. Infill orientations of 0° and 45° were chosen due to this infill pattern having four axes of symmetry,

meaning those orientations would result in the largest amount of variation. The width selection required more analysis since the impact of the outer border on the overall characteristics was unknown. The border consisted of solidly printed lines of filament in the vertical and horizontal directions, which could skew the results of the test if the border represented a large part of the overall strength of the samples. 3-Point Bending tests were conducted on samples with a length of 96 mm, a thickness of 3 mm, and widths of 12.5 mm, 25 mm, and 37.5 mm. Even though the border made up a slightly larger portion of the smaller samples, all three widths resulted in almost identical properties. As a result, the width was chosen based on how consistently it was able to be printed, and that was determined using the coefficient of variation, a measurement of the variation and discrepancy in a data set. The 25 mm width resulted in samples that had the least coefficient of variation between the trials and thus was chosen to be used in this study.

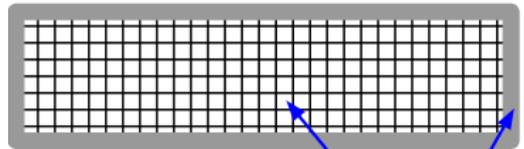
C. Validity

The number of external factors that affected the results were minimized, and five replicates were done for each orientation to ensure that outliers would not heavily impact the data. Due to ABS being a polymer that shares many characteristics with a multitude of other polymers, the overall results of this study could be applied to FDM parts made from those similar polymers as well. This study focused on the general trends between the flexural properties of samples in different orientations and less on the individual strength values themselves. This means that if a part has a similar structure and infill to the samples in this experiment, the results would provide some insight into its properties. The tools used in this study were a dial caliper and a micrometer, both of which determined the dimensions of the samples with an accuracy of 0.001 mm. An Instron Universal Testing machine with a 3-Point Bending fixture was used to gather data about the strength and displacement of each of the samples. The Instron machine was accurate to within 0.5% of the measured force.

D. Procedure

In this study, ten samples made from ABS with dimensions of 96 mm by 25 mm by 3 mm were printed through FDM in a Stratasys© uPrint© SE. The samples had 100% infill density and the grid infill pattern. Five samples had a 0° infill orientation and five had a 45° infill orientation. The border width was set to the smallest setting, 0.1 mm. **Figure 1** provides a diagram illustrating the samples' infill pattern and orientation. The ten samples were labeled and then measured with a dial caliper, micrometer, and scale to determine the actual dimensions and weight of each sample. A 3-Point Bending test, depicted in **Figure 2**, was performed with an Instron Universal Testing machine with a 100 N load cell according to the specifications in the standard provided by ASTM [21]. Following those specifications, the test used a support span of 72 mm, had a 2.54 mm/min loading velocity, and ran until the samples had passed their point of ultimate strength. The tests provided values for time and force at displacement intervals of 0.0254 mm. The failure mode, which is the way that each sample broke, was manually recorded.

0° Infill Orientation:



45° Infill Orientation: Grid Pattern Border

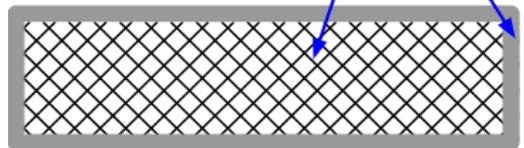


Figure 1. Diagram of the samples detailing the grid infill pattern at 0° and 45° infill orientations. Note that the space in between the lines is only to emphasize the infill structure and no perceptible gaps were in the actual samples tested.

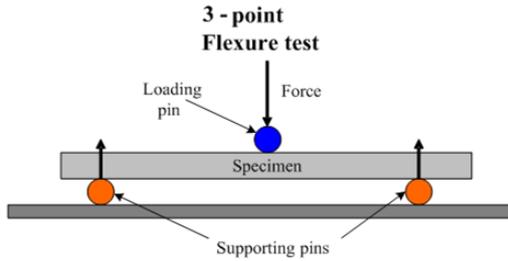


Figure 2. Diagram of 3-Point Bending test used in this study. Adapted from [22]

E. Data Analysis

The standard deviations, averages, and coefficients of variation were used to analyze the results of this study. The equations in the standard test method [21] were used to determine the flexural strength, flexural chord modulus of elasticity, and other related properties from the collected data. Graphs displaying force vs displacement and flexural stress vs flexural strain were created.

III. Results

A. General Characteristics

The basic dimensions and characteristics of the samples are displayed in **Table 1**. The graph of force vs displacement for all samples is displayed in **Figure 3**. The maximum force, displacement at the maximum force, and displacement at the point of failure for each sample are displayed in **Table 2**.

Sample No	Infill Orientation (degrees)	Length (cm)	Width (cm)	Thickness (cm)	Weight (g)	Density (g/cm ³)
1	0	9.589	2.494	0.320	6.441	0.841
2	0	9.602	2.506	0.318	6.441	0.842
3	0	9.589	2.497	0.321	6.441	0.839
4	0	9.611	2.499	0.315	6.441	0.851
5	0	9.608	2.497	0.321	6.441	0.836
6	45	9.601	2.501	0.325	6.532	0.836
7	45	9.595	2.492	0.317	6.532	0.862
8	45	9.605	2.498	0.319	6.486	0.848
9	45	9.590	2.496	0.321	6.532	0.850
10	45	9.580	2.499	0.332	6.532	0.822

Table 1. The infill orientations, dimensions, weight, and density of tested samples.

EFFECT OF INFILL ORIENTATION ON THE FLEXURAL PROPERTIES IN FDM PARTS

Sample No	Infill Orientation (degrees)	Max Force (N)	Displacement at Max Force (mm)	Displacement at Failure Point (mm)
1	0	89	9.1897	11.4732
2	0	92	8.8392	9.9263
3	0	90	9.6393	11.1430
4	0	94	8.5471	9.4615
5	0	90	9.3472	11.2979
6	45	91	11.8237	> 16
7	45	98	11.5138	> 16
8	45	98	11.5291	> 16
9	45	98	11.3614	> 16
10	45	91	12.1615	> 16

Table 2. The maximum force exerted, the displacement at the point of maximum force, and the displacement at the point of failure for each sample during the 3-Point Bending tests.

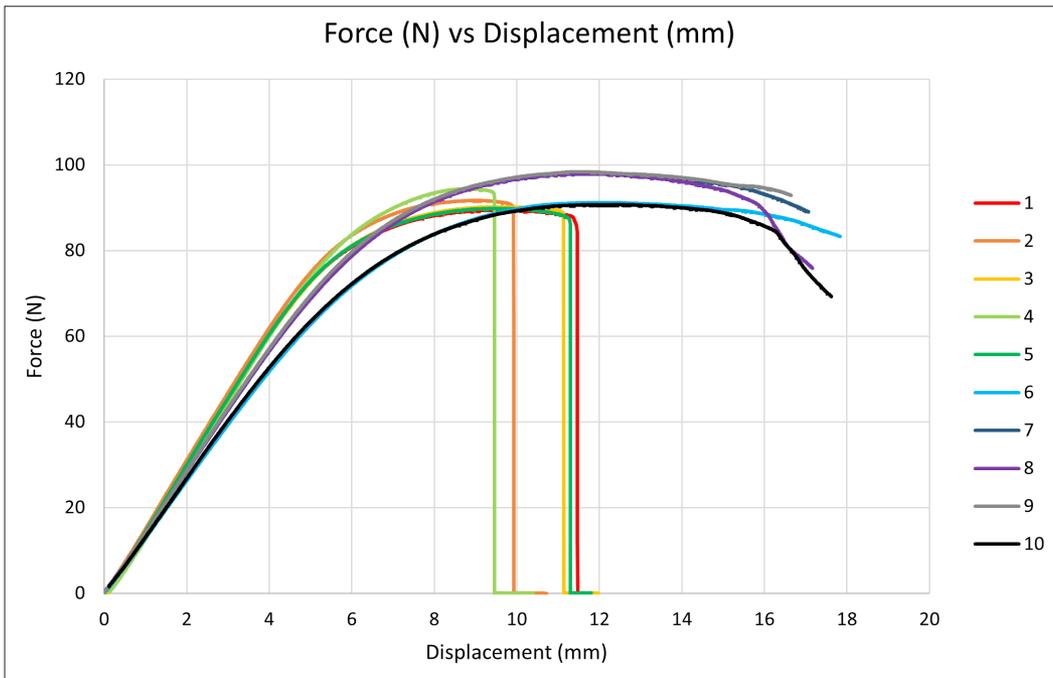


Figure 3. Force vs displacement in a 3-Point Bending test, describing how much force was required to continuously push down the sample with the loading pin at a rate of 0.254 mm/min. Points were recorded in intervals of 0.05 mm or less of displacement.

Samples 1 to 5, with a 0° infill orientation, exhibited a brittle failure mode, breaking cleanly into two pieces while or slightly after reaching its maximum loading force. In **Figure 3**, between a displacement of 9 mm and 12 mm, the value of force for samples 1 to 5 sharply dropped to 0 N. Alternatively, samples 6 to 10, with a 45° infill orientation, exhibited a ductile failure mode, deforming further instead of breaking after reaching its peak load. The values of force for samples 6 to 10 gradually decrease after the samples have reached their maximum load. **Figure 4** and **Figure 5** show pictures of samples 1 and 6, respectively, and visually display the difference in failure modes between the two samples.

All samples showed visual signs of plastic deformation visible as whitening at the central break location in the 0° samples and the middle of the 45° samples. However, plastic deformation was much more extensive in the 45° samples, as shown in **Figure 5**. The plastic deformation could be observed mainly on the tension side, where the material was being stretched apart. The plastic deformation across the center of the span affected both the material composing the border and the infill of the sample. The most deformation could be seen at the center of the samples.

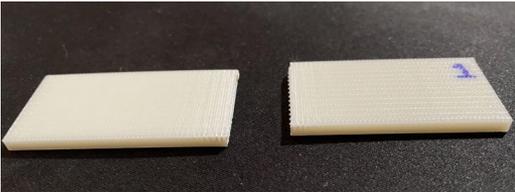


Figure 4. Sample 1 (0° orientation) after it underwent its 3-Point Bending test and a brittle failure mode.



Figure 5. Sample 6 (45° orientation) after it underwent its 3-Point Bending test and a ductile failure mode.

B. Calculated Mechanical Properties

All the properties calculated for the samples were found using the equations found in the standard method [21]. The flexural stress was calculated using (1) and the flexural strain was calculated using (2).

$$\sigma = \frac{5PL}{2bh^2} \quad (1)$$

where σ is the stress at the outer surface at mid-span, P is the applied force, L is the support span, b is the width of the beam, and h is the thickness of the beam.

$$\varepsilon = \frac{6\delta h}{L^2} \quad (2)$$

where ε is the maximum strain at the outer surface, δ is the mid-span deflection, L is the support span, and h is the thickness of the beam.

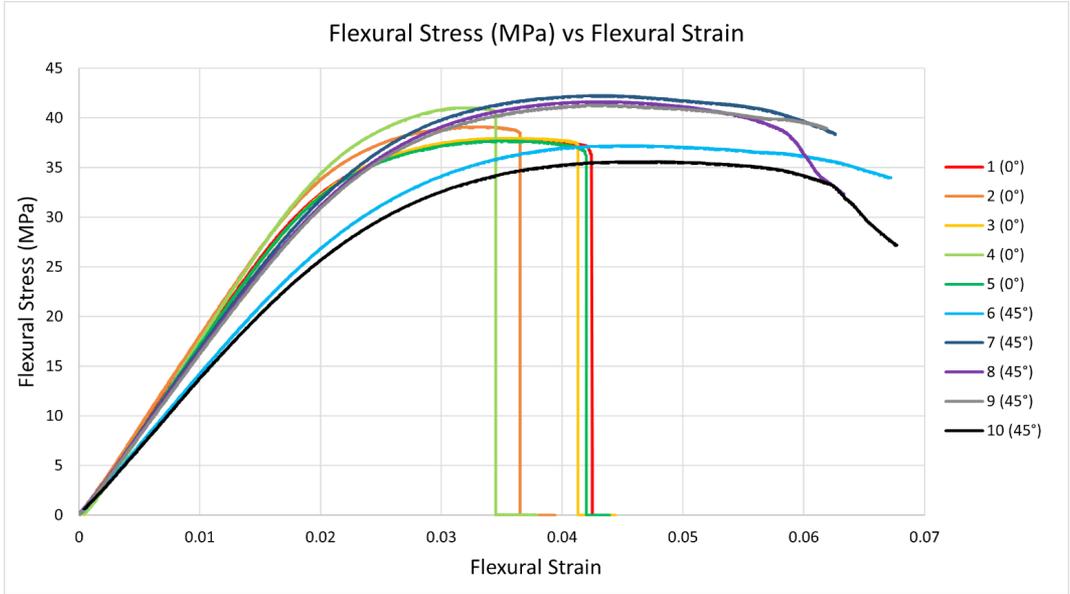


Figure 6. Graph of flexural stress vs flexural strain during the 3-Point Bending test.

Figure 6 shows the graph of flexural stress vs flexural strain. The 0° samples all followed a linear pattern, transitioned to a non-linear pattern until peaking around 39 MPa, and then suddenly decreased to 0 MPa. The 45° samples originally followed the same pattern as the 0° samples, starting linear and becoming non-linear, but did not experience any sudden drops, and instead more gradually decreased after peaking around 40 MPa.

The flexural strength, the flexural strain at the point of maximum flexural stress, the 0.2% offset yield stress, and the flexural chord modulus of elasticity for the samples are displayed in **Table 3**. The flexural chord modulus of elasticity was calculated according to (3).

$$E_f^{chord} = \frac{\Delta\sigma}{\Delta\varepsilon} \quad (3)$$

where E_f^{chord} is the flexural chord modulus of elasticity, $\Delta\sigma$ is the difference in flexural stress between the two selected strain points, and $\Delta\varepsilon$ is the difference between the two selected strain points.

The 0.2% offset line was calculated by finding the linear slope of the original data between the strain values of 0.1% and 0.3% and offsetting a line with that slope to the right by 0.2% flexural strain. The 0.2% offset yield stress was then found by determining the flexural stress value for the intersection point between the flexural stress vs flexural strain curve and the linear offset line. That 0.2% yield stress represents the amount of force required to cause permanent deformation of 0.2% in the sample. The mean, standard deviation, and coefficient of variation for each of the four properties displayed in **Table 3** are shown in **Table 4**.

EFFECT OF INFILL ORIENTATION ON THE FLEXURAL PROPERTIES IN FDM PARTS

Sample No	Infill Orientation (degrees)	Flexural Strength (MPa)	Flexural Strain at Max Flexural Stress	0.2% Offset Yield Stress (MPa)	Flexural Chord Modulus of Elasticity (MPa)
1	0	37.78	0.034	33.89	1,683
2	0	39.11	0.033	36.08	1,746
3	0	37.98	0.036	33.99	1,685
4	0	41.02	0.031	38.70	1,693
5	0	37.69	0.035	34.02	1,671
6	45	37.21	0.045	29.32	1,416
7	45	42.26	0.042	36.26	1,612
8	45	41.64	0.043	35.57	1,590
9	45	41.28	0.042	35.17	1,586
10	45	35.59	0.047	28.99	1,327

Table 3. Flexural strength, flexural strain at the point of maximum flexural stress, the 0.2% offset yield stress, and the flexural chord modulus of elasticity are displayed for samples 1 to 5 during the 3-Point Bending tests.

Infill Orientation	Flexural Strength (MPa)		Flexural Strain at Maximum Flexural Stress		0.2% Offset Yield Stress (MPa)		Flexural Chord Modulus of Elasticity (MPa)	
	Mean ± SD	CV	Mean ± SD	CV	Mean ± SD	CV	Mean ± SD	CV
0 degrees	38.72 ± 1.26	3.25%	0.0336 ± .0016	4.76%	35.34 ± 1.87	3.25%	1,695.56 ± 26.09	1.54
45 degrees	39.60 ± 2.68	6.76%	0.0436 ± .0018	4.04%	33.06 ± 3.21	6.76%	1,506.09 ± 114.15	7.58

Table 4. The means, standard deviations, and coefficients of variation (CV) of the flexural strength, flexural strain at the point of maximum flexural stress, the 0.2% offset yield stress, and the flexural chord modulus of elasticity are displayed for each of the infill orientations evaluated.

C. Summary of Research

Overall, samples 6 to 10 had higher maximum loads and higher flexural properties than samples 1 to 5, but lower flexural chord moduli of elasticity. The failure modes were dependent on the orientation of the infill pattern, as the samples printed with the grid pattern in the 0° orientation had a brittle failure mode and the samples with the grid pattern in the 45° orientation had a ductile failure mode. Across most of the samples, the curve of any calculated property followed the same approximate path up until near the point of failure.

IV. Discussion

A. Property Analysis

The primary difference in the samples were their failure modes, and the results strongly suggest that the infill orientation was the cause of it. There is imperfect bonding between the parallel layers in FDM parts that creates a weakness, and the 0° samples had half of their internal filament in layers parallel to the y/z plane provided in **Figure 7**, which depicts the forces active in a 3-Point Bending test. Due to their orientation, the forces were in the same direction as the vulnerability, resulting in the complete failure of the interlayer bonds and a brittle failure mode. On the other hand, in the 45° samples, both layers of filament had their weaknesses offset by 45° to the direction of force and logically the interlayer bonds were more resistant to completely failing. Harpool et al. [13], instead of using flexural tests, conducted tensile tests on FDM samples. They used samples with a 15% infill density and observed that both 0° and 45° samples resulted in ductile failure modes under tensile force. The lack of change in failure modes may have been due to the low infill density resulting in samples that were less stiff and acted similarly to the less stiff 45° samples in the study at hand. However, they did evaluate solid samples with 100% infill density and found that those samples were very stiff and exhibited a brittle failure mode. Based on their results [13], the conclusions made in the study at hand apply primarily to FDM parts that have higher infill densities and that the infill density's influence on the part cannot be ignored.

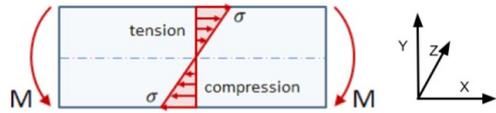


Figure 7. Diagram of the forces active in a 3-Point Bending test. Adapted from [23]

Since both 0° and 45° samples in the study at hand had no measured differences in their weights and the 45° samples were able to support a maximum load that was 5% higher on average than the 0° samples, it suggests that using a 45° infill orientation results in slightly stronger FDM objects than a 0° infill orientation for the same amount of material used. One explanation for this is that due to the chosen infill pattern consisting of two primary filament directions perpendicular to each other, the 45° samples were able to utilize to some extent filament in both directions to support the load, whereas the 0° samples were unable to use the filament in the direction perpendicular to its length to support almost any part of the load.

Alternatively, despite the 45° samples being able to support heavier loads and having higher flexural strengths, the 0° samples had a 7% higher 0.2% offset yield stress on average compared to the 45° samples. Due to the definition of 0.2% offset yield stress, this means that 0° samples were able to support a 7% higher amount of force before permanently deforming by 0.2%, which indicates that having a 0° infill orientation results in parts that are initially stiffer than parts with a 45° infill orientation. Additionally, the 0° samples had a flexural chord modulus of elasticity that was 13% higher than the 45° samples, which means that 13% heavier loads are required to deform or bend the 0° samples by the same amount as the 45° samples during the linear part of the curve. That suggests that a 0° infill orientation is better in situations where avoiding or minimizing permanent deformation is the priority, but in situations where deformation is acceptable and the ability to withstand a heavier load is prioritized, a 45° infill orientation is better.

The engineering stress vs engineering strain graph in Harpool et al. [13] showed that the samples with a 0° infill orientation were able to support much higher

engineering stress values during a tensile test compared to the samples with a 45° infill orientation. This is contrary to the results of the study at hand where both infill orientations resulted in similar flexural stress values. ABS in general tends to have higher flexural strength than tensile strength, which might have caused it to be more sensitive to the tensile forces in [13].

Another study, [19], concluded that when the primary load direction is parallel to the primary filament direction, the mechanical properties are higher. In the study at hand, the 0° samples had 7% higher 0.2% offset yield stress and 13% higher flexural chord moduli of elasticity, which supports the conclusion made in that study. While that conclusion is not always true for the samples evaluated in this study, in the few properties where the 0° samples are weaker, the values for the 45° samples are at most 5% greater than the values for the 0° samples. This is not a negligible difference, but since two of the main flexural properties are much greater in the 0° samples than in the 45° samples, it can be concluded that generally, the conclusion in the earlier study [19] is supported by the results of the study at hand.

Since the 0° samples had coefficients of variation that were much lower than the 45° samples in all calculated flexural properties except flexural strain, where they were only 1% higher, it suggests that fabricating parts with a 3D printer is a much more consistent process when using a 0° infill orientation rather than a 45° infill orientation. One explanation for this is that 3D printers may be more consistent at printing in an isolated x or y direction and less consistent at printing in a diagonal direction. This means that when parts need to be made to specific tolerances and specifications, using a 0° infill orientation is better, but when precision can be sacrificed for additional strength, then a 45° infill orientation is better.

Overall, the data collected in this study suggests that the 0° infill orientation is superior in situations where a stiffer, more precise part resistant to deformation is required, but a 45° infill orientation is superior in situations where the ability to support a larger amount of weight is the highest priority for the part. Additionally, it indicates that using a 0° infill orientation creates risks for a part to snap abruptly and brittlely when it is overloaded but using a 45° infill orientation removes the risk of snapping and results in a

ductile failure mode where the part would get weaker as it bends further.

B. Alternative Explanations and Limitations

An alternative explanation for these results is that the differences in the way that the 3D printer prints the 0° infill orientation versus a 45° infill orientation are the cause of the differences in the flexural properties. To elaborate, the printer needs to move horizontally and vertically simultaneously to print a diagonal section as opposed to only needing to move in one direction to print a vertical or horizontal section. That might have caused the 45° parts to have more inaccuracies and weaknesses than the 0° parts, as they are printed primarily with filament running along a diagonal path that is more complicated for a printer to follow. That could mean that the results of this study were only due to the flaws in the printer itself and the effect of infill orientation had little to no impact on a part's flexural properties.

Further research can be utilized to come to a better conclusion about the validity of the alternative explanation. Specifically, another study could investigate whether printing the 45° samples with a 45° angle offset, so that the infill would still be in line with the printer's x and y axes, results in samples that differ greatly from the 45° samples in the study at hand. That would be necessary to determine if the weaknesses in the 45° samples in this study are due to weaknesses in the printer, or if the infill orientation is the primary cause of those weaknesses.

While the conclusions made in this study can be applied to other infill patterns that are structured similarly to the grid infill pattern, little can be concluded about other infill patterns that do not share a similar structure. Additionally, while the 0° and 45° orientations provided the largest variation, it is still necessary to test the angles between those orientations for a more thorough understanding of the anisotropic behaviors. Also, the infill density has been shown to have a large enough influence on the properties of the samples to cause the failure mode to switch from a brittle to a ductile failure mode. Additional studies can investigate how the conclusions made in this study are affected by changes in infill density.

V. Conclusion

The FDM parts with a 0° infill orientation were most resistant to bending and had less variation between replicates of the samples, and the FDM parts with a 45° infill orientation were able to support a higher maximum force. When a bending force was applied, the 0° infill orientation resulted in a brittle failure mode, while the 45° infill orientation resulted in a ductile failure mode. These conclusions are important to the field of study because they provide insight into the extent of influence that the infill orientation has on various flexural properties and thus can be used to make decisions when manufacturing parts with FDM. Furthermore, if these conclusions are not considered when manufacturing an object using FDM, then the object may fail to meet the required standards for its properties, which is extremely dangerous in many industries.

VI. Acknowledgements

Many thanks to Dr. John Gillespie for his knowledgeable guidance and advise throughout the research process. I would like to thank Johnny Thiravong for his instruction regarding the testing of the samples in this study. I would also like to recognize Christopher Meanor for providing access to a 3D printer and his role in building the samples.

References

- [1] K. V. Wong and A. Hernandez, "A review of additive manufacturing," *ISRN Mechanical Engineering*, 16-Aug-2012. [Online]. Available: <https://doi.org/10.5402/2012/208760>. [Accessed: 01-Nov-2021].
- [2] M. Attaran, "The rise of 3-D printing: The advantages of additive manufacturing over traditional manufacturing," *Business Horizons*, 28-Jun-2017. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0007681317300897>. [Accessed: 28-Sep-2021].
- [3] M. Habrman, "Influence of the infill on the tensile strength and the economic factors of 3D printing," [Online]. Available: https://dspace5.zcu.cz/bitstream/11025/36423/1/DAAAM_2019_Habrman.pdf. [Accessed: 01-Oct-2021].
- [4] K. Wang, X. Xie, J. Wang, A. Zhao, Y. Peng, and Y. Rao, "Effects of infill characteristics and strain rate on the deformation and failure properties of additively manufactured polyamide-based composite structures," *Results in Physics*, 20-Aug-2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2211379720318131>. [Accessed: 28-Sep-2021].
- [5] J. C. Camargo, A. R. Machado, E. C. Almeida, and E. F. M. S. Silva, "Mechanical properties of pla-graphene filament for FDM 3D printing," *The International Journal of Advanced Manufacturing Technology*, 22-Apr-2019. [Online]. Available: <https://link.springer.com/article/10.1007/s00170-019-03532-5>. [Accessed: 01-Oct-2021].
- [6] B. Aloyaydi, "Investigation of infill-patterns on mechanical response of 3D printed poly-lactic-acid," *Polymer Testing*. [Online]. Available: <https://doi.org/10.1016/j.polymertesting.2020.106557>. [Accessed: 01-Oct-2021].
- [7] G. Ehrmann and A. Ehrmann, "Investigation of the shape-memory properties of 3D printed PLA structures with different infills," *MDPI*, 05-Jan-2021. [Online]. Available: <https://doi.org/10.3390/polym13010164>. [Accessed: 01-Oct-2021].
- [8] A. Elkaseer, S. Schneider, and S. G. Scholz, "Experiment-based process modeling and optimization for high-quality and resource-efficient FFF 3D printing," *MDPI*, 22-Apr-2020. [Online]. Available: <https://www.mdpi.com/2076-3417/10/8/2899/htm>. [Accessed: 27-Sep-2021].
- [9] W. M. H. Verbeeten, R. J. Arnold-Bik, and M. Lorenzo-Bañuelos, "Print velocity effects on strain-rate sensitivity of acrylonitrile-butadiene-styrene using material extrusion additive manufacturing," *MDPI*, 01-Jan-2021. [Online]. Available: <https://www.mdpi.com/2073-4360/13/1/149/htm>. [Accessed: 14-Sep-2021].

EFFECT OF INFILL ORIENTATION ON THE FLEXURAL PROPERTIES IN FDM PARTS

- [10] A. El Moumen, M. Tarfaoui, and K. Lafdi, "Modelling of the temperature and residual stress fields during 3d printing of polymer composites," *The International Journal of Advanced Manufacturing Technology*, 12-Jun-2019. [Online]. Available: <https://link.springer.com/article/10.1007%2Fs00170-019-03965-y>. [Accessed: 14-Sep-2021].
- [11] A. K. Sood, R. K. Odar, and S. S. Mahapatra, "Experimental investigation and empirical modelling of FDM process for compressive strength improvement," *Journal of Advanced Research*, 02-Jun-2011. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S209012321100066X>. [Accessed: 28-Sep-2021].
- [12] H. Gonabadi, A. Yadav, and S. J. Bull, "The effect of processing parameters on the mechanical characteristics of PLA produced by a 3D FFF Printer," *The International Journal of Advanced Manufacturing Technology*, 02-Oct-2020. [Online]. Available: <https://link.springer.com/article/10.1007/s00170-020-06138-4>. [Accessed: 28-Sep-2021].
- [13] T. D. Harpool, I. M. Alarifi, B. A. Alshammari, A. Aabid, M. Baig, R. A. Malik, A. Mohamed Sayed, R. Asmatulu, and T. M. A. A. EL-Bagory, "Evaluation of the infill design on the tensile response of 3D printed polylactic acid polymer," *MDPI*, 25-Apr-2021. [Online]. Available: <https://doi.org/10.3390/ma14092195>. [Accessed: 25-Sep-2021].
- [14] O. Basurto-Vázquez, E. P. Sánchez-Rodríguez, G. J. McShane, and D. I. Medina, "Load distribution on PET-g 3D prints of honeycomb cellular structures under compression load," *MDPI*, 17-Jun-2021. [Online]. Available: <https://www.mdpi.com/2073-4360/13/12/1983>. [Accessed: 01-Oct-2021].
- [15] J. Sárosi, "Investigation of 3D printing parameters affecting the impact strength," *ResearchGate*, Jul-2021. [Online]. Available: https://www.researchgate.net/publication/353286074_investigation_of_3D_printing_parameters_affecting_the_impact_strength. [Accessed: 14-Sep-2021].
- [16] L. Baich, "Impact of infill design on mechanical strength and production cost in material extrusion Based Additive Manufacturing," 01-Jan-1970. [Online]. Available: <https://www.semanticscholar.org/paper/Impact-of-Infill-Design-on-Mechanical-Strength-and-Baich/a53d8f604c0a99e907dccc55503c7c91ec72cf1>. [Accessed: 01-Oct-2021].
- [17] M. Lalegani Dezaki and M. K. A. Mohd Ariffin, "The effects of combined infill patterns on mechanical properties in FDM process," *MDPI*, 26-Nov-2020. [Online]. Available: <https://doi.org/10.3390/polym12122792>. [Accessed: 01-Oct-2021].
- [18] C. Lubombo and M. A. Huneault, "Effect of infill patterns on the mechanical performance of lightweight 3D-printed cellular PLA parts," *Materials Today Communications*, 21-Sep-2018. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S2352492818301600>. [Accessed: 07-Nov-2021].
- [19] C. Casavola, A. Cazzato, V. Moramarco, and G. Renna, "Mechanical behaviour of ABS-fused filament fabrication compounds under impact tensile loadings," *MDPI*, 19-Apr-2019. [Online]. Available: <https://www.mdpi.com/1996-1944/12/8/1295>. [Accessed: 14-Sep-2021].
- [20] J. Podroužek, M. Marcon, K. Ninčević, and R. Wan-Wendner, "Bio-inspired 3D infill patterns for additive manufacturing and structural applications," *MDPI*, 06-Feb-2019. [Online]. Available: <https://doi.org/10.3390/ma12030499>. [Accessed: 01-Oct-2021].
- [21] Standard Test Method for Flexural Properties of Polymer Matrix Composite Materials, D7264/D7264M – 21, 2021.
- [22] "Materials engineering," *Flexural strength tests of ceramics [SubsTech]*, 01-Jun-2012. [Online]. Available: https://www.substech.com/dokuwiki/doku.php?id=flexural_strength_tests_of_ceramics. [Accessed: 29-Apr-2022].
- [23] "Strength of materials," *MechaniCalc*. [Online]. Available: <https://mechanicalcalc.com/reference/strength-of-materials>. [Accessed: 26-Jun-2022].

Analyzing Twitter Data to Understand COVID-19 Vaccine Hesitancy Following the Federal Private Sector Mandate

Manisha Palaniappan

Since the rollout of the COVID-19 vaccine, vaccine hesitancy has been prevalent in the US, particularly on social media such as Twitter. Significant efforts to understand the reasons behind anti-vaccine sentiments have been concentrated in the earlier stages of vaccine distribution and research suggests personal freedom to be a primary driver of vaccine hesitancy. This study aims to analyze anti-vaccine sentiment on Twitter following the first federal private sector mandate in November of 2021 to better understand how mandates impact vaccine sentiment in the later stages of the pandemic. Two machine learning models, roBERTa and LDA, were used to analyze a pre-curated Kaggle dataset of COVID-19 vaccine-related tweets. Results suggest that there is an overwhelmingly negative sentiment towards the vaccine. Reasons include efficacy and side-effect concerns, mistrust in institutions, and anti-mandate sentiments. This study identified common reasons behind vaccine hesitancy which future research can hopefully use to create targeted outreach programs.

Key Words: COVID-19, vaccine sentiment, vaccine hesitancy, machine learning

Introduction

COVID-19 has had devastating effects worldwide. According to the World Health Organization, at the time of writing, there have been close to 300 million confirmed cases, including more than 5 million deaths globally (World Health Organization, 2021). In response, there were extensive efforts to develop a vaccine with research beginning in January of 2020. Less than 11 months later, vaccine distribution started, making it the fastest developed vaccine in history (Kuter et al, 2021). Despite such rapid development, vaccination rates have not increased sufficiently mainly due to vaccine hesitancy. According to renowned Canadian physician and professor Noni McDonald, vaccine hesitancy is defined as:

“A delay in acceptance or refusal of vaccination despite availability of vaccination services. Vaccine hesitancy is complex and context-specific, varying across time, place, and vaccines. It is influenced by factors such as complacency, convenience, and confidence” (MacDonald, 2015).

Vaccine hesitancy is not a new phenomenon; in 2017, a measles outbreak occurred in Minnesota due to low vaccination rates (Petraco, 2019). However, vaccine hesitancy has been particularly widespread in the case of COVID-19 vaccines because of the political polarization that has been associated with it (Jiang et al, 2020). To combat low vaccination rates, the federal government implemented vaccine mandates. On November 4, 2021, the white house announced the first federal level vaccination mandate for the private sector (The United States Government, 2021). This

was met with considerable controversy, and on January 13, 2022 the Supreme Court blocked this mandate (Breuninger & Kimball, 2022). Even so, individual states as well as private businesses are still enforcing vaccinations and therefore, it is important to understand how these mandates play into vaccine hesitancy. Proponents of vaccine mandates argue that these mandates will normalize the COVID-19 vaccine and encourage people to get vaccinated

(Ashwell et al, 2021). However, there has still been much opposition from those who are vaccine hesitant. In order to investigate this phenomenon, social media has been used extensively to understand public perception, specifically in the case of the COVID-19 vaccine.

Social Media Analysis

Social media has proven to be a very useful tool in understanding public sentiment towards COVID-19 vaccines because so many Americans use it frequently. According to the Pew Research Center, over 70% of American adults use at least one social media site

(Pew Research Center, 2021). Specifically, Twitter has been used to understand public sentiment in regards to public health and infectious diseases due to low costs associated with this type of research as well as its effectiveness (Sinnenberg et al, 2017). With respect to the pandemic, it has fueled vaccine talk and opinions, which makes it a prominent tool in helping to analyze COVID-19 vaccine sentiment within the US (Featherstone et al, 2020a,b). Particularly, it has aided in understanding vaccine hesitancy. However, because of the extensive amounts of data provided on this platform, it becomes impossible to conduct analyses by hand. Instead, researchers frequently use machine learning models to understand vaccine sentiment and themes behind vaccine hesitancy.

Literature Review

Overview of Literature

Because of the controversial yet essential nature of the COVID-19 vaccinations, it is important to conduct research to understand vaccine sentiments and underlying reasons for vaccine hesitancy. By understanding the causes of vaccine hesitancy, a first step can be taken to finding effective ways to communicate the necessity of getting vaccinated to the unvaccinated population. Therefore, in order to comprehend this topic, researchers have been trying to study vaccine hesitancy since the vaccine was first manufactured (Boucher et al, 2021). However, results have been mixed. Some past research has suggested that negative vaccine sentiment has generally declined as time progressed (Hu et al, 2021). However, others have revealed that there is a fluctuation in negative sentiment, with it declining as positive vaccine efficacy news came out, but then declining afterwards despite no major correlation factors (Fazel et al, 2021). As such, it has been difficult to determine the level of negative public opinion towards COVID-19 vaccines. Furthermore, there has also been research conducted, especially through social media, to discern specific reasons for vaccine hesitancy with respect to COVID-19 vaccines. A review of literature regarding COVID-19 vaccine sentiment conducted at the Sultan Idris Education University revealed three major themes for vaccine hesitancy: concerns regarding the vaccine itself, mistrust in institutions, and individuals' social attributes

(Almoodi et al, 2021). These will be further discussed in the following sections.

Vaccine-related

The first major theme—also the largest one—is specifically vaccine-related. This includes efficacy of COVID-19 vaccines, people's confidence in them, and vaccine safety (Almoodi et al, 2021). One study led by Jean-Christophe Boucher, a reputable researcher at the University of Calgary used sentiment and semantic analysis of Twitter data using machine learning at the time when successful vaccine trials were first announced to understand vaccine hesitancy. They

found that there were many concerns online regarding harmful side effects of vaccines and also a lack of confidence that the vaccine would be effective in ridding the population of COVID-19 (Boucher et al, 2021). These reasons largely stem from the fact that the vaccine was created so quickly, and as such, many people who are vaccine hesitant are afraid that there may not have been enough time to properly create an effective vaccine (Paul et al, 2021). Advocates of the vaccine argue that this is untrue since the fast development was due to previous research accumulated over years, in addition to cooperation of world leaders and plenty of funding (Kuter et al, 2021). Overall, these studies were mostly conducted when the vaccine was first developed, and as such, there has been high uncertainty regarding the long-term impacts of the vaccine, though it is likely to change over time. Therefore, more research needs to be done on vaccine sentiment at the later stages of the pandemic when the vaccine has been more widely distributed.

Mistrust in Institutions

Another major theme present in previous literature regarding reasons for vaccine hesitancy is mistrust in institutions which includes mistrust of the media, international organizations (e.g., pharmaceutical companies), as well as the government—local and federal. In Boucher's social media analysis, he found that there was a large proportion of tweets which expressed criticism of the government and multinational corporations' actions to limit individuals' freedom to get vaccinated (Boucher et al, 2021). This, in turn, led many to be distrustful of large businesses such as those in the pharmaceutical industry and the government. Specifically, one longitudinal study conducted by Ariel Fridman looked at the influence of political parties on trust in the media and government. Results revealed that Republicans were more likely to have mistrust in the media and government, which was closely correlated with their lack of confidence in the COVID-19 vaccine, while the opposite was true for Democrats (Fridman et al, 2021). Thus, due to widespread mistrust, which is influenced by the politicization of the vaccine, forceful actions taken by the government to promote the vaccine, may have major negative consequences in regards to public perception of the vaccine.

Individuals' social attributes

Individuals' social attributes also play a big role in vaccine hesitancy. This ranges from one's socioeconomic status, level of education, to political affiliation. A study conducted by Chen Luo analyzed Chinese and American sentiment toward the COVID-19 vaccine through a cultural perspective. In general, US citizens' individualism, as opposed to the collectivism present in China, are one of the main drivers of vaccine hesitancy (Luo et al, 2021). While level of education and finances do not have a strong correlation with vaccine hesitancy, social cohesion and politicization, which are further influenced by the fact that Americans have more individualistic views, play an important role in vaccine hesitancy (Paul et al, 2021). More specifically, it has been revealed that liberals express the least amount of distrust and vaccine hesitancy, while the opposite is true of conservatives (Jiang et al, 2021). Because of the heavy influence of political polarization fueled by the Americans' individualistic mindsets, critical announcements made by political leaders are one of the major contributors to swaying public opinion on vaccines one way or another (Hu et al, 2021). Therefore, the vaccine mandate that President Biden implemented in November of 2021, could have had major impacts on public sentiment of the vaccine.

Although there has been research conducted on vaccine hesitancy in the past, it still needs to be investigated more thoroughly in order to get a more complete understanding of causes behind vaccine hesitancy. According to the Centers for Disease Control and Prevention (CDC) at the time of writing, close to 40% of the US population have not been fully vaccinated while there is still an average of more than 500,000 COVID-19 cases each week (Centers for Disease Control and Prevention, 2022). As such, it can be seen that vaccine hesitancy is still prevalent and needs to be more thoroughly investigated in order to implement efficient interventions to increase vaccination. Furthermore, because of the politicized nature of the COVID-19 vaccines, vaccine mandates, implemented to increase vaccination rates, may not be very effective. Specifically, there has been speculation that vaccine mandates could potentially be related to negative sentiments toward the vaccines themselves (Paul et al, 2021). Despite this, there has been little research exploring vaccine sentiment with respect to the fed-

ANALYZING TWITTER DATA TO UNDERSTAND COVID-19 VACCINE HESITANCY

eral vaccine mandates passed in the US. To address this gap, this study aims to look at American vaccine sentiment on Twitter, and particularly understand the reasons behind vaccine hesitancy at the time of the first private sector federal vaccine mandate in the US. Therefore, the question emerges: What are the reasons behind negative sentiment towards the COVID-19 vaccine as expressed by Americans on Twitter at the time when the first federal vaccine mandate for the private sector was implemented?

Methods

To address the question, similar to previous social media analysis research, this study also makes use of machine learning models in order to understand vaccine sentiment and hesitancy. Particularly, this research uses two pre-trained models which perform sentiment analysis and topic modeling. Sentiment analysis is the process of identifying the overall sentiment, while topic modeling analyzes major themes of a corpus. In accordance with the aim of this study, I used a sentiment analysis model to determine the overall sentiment of each tweet in the dataset and then proceeded to conduct topic modeling on only the negative tweets to understand reasons behind vaccine hesitancy.

Data Extraction and Cleaning

The Twitter data that was collected for this particular study came from between the dates November 04, 2021—when the federal private sector mandate was first passed—to December 04, 2021—a month after the mandate was implemented. Although researchers frequently use the Twitter API to extract their data, many restrictions such as number of tweets and lack of access to historical data, made this a poor choice for the requirements of this particular study. Instead, the data used for this particular study comes from Kaggle, a reputable website which consists of numerous pre-curated datasets that are easily accessible. This specific Kaggle dataset contains just over 730,000 English tweets with the keyword *covid_vaccine* curated from October 27, 2021 to December 31, 2021 (Ozturk, 2021). This study used the information columns of Tweet Content, User Location, and Date that were provided with the dataset. I first filtered out

the rows which contained tweets from November 04, 2021 to December 04, 2021 and came from the United States since I am only focusing on American Twitter sentiment. I filtered out the tweets using a combination of the Microsoft Excel filtering function and Python code. Next, I had to clean my data by performing a series of preprocessing techniques—removing all at-mentions, hashtags, links, and special characters from my data—in order to have it in a format that my data analysis models can understand. To perform this cleaning process, I used a few specific python libraries, which include Pandas and Regex, in order to discard the irrelevant information.

Procedure/Analysis

After filtering and cleaning the data, the final dataset consisted of a little over 57,000 tweets. Due to the large amount of data, it was impossible to analyze it by hand. As such, machine learning and artificial intelligence (AI) was needed to analyze the data. Specifically, this research made use of natural language processing (NLP), a subfield of AI that deals with the interaction between computers and human language. Two different tasks within NLP were performed: sentiment analysis and topic modeling. Sentiment analysis is looking at the overall sentiment of each tweet, specifically understanding whether the tweet is positive, negative, or neutral. Afterwards, particularly focusing on the negative sentiment tweets, topic modeling was conducted to understand major topics being discussed amongst those who are vaccine hesitant. Topic modeling is a model that is used to discern prevalent themes or topics being discussed within large bodies of text.

The method used in this study was modeled after Chad A. Melton and others' research who analyze vaccine hesitancy on Reddit by sentiment analysis and topic modeling

(Melton et al). However, in contrast to their use of the Textblob sentiment analysis model, this study uses the roBERTa pre-trained sentiment analysis model which was trained specifically on Twitter data and thus, has a high accuracy rate. Additionally, Latent Dirichlet Allocation (LDA) Topic Modeling through the Gensim library on Python was used as previous research has identified it to have the most relevant results for short-text data (Albalwi et al).

Results

Sentiment Analysis

After conducting Sentiment Analysis using the roBERTa model, there were three output values for each tweet. The output consisted of the percent positive, negative, and neutral sentiment that each tweet has on a scale from 0 to 1 (Figure 1).

- 1) neutral 0.5942
- 2) negative 0.3839
- 3) positive 0.0219

Figure 1: This is a sample output for a tweet. Each value is on a scale from 0 to 1. Neutral+Negative+Positive=1

Sentiment Analysis revealed that the majority of tweets in my dataset, 58.3% (n=33,679), were negative. The second largest group was neutral at 33.2% (n=19,166), and the smallest was positive at only 8.5% (n=4,937) (Figure 2).

The mean and variance of scores for positive, negative, and neutral tweets were also calculated (Table 1). It can be seen that negative and positive tweets had similar score means which indicate that they had relatively the same strength of sentiment while it was a little lower for neutral. It also seems that positive and negative tweets have similar variances while, again, the neutral tweet scores have a lower variance.

Table 1

Sentiment Percentages

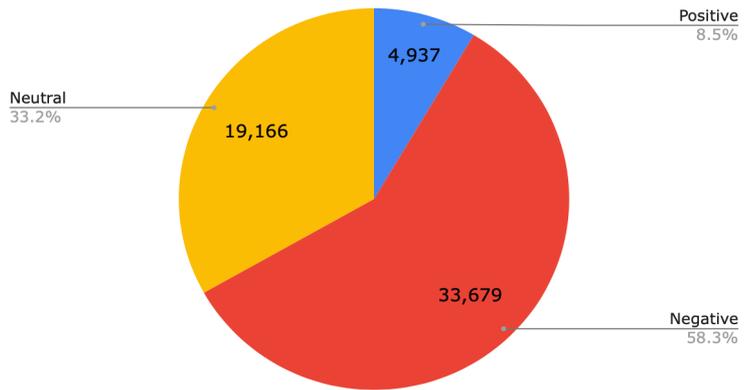


Figure 2. Using the roBERTa model, sentiment analysis results showed that a majority of tweets were negative, with neutral being the second largest, and positive being the smallest group.

	Negative	Positive	Neutral
Mean	0.74	0.72	0.63
Variance	0.020	0.026	0.016

Table 1. Means for the negative and positive tweets were relatively similar with the score for neutral being a little lower. Variance was overall very similar for all three groups, but with neutral again being a little lower than the other two.

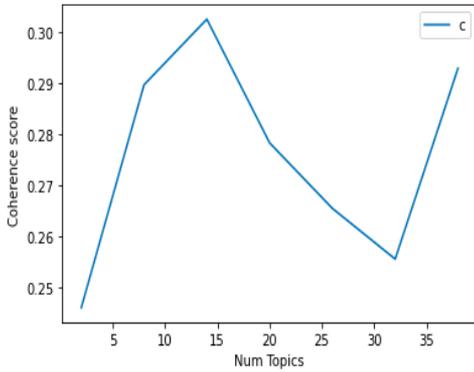


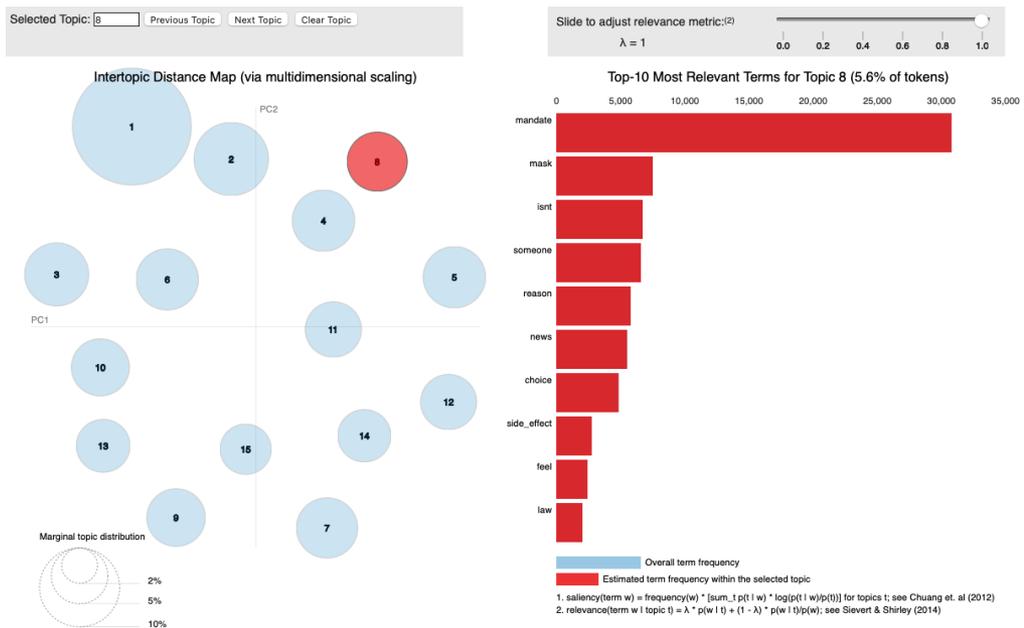
Figure 3. Graph of number of topics vs. topic coherence score. The optimal number of topics can be seen at 15, the highest peak. The coherence score at this point is around 0.30.

Figure 4. Example of “Topic 8” is shown. To the left, the circles represent each topic with the size of the circle denoting its relevance to the corpus; the larger the circle, the more relevant. To the right, since we have selected Topic 8, the top 10 words in this topic are displayed.

Latent Dirichlet Allocation (LDA) Topic Modeling

The sentiment analysis was followed by LDA topic modeling. This was used to understand specific topics being discussed in the negative tweets to better analyze vaccine hesitancy. The LDA model yielded a total of 15 latent topics. This number, which was a parameter passed into the LDA model, was determined using the topic coherence score. Topic coherence measures the semantic similarity between top words in a topic to determine how closely related they are. Figure 3 shows a graph which plots the number of topics against the coherence score. From this, it can be seen that the highest topic coherence score is at 15 topics and therefore, this was the optimal number to use for the LDA model.

The LDA model essentially generates multiple topics, each of which represent a group of words within the corpus. In this case, for each of the 15 topics, the first 10 words were also displayed (Figure 4). However, the overarching themes for each were not immediately recognizable. To interpret the themes, I looked at specific tweets that contained a few of the



ANALYZING TWITTER DATA TO UNDERSTAND COVID-19 VACCINE HESITANCY

	Overall Interpreted Theme	Top 10 Words
Topic 1	Child immunization and side-effects	vaccine, covid, death, child, keep, risk, state, trump, myocarditis, injury
Topic 2	Side-effects and mistrust in pharmaceutical companies	die, anyone, number, show, flu, money, lot, vax, school, drug
Topic 3	Efficacy and mistrust in pharmaceutical companies	fact, spread, life, country, family, nothing, month, booster, world, know
Topic 4	Efficacy and mistrust in government and pharmaceutical companies	don't, year, infection, vaccination, see, protection, end, study, natural_immunity, governor
Topic 5	Mistrust in government and Work-related	get, everyone, didn't, cause, let, they're, age, worker, tweet, kind
Topic 6	Mistrust in government and Work-related	virus, prevent, case, anything, lie, kill, won't, tell, guy, employee
Topic 7	Efficacy	doesn't, work, you're, person, protect, help, source, hospital, start, talk
Topic 8	Side-effects and Anti-vaccine mandates	mandate, mask, isn't, someone, reason, news, choice, side_effect, feel, law
Topic 9	Mistrust in pharmaceutical companies and work-related	amp, pfizer, health, immunity, rate, information, business, jab, question, folk
Topic 10	Mistrust in science and news	time, way, issue, problem, look, medium, treatment, article, ppl, research
Topic 11	Side-effects and Mistrust in government	thing, can't, go, today, trust, cdc, report, cancer, week, none
Topic 12	Child immunization and Side-effects	kid, government, doctor, body, need, hospitalization, parent, remember, happen, response
Topic 13	Mistrust in scientific community and Anti-vaccine mandates	take, day, something, think, care, science, freedom, refuse, community, line
Topic 14	Mistrust in news and scientific community	stop, shot, disease, point, misinformation, post, scientist, yesterday, gonna, consequence
Topic 15	Efficacy	i'm, part, make, word, bc, god, dr, wouldn't, save_life, lack

Table 2. This table lists the interpreted theme and the first 10 words for each of the 15 topics. Efficacy, mistrust, side-effects, and anti-vaccine mandate sentiment themes seem to be the most common.

	Tweet Examples
Efficacy	“For the millionth time, vaccines do not prevent you from getting Covid. Vaccines do not prevent you from spreading Covid & it is becoming very clear that the vaccines don’t prevent you from dying from Covid. Don’t believe me? Look at the number of breakthrough cases & deaths in MA.”
Side-effects	“Yet there are people dying from the vaccine and getting serious side effects and we do not know the long term side effects.”
Mistrust	“The main data from which [vaccine proponents] form their opinion are all from CDC, which is in the business of pushing vaccines at the behest of Big Pharma under the guise of a public health agency. Do you not see the conflict of interest?”
Anti-mandate sentiment	“Everytime a new mandate or mask/vax policy is introduced the longer I will refuse to take the vaccine.” “End restrictions. All of them. It’s over. Covid is endemic. The vaccines didn’t work. It is long past time to shift from failed government authoritarianism to personal choice.”

Table 3. The four major themes that were displayed from the 15 topics were efficacy, side-effects, mistrust, and anti-mandate sentiment. This table provides specific examples of tweets illustrating each of these themes.

top words (as generated by the model) in each topic. Table 2 displays the results for each topic. From the interpreted themes of all 15 topics, four major concepts emerged. These were: concerns regarding efficacy and side-effects, mistrust in institutions, and strong anti-mandate sentiments. This was evidenced by multiple tweets in the dataset during qualitative analysis. Examples of tweets that portray each of the four ideas are given in Table 3.

From the fifteen topics, four major themes emerged. The first theme was composed of four topics out of the fifteen (26.67%) and discussed concerns regarding efficacy. The second theme of side-effects consisted of four topics (26.67%) with two of these specifically focused on child immunization. The third major theme, made up of 10 topics (66.67%) was mis-

trust in institutions. Within this theme, 40% of topics were mistrustful of the government, 40% mistrustful of the pharmaceutical industry, and 30% were mistrustful of the news and scientific community at large. The final theme of anti-vaccine mandate sentiment was composed of five topics (33.33%) with three of these focused on specifically work-related mandates. Percentages do not add up to one hundred since some topics reflected multiple themes.

Discussion

Results indicate that there was an overwhelming amount of negative sentiment regarding the COVID-19 vaccine at the time of the mandate on Twitter whereas positive sentiment made up only a small fraction of tweets. Prior research conducted from March of 2020—the beginning of the pandemic—to January of 2021 has indicated that negative vaccine sentiment displayed on Twitter has generally declined as time progressed (Lyu et al, 2021). However, this study suggests this may not really be the case and instead, negative vaccine sentiment may still be prevalent in the later stages of the pandemic. Therefore, the phenomenon of vaccine hesitancy is only becoming more important to understand with regards to the mandate and how discourse over the mandate manifests on social media platforms like Twitter.

Qualitative analysis of the LDA topics revealed 4 overarching themes. These were concerns regarding the efficacy of the vaccine, side-effects, mistrust in institutions, and anti-vaccine mandate sentiments. Most previous literature on vaccine sentiment has been conducted in the earlier stages of vaccine rollout (Hu et al, 2021). Because of the rapid development of the vaccine, past research has identified efficacy to be one of the biggest drivers of vaccine hesitancy (Paul et al). This study, which used Twitter data from the later months of 2021, almost a year after vaccines were first distributed, still found that efficacy remains one of the major drivers of vaccine hesitancy. Previous research found that efficacy was a reason behind vaccine hesitancy, but was unable to identify explicitly where this sentiment stemmed from. However, further analysis of data from this study revealed breakthrough infections to be a leading cause of vaccine hesitancy. Because there have been many particular instances where people who were vaccinated still contracted the virus, there has been increased speculation regarding efficacy. Furthermore, earlier research has indicated uncertainty regarding side-effects to be a factor for anti-vaccine sentiments (Boucher et al, 2021). As evidenced by the tweet associated with side-effects in Table 3, this concern over side-effects has also been prevalent in this study's results. In addition, the results of this study are unique in that they show there is an increased emphasis on potential side-effects for children who are vaccinated. This is particularly evident

in topics 1 and 12. A potential reason for this could be because of the increasing pressure to get children vaccinated and more widespread availability of the vaccine for younger ages. As news sources have highlighted, later variants of COVID-19 such as Omicron have resulted in much greater hospitalization rates for children (Kozlov, 2022). As such, a larger emphasis on child vaccination could have ensued due to this risk.

Mistrust in institutions that include the government, the news, pharmaceutical companies, and the scientific community has played a large part in the negative sentiments. From analysis, it seems that mistrust of the government and the pharmaceutical industry is most prevalent. Because there has been an increased association between the private sector and the federal government through funding of vaccine development in addition to the mandates being imposed, this mistrust has only grown. The tweet associated with the theme of mistrust in Table 3 clearly illustrates this issue by discussing the possibility of a conflict of interest between the government and pharmaceutical companies. While in the past, there has been a greater emphasis on mistrust of the government and the pharmaceutical industry, this study identifies mistrust of the media and the scientific community to be present as well, although to a lesser extent.

Since this study was conducted in the one-month time frame right after the vaccine mandate for the private sector was announced, there is a larger emphasis on personal freedom and choice when it comes to the vaccine. In particular, topics eight (Figure 4) and thirteen are particularly interesting. The overarching theme for those topics conveyed that many are against the vaccine *mandates* in particular, which could potentially be a driving force behind their anti-vaccine sentiments. Prior studies have speculated that negative vaccine sentiment could potentially be correlated with negative *mandate* sentiment (Paul et al, 2021). This has been explicitly identified in my results and can be evidenced through the example tweets for anti-mandate sentiment in Table 3. Moreover, research has indicated that confidence in the vaccine itself can influence support for mandates (Ashwell et al, 2021). Therefore, it comes as no surprise that there has been such strong negative sentiment towards the federal mandate, since there has been a lot of uncertainty regarding the efficacy and many concerns regarding side-effects.

Moreover, one study wherein the researchers analyzed twitter data proposes that the individualistic mindset of Americans as opposed to a communal mindset in countries like China have played an important role in whether the people got vaccinated or not (Luo et al, 2021). Americans tend to be much more self-reliant and in health-related contexts such as this one, they wanted to have personal control over health risks. This could explain the results from this research because this particular study was solely focused on American sentiment particularly when the mandate was imposed. Since Americans have a more individualistic mindset, they value personal choice and by imposing the mandate it could have potentially had a negative emotional impact on many.

Topics 6 and 9 are interesting in that they encompass sentiments wherein vaccine hesitancy has been displayed in the context of the workplace. Specific references to employers, workers, and businesses have shown that vaccination requirements at work could be potential reasons for vaccine hesitancy as well. Previous research has not identified implications of mandates in businesses. Because this federal mandate has been centered around the private sector, it has likely fueled concerns among the vaccine hesitant with regards to how their income and work life will be affected.

Conclusion

Sentiment analysis of Twitter data following the announcement of the first federal private sector COVID-19 vaccine mandate in the US showed that there was overwhelming negative sentiment towards the vaccine. LDA portrayed 15 topics that were then qualitatively analyzed and broken down into four major themes which consisted of concerns regarding efficacy, side-effects, mistrust, and personal freedom. Overall, my results are in alignment with Almoondi and others' identified themes of the causes of negative vaccine sentiment

(Almoondi et al, 2021). Ultimately, this research has been effective in building a greater understanding of the unvaccinated perspective which will aid in targeted efforts at increasing vaccination.

Limitations

In conducting this research study, it is important to acknowledge the limitations that may have impacted the work done. For this study, twitter data taken from a Kaggle dataset was used. Kaggle is a very reputable source from which many data scientists and researchers get their data. However, it is pre-curated by others and as such, I do not have the means to definitively identify the methods through which the data was scraped. As such, there could be implicit biases inherent in the types of tweets collected. Furthermore, the specific sentiment analysis model that was used in this study called roBERTa has one of the highest accuracy rates when it comes to pre-trained sentiment analysis models. Despite this, since I had neither the capabilities nor the time to train my own model, sentiment analysis was not completely accurate. Specifically, in conducting the qualitative analysis, it was revealed that in certain cases, sentiment analysis could have reflected the underlying connotation of the tweets rather than their attitude towards the COVID-19 vaccine. Thus, this resulted in some negatively labeled tweets actually being positive. For example, one tweet reads: *"I don't push vaccines because I enjoy the fight. I don't push them because I hate freedom. I push them so I don't have to see another person I know or take care of die from something easily preventable."* This tweet received a negative score of 0.7482.

Since the Natural Language Processing sector of AI and Machine Learning is in its beginning phases of development, the LDA model, although very useful in analyzing topics, can be difficult to interpret using only the top words in each topic. Specifically, the ability to understand the overarching theme of a topic right away is sometimes difficult. Due to this, qualitative analysis was also conducted. Because it would be physically impossible to go through each tweet and qualitatively analyze it, I was only able to look at a sample of a few of these. Therefore, it may not capture the full picture in regards to topics relating to vaccine hesitancy.

Implications

The findings of this research have assisted in furthering the understanding of vaccine hesitancy. As

mentioned previously, research such as Boucher and others' study has focused on the earlier stages of vaccine rollout (Boucher et al, 2021). However, only around 65% of the US population have been fully vaccinated to date while herd immunity is estimated to be achieved only when 70-85% of the population are vaccinated. As such, vaccine hesitancy remains a large issue. The findings of this study are important to understand anti-vaccine perspectives. Only through recognizing the opposing points of views, will government officials, scientists, and news agencies be able to take effective measures to further combat low vaccination rates. Aside from vaccine mandates, other interventions have also been implemented to varying degrees of success. Previous research has revealed interventions such as offering monetary incentives to get vaccinated have had positive impacts on vaccination rates (Mercade et al, 2021). On the other hand, news sources have highlighted the fact that advertisements regarding vaccination on television have not been very successful in increasing vaccination rates nationally (Christensen). Based on the results from this study, it seems that vaccine mandates are not very effective in encouraging people to get vaccinated. Previous research has indicated that implementing mandates in the workplace could normalize the vaccine, thus increasing vaccination rates (Ashwell et al). However, it seems that in reality, this may not be the case, based on the results of this study. Therefore, mandates may not be the best approach to tackle the issue of vaccine hesitancy. Although the federal vaccine mandate has been blocked by the Supreme Court, many states and employers have enforced rules regarding vaccination. Since there has been strong opposition to federal vaccine mandates, there is a very likely possibility that there will also be negative sentiments towards state government and workplace mandates as well. In addition, based on the results, it is important to note that continued efforts in displaying accurate information regarding side-effects and efficacy rates are important to combat rampant disinformation. Moreover, it is crucial that greater attempts be made to build more trust between the people and government, pharmaceutical companies, and the scientific community. It is likely that decreased political polarization and more affirmations of unity are essential to this endeavor.

Future Directions

While this study focused on Twitter data from the US nationally, future research could focus on a comparison of vaccine hesitancy between states with lower vaccination rates and those that have higher vaccination rates. This could allow for a more focused analysis of causes behind anti-vaccine sentiment. It would aid in improving vaccination rates of states with lower vaccination rates through implementing effective methods that have been used in states with higher vaccination rates. Further, this study used pre-curated Twitter data from Kaggle. Future works could use Twitter data scraped directly from the Twitter API to get a more comprehensive and specified list of tweets to analyze. Not only that, but researchers could compare tweets from different demographics which could provide greater insight into views of different age groups. Overall, more research will need to be conducted in order to develop effective, targeted outreach programs to combat low vaccination rates.

References

- Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). Using topic modeling methods for short-text data: A comparative analysis. *Frontiers in Artificial Intelligence*, 3, 42. <https://doi.org/10.3389/frai.2020.00042>
- Alamoodi, A. H., Zaidan, B. B., Al-Masawa, M., Taresh, S. M., Noman, S., Ahmaro, I. Y. Y., Garfan, S., Chen, J., Ahmed, M. A., Zaidan, A. A., Albahri, O. S., Aickelin, U., Thamir, N. N., Fadhil, J. A., & Salahaldin, A. (2021). Multi-perspectives systematic review on the applications of sentiment analysis for vaccine hesitancy. *Computers in Biology and Medicine*, 139, 104957. <https://doi.org/10.1016/j.combiomed.2021.104957>
- Ashwell, D., Cullinane, J., & Croucher, S. M. (2021). Vaccine hesitancy and support for employer vaccine mandates. *Frontiers in Communication*, 6, 10. <https://doi.org/10.3389/fcomm.2021.780415>
- Boucher, J.-C., Cornelson, K., Benham, J. L., Fullerton, M. M., Tang, T., Constantinescu, C., Mourali, M., Oxoby, R. J., Marshall, D. A., Hemmati, H., Badami, A., Hu, J., & Lang, R. (2021). Analyzing social media to explore the attitudes and behaviors following the announcement of successful COVID-19 vaccine trials: Infodemiology study. *JMIR Infodemiology*, 1(1), e2880. <https://doi.org/10.2196/28800>
- Breuninger, K., & Kimball, S. (2022, January 14). *Supreme Court blocks Biden Covid vaccine mandate for businesses, allows health-care worker rule*. CNBC. Retrieved April 22, 2022, from <https://www.cnbc.com/2022/01/13/supreme-court-ruling-biden-covid-vaccine-mandates.html>
- Campos-Mercade, P., Meier, A. N., Schneider, F. H., Meier, S., Pope, D., & Wengström, E. (2021). Monetary incentives increase covid-19 vaccinations. *Science*, 374(6569), 879–882. <https://doi.org/10.1126/science.abm0475>
- Centers for Disease Control and Prevention. (2022). *CDC Covid Data tracker*. Centers for Disease Control and Prevention. Retrieved January 12, 2022, from https://covid.cdc.gov/covid-data-tracker/?utm_campaign=Private%20Equity%20Newsletter&utm_source=hs_email&utm_medium=email&_hsenc=p2ANqtz-_9OSt7jgR0ag2vPQVMdjmHOJhtBfAKqb3OsXS0Pr1RdsT3Wm9YvOe0XcLivUaeChn93xw#vaccinations_vacc-people-onedose-pop-5yr
- Chandrasekaran, R., Mehta, V., Valkunde, T., & Moustakas, E. (2020). Topics, trends, and sentiments of tweets about the COVID-19 pandemic: Temporal infoveillance study. *Journal of Medical Internet Research*, 22(10), e22624. <https://doi.org/10.2196/22624>
- Christensen, J. (2021, August 6). *Sharp decline in ads for covid-19 vaccines, as the number of 'persuadable' Americans diminishes*. CNN. Retrieved May 2, 2022, from <https://www.cnn.com/2021/08/06/health/covid-19-ads-why-theyve-gone-away/index.html>
- Fazel, S., Zhang, L., Javid, B., Brikell, I., & Chang, Z. (2021). Harnessing twitter data to survey public attention and attitudes towards covid-19 vaccines in the UK. *Scientific Reports*, 11(1), 1-5. <https://doi.org/10.1038/s41598-021-02710-4>
- Featherstone, J. D., Barnett, G. A., Ruiz, J. B., Zhuang, Y., & Millam, B. J. (2020). Exploring childhood anti-vaccine and pro-vaccine communities on Twitter – a perspective from influential users. *Online Social Networks and Media*, 20, 100105. <https://doi.org/10.1016/j.osnem.2020.100105>
- Featherstone, J. D., Ruiz, J. B., Barnett, G. A., & Millam, B. J. (2020). Exploring childhood vaccination themes and public opinions on Twitter: A Semantic Network analysis. *Telematics and Informatics*, 54, 101474. <https://doi.org/10.1016/j.tele.2020.101474>
- Fridman, A., Gershon, R., & Gneezy, A. (2021). Covid-19 and vaccine hesitancy: A longitudinal study. *PLOS ONE*, 16(4), e0250123. <https://doi.org/10.1371/journal.pone.0250123>
- Hu, T., Wang, S., Luo, W., Zhang, M., Huang, X., Yan, Y., Liu, R., Ly, K., Kacker, V., She, B., & Li, Z. (2021). Revealing public opinion towards covid-19 vaccines with Twitter data in the United States: Spatiotemporal Perspective. *Journal of Medical Internet Research*, 23(9), e30854. <https://doi.org/10.2196/30854>
- Jiang, J., Chen, E., Yan, S., Lerman, K., & Ferrara, E. (2020). Political polarization drives online conversations about COVID -19 in the United States. *Human Behavior and Emerging Technologies*, 2(3), 200–211. <https://doi.org/10.1002/hbe.2.202>
- Koslap-Petraco, M. (2019). Vaccine hesitancy: Not a new phenomenon, but a new threat. *Journal of the American Association of Nurse Practitioners*, 31(11), 624–626. <https://doi.org/10.1097/jnx.0000000000000342>
- Kozlov, M. (2022, February 4). *Does Omicron Hit Kids harder? scientists are trying to find out*. Nature News. Retrieved May 2, 2022, from <https://www.nature.com/articles/d41586-022-00309-x>
- Kuter, B. J., Offit, P. A., & Poland, G. A. (2021). The development of covid-19 vaccines in the United States: Why and how so fast? *Vaccine*, 39(18), 2491–2495. <https://doi.org/10.1016/j.vaccine.2021.03.077>

ANALYZING TWITTER DATA TO UNDERSTAND COVID-19 VACCINE HESITANCY

- Luo, C., Chen, A., Cui, B., & Liao, W. (2021). Exploring public perceptions of the COVID-19 vaccine online from a cultural perspective: Semantic network analysis of two social media platforms in the United States and China. *Telematics and Informatics*, 65, 101712. <https://doi.org/10.1016/j.tele.2021.101712>
- Lyu, J. C., Han, E. L., & Luli, G. K. (2021). Covid-19 vaccine-related discussion on Twitter: Topic Modeling and sentiment analysis. *Journal of Medical Internet Research*, 23(6), e24435. <https://doi.org/10.2196/24435>
- MacDonald, N. E. (2015). Vaccine hesitancy: Definition, scope and determinants. *Vaccine*, 33(34), 4161–4164. <https://doi.org/10.1016/j.vaccine.2015.04.036>
- Melton, C. A., Olusanya, O. A., Ammar, N., & Shaban-Nejad, A. (2021). Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence. *Journal of Infection and Public Health*, 14(10), 1505–1512. <https://doi.org/10.1016/j.jiph.2021.08.010>
- Oktay Ozturk. (2021). <i>800k+ COVID-19 Vaccine Tweets</i> [Data set]. Kaggle. <https://doi.org/10.34740/KAGGLE/DSV/2992050>
- Paul, K. T., Eberl, J.-M., & Partheymüller, J. (2021). Policy-relevant attitudes toward covid-19 vaccination: Associations with demography, Health Risk, and social and political factors. *Frontiers in Public Health*, 9. <https://doi.org/10.3389/fpubh.2021.671896>
- Pew Research Center. (2021, April 7). *Demographics of social media users and adoption in the United States*. Pew Research Center: Internet, Science & Tech. Retrieved January 12, 2022, from <https://www.pewresearch.org/internet/fact-sheet/social-media/#social-media-use-over-time>
- Saha, K., Torous, J., Caine, E. D., & De Choudhury, M. (2020). Psychosocial effects of the COVID-19 pandemic: Large-scale quasi-experimental study on social media. *Journal of Medical Internet Research*, 22(11), e22600. <https://doi.org/10.2196/22600>
- Sinnenberg, L., Buttenheim, A. M., Padrez, K., Mancheno, C., Ungar, L., & Merchant, R. M. (2017). Twitter as a tool for Health Research: A Systematic Review. *American Journal of Public Health*, 107(1), e1-e8. <https://doi.org/10.2105/ajph.2016.303512>
- The United States Government. (2021, November 4). *Fact Sheet: Biden Administration Announces Details of Two Major Vaccination Policies*. The White House. Retrieved January 12, 2022, from <https://www.whitehouse.gov/briefing-room/statements-releases/2021/11/04/fact-sheet-biden-administration-announces-details-of-two-major-vaccination-policies/>
- World Health Organization. (2021). *Who coronavirus (COVID-19) dashboard*. World Health Organization. Retrieved January 12, 2022, from <https://covid19.who.int/>

Unfounded Confidence? Gender Differences in Overconfidence Across Toronto Adolescents

Jessica Yu

Most people are overconfident and miscalibrated in their probability estimates. Miscalibration can have severe implications, especially in fields like medicine and finance where professionals make consequential probability judgments, and thus studying this topic has useful applications in the real world. This study examines confidence and calibration across gender and age and is guided by two inquiries: do previous findings in the gender and confidence literature apply in the adolescent age group, and does the effectiveness of internal calibration vary across Toronto adolescent males and females? A one-group pretest-posttest experiment was conducted on a sample of 70 Toronto high school students aged 14-18, with the treatment being an internal calibration method that corrects for miscalibration. There are three central findings: both male and female adolescents were underconfident, there is no statistically significant difference between male and female confidence, and internal calibration was effective for male but not female participants.

Keywords: confidence, calibration, gender, adolescents, behavioural economics

Unfounded Confidence? Gender Differences in Overconfidence Across Toronto Adolescents

We make many decisions every day. What do I want to eat for breakfast: oatmeal or toast? Should I bring an umbrella to school in case it rains? Which team do I think will win the Super Bowl? We often use probability estimates to make these decisions – if we were wondering whether to bring an umbrella or not, we need to estimate the probability that it will rain. Despite the many decisions we make, most people are grossly miscalibrated in their estimates. Calibration measures the correctness of probability assessments. Someone is well-calibrated if their probability esti-

mate equals the true proportion, and miscalibrated if their probability estimate is far from the true proportion. Generally, most people are miscalibrated and overconfident, meaning most people believe they are better calibrated than they actually are. By definition, overconfidence is “the difference between mean confidence and overall accuracy” (Brenner et al., 1996, p. 212).

Miscalibration and overconfidence have grave implications in the real world, as these estimates guide actions that have significant and consequential outcomes. For example, miscalibration in the financial world results in excessive risk-taking among finance professionals and consequently the mismanagement of funds. It therefore becomes crucial to understand how people are miscalibrated and how to correct such miscalibrations. This paper examines confidence and

calibration across gender and age, and the goal is to investigate whether previous findings in the gender and confidence literature apply in the adolescent age group.

Literature Review

Overconfidence

There is seemingly unanimous agreement among researchers that people are overconfident in their probability estimates. In a summary of research on calibration done before the 1980s, Lichtenstein et al. (1982) note that for calibration where the outcome is discrete (e.g., the estimate for “they are planning an attack”), participants are prone to miscalibration and overconfidence and overestimate how much they know. Where the outcome is continuous (e.g., the estimate for “how much will this project cost?”), the authors find overconfidence as well. For continuous outcomes, subjects tend to estimate density curves that are too narrow (i.e., density curves graph the probability of different outcomes, and a narrow density curve means that the estimated range of different outcomes is too small). This bias means that the overconfidence in continuous outcomes stems from subjects thinking they know more about uncertain quantities than they actually do. Subjects’ tendency to estimate narrower density curves than the true density curve is supported by Tversky & Kahneman (1974), who argue that the adjustment and anchoring effects are responsible for the phenomenon in question. The adjustment and anchoring effects occur when one makes estimates by starting with an initial value and then adjusting that value. Different initial values will yield different estimates, each biased towards their initial value. The anchoring effect is responsible for narrow density curves as most participants will estimate the 0.9 and 0.1 marks by adjusting up and down respectively from their initial estimates. The adjustment effect, Tversky & Kahneman (1974) argues, causes the 0.9 mark to be insufficiently large and the 0.1 mark to be insufficiently small, resulting in a narrow confidence interval.

Another recurring trend is the relationship between overconfidence and difficulty. In their experiment, Brenner et al. (1996) find that the major determinant of overconfidence is the difficulty of the

questions and not the order or manner in which they were selected, and for both random and non-random selection of questions, overconfidence increases as difficulty increases. The positive relationship between overconfidence and difficulty is corroborated by Lichtenstein et al. (1982) who argue that overconfidence is a function of the difficulty of the task when the outcome of probability estimates is discrete. As tasks get easier, overconfidence decreases. Furthermore, Lichtenstein & Fischhoff (1977) observe that the most knowledgeable subjects responding to the easiest items were actually underconfident. This is called the hard-easy effect, where there is overconfidence with hard items and underconfidence with easy items (Lichtenstein et al., 1982; Bol, 2012).

Correcting Distortions

Given the overconfidence in human probability estimates, researchers have identified internal and external calibration methods to correct these distortions. Internal calibration refers to debiasing the way a person makes decisions so that the person can make better-calibrated estimates.

Oskamp (1962) conducted experiments where subjects trained for accuracy in estimation, and subjects increased their accuracy from 67% to 73%. This brought their accuracy closer to their confidence level of 0.78, which remained the same before and after the training. Other subjects who were trained for calibration instead of accuracy lowered their confidence from 0.78 to 0.74. This new confidence approaches their accuracy of 68% which also remained the same before and after the training. Moreover, Koriat et al. (1980) conducted a study where subjects first answered general-knowledge questions. Then, additional questions were presented to the subjects but they were asked to list all the reasons they could think of that supported or contradicted the possible answers before making probability evaluations. The authors find that the process of listing reasons why one might be wrong greatly reduces overconfidence.

Another approach researchers have used is to frame the task in a way that discourages overconfidence. Lichtenstein & Fischhoff (1981) tested whether the length of instruction reduced overconfidence. One instruction was long and gave participants explanations of and warnings against overconfidence,

while the other was shorter and did not include explanations nor warnings. The authors find that there was no significant impact of instruction on overconfidence. Similarly, in one experiment, Kaustia & Perttula (2012) observe that warnings about the existence of overconfidence and narrow density curves do not increase calibration accuracy. The two groups of participants assigned normal and debiasing treatments did not produce statistically significant differences. The other two debiasing treatments conducted by the authors also did not impact overconfidence in probability assessments. Nonetheless, it is important to note that Kaustia & Perttula's (2012) experiments were composed of little data. For instance, in one of the debiasing treatments, participants were asked only four questions.

Novelty of Present Study: The Question of Gender

The topic of confidence and gender is one that concerns many fields such as psychology, sociology, economics, and education and is not limited to behavioral economics. This topic becomes essential to examine when one considers the gender inequality that persists in current society. Recent literature has also suggested that such gender gaps can be attributed to confidence in addition to discrimination and capital allocation (Sarsons & Xu, 2016). It thus becomes important to understand how gender fits into the existing literature on calibration and confidence. The current literature on confidence across gender is inconclusive; some researchers argue that there is no difference across genders while others assert that there exists difference.

Deaves et al. (2008), Lichtenstein & Fischhoff (1981), Kaustia & Perttula (2012), and Lichtenstein et al. (1982) posit that there is no statistically significant difference in confidence across genders and that both genders are overconfident. On the other hand, Sarsons & Xu (2016), Barber & Odean (2001), Jakobsson (2012), Gutierrez & Price (2016), Beyer & Bowden (1997), Fitzsimmons et al. (2021) and Boekaerts & Rozendaal (2010) argue that deeply embedded gender stereotypes, such as the idea that women and girls are bad at math and sciences, create a gender difference in confidence. Researchers who argue there to be a difference further disagree on how confidence across

genders varies. Boekaerts & Rozendaal (2010), Sarson & Xu (2016), and Barber & Odean (2001) argue that men are overconfident and women are well-calibrated in terms of their confidence, while Jakobsson (2012) and Gutierrez & Price (2016) argue that men are overconfident and women are underconfident.

However, some research that has concluded that gender differences involved confounding variables. Sarsons & Xu (2016) note that limitations exist in their research. The authors asked the male and female tenured economists at top US universities questions about macroeconomic policy. However, the answers are made public after being submitted, which influences how participants answer the questions. The authors note this may be the reason why women stay away from extremes as they may be concerned about being penalized for holding strong views. Moreover, Deaves et al. (2008) point out confounding variables in Barber & Odean's (2001) research, which uses the empirically established theory that overconfidence will result in more stock trading activity because overconfident investors overestimate their knowledge of securities and end up trading excessively (Odean, 1998). Barber & Odean (2001) find that men trade 45 percent more than women, demonstrating that men are overconfident. In direct response to Barber & Odean (2001), Deaves et al. (2008) investigate the same question but observe the same level of overconfidence and stock trading activity for men and women. Deaves et al. (2008) contend that the discrepancy between their results is likely due to Barber & Odean (2001) omitting variables such as financial education.

Nonetheless, other researchers have attempted to explain these discrepancies. Beyer & Bowden (1997) introduced the concepts of gendered tasks, with there being masculine (e.g., questions about sports), feminine (e.g., questions about soap opera TV), and neutral tasks (e.g., general knowledge questions). The authors find that female participants significantly underestimated their overall performance and showed poorer calibration in masculine tasks, while no significant difference in calibration was found for feminine or neutral tasks. Beyer & Bowden (1997) theorize that because calibration research (mentioned in the first section of this review) utilizes neutral task questions, no differences in gender are observed. Similarly, Lundeberg et al. (1992) note that differences in confidence across genders depend on domain. For

instance, in a domain like math, men were more confident than women. Another possible explanation for discrepancies in gender and confidence is the method in which confidence estimates are elicited. Lundeberg et al. (1992) find that “women and men give very different confidence scores when prospectively estimating general feelings of confidence than they do in estimating their confidence in the accuracy of their answers to specific items” (9). Beyer & Bowden (1997) further contend that female underconfidence only occurs on item-specific evaluations of confidence, and for general evaluations of confidence, both men and women are overconfident. The discrepancy in confidence and gender research may therefore be explained by whether participants are asked to evaluate general versus specific confidence.

It is worth noting that the aforementioned research has studied adults, undergraduate and graduate students, or children in primary education, and there is limited research on confidence within the adolescent time frame. However, adolescence proves to be an important period for changes in confidence. Robins et al. (2002), who traced 326,641 individuals aged 9-90 years, find that gender gaps in self-esteem emerged in adolescence for every subgroup examined. Furthermore, there has been very limited research on whether genders react to internal calibration differently. Lichtenstein & Fischhoff (1981) find that gender had no significant impact on the effectiveness of internal calibration in producing better-calibrated results. Other research, however, rarely focuses on testing internal calibration across gender, demonstrating a gap in the research. Thus, this study intends to examine whether findings in the gender and confidence literature appear in the adolescent age group and whether the effectiveness of internal calibration varies across adolescent males and females.

Method

Two inquiries are explored in this research: is there a difference in male and female confidence in the adolescent age group, and does the effectiveness of internal calibration vary across male and female adolescents? The method is designed to answer both inquiries.

Experimental Procedure

A pre-experimental one-group pretest-posttest design, which consists of a pretest measure, a treatment, and a posttest for a single group, was conducted. The pretest serves as an initial confidence assessment, and the overconfidence level is the dependent variable. The treatment, or the independent variable, is internal calibration, where participants write out one reason why they might be wrong and one reason they might be right. This treatment was first proposed by Koriati et al. (1980), who asked participants to list the reasoning behind their answers to general knowledge questions. This method has been shown to be effective in reducing overconfidence. In this study, the treatment was shortened and participants were only asked to list one reason for and one reason against their answer, instead of all the reasons for and against their answer.

The experiment had two phases. Phase One consisted of the pretest measure and addressed the first inquiry of whether there is a difference in male and female confidence. Phase Two consisted of the treatment and posttest measure and addressed the second inquiry of whether there is a difference in male and female reactions to internal calibration. In Phase One, participants answered twelve questions and assessed their confidence in the correctness of their answers for each question. In Phase Two, participants answered twelve more questions. After each question, participants wrote out one reason their chosen answer might be wrong, one reason it might be right, and then their confidence in the correctness of their chosen answer. This experimental procedure is similar to the debiasing and calibration experiments conducted by Kaustia & Perttula (2012) and Lichtenstein & Fischhoff (1981), who both applied calibration methods and measured their effectiveness after an initial confidence assessment.

Participants

The sample is composed of Toronto male and female high school students aged 14-18, inclusive. The number of participants in previous studies (e.g. Koriati et al., 1980; Kaustia and Perttula, 2012; Lichtenstein & Fischhoff, 1981; Sarson & Xu, 2016) ranged from 40 to 150, and thus it was ideal to have a number of participants within that range. This study had at least thirty

female and thirty male participants, which satisfies the Normality condition and allows for assumption of Normality for statistical analysis.

Instrument: Questionnaire

Participants filled out a Google Form questionnaire containing 24 dichotomous multiple-choice general knowledge questions. It is assumed that participants have no expert knowledge that helps them answer the questions and that participants did not search for answers.

Different types of questions can differently influence participants' confidence levels. Neutral tasks induce less bias in gender confidence while gendered tasks induce more bias (Beyer & Bowden, 1997). Previous calibration research has also used neutral tasks in the form of general knowledge questions. Thus, to avoid incorporating bias into the questionnaire, general knowledge questions were used.

The questions in this research are selected from Nelson & Narens (1980), who used general knowledge questions similar to those used by Lichtenstein & Fischhoff (1997) and Lichtenstein & Fischhoff (1981). It is also necessary to ensure that the selected questions had varying difficulties. This was possible because Nelson & Narens (1980) ranked their 300 questions by difficulty (through displaying the percentage of participants who answered a particular question correctly). Questions in this research were selected from four categories: those that had 0.3-0.4, 0.4-0.5, 0.5-0.6, and 0.6-0.7 percent of Nelson & Narens' (1980) participants answer correctly. Using a random number generator, three questions were chosen from each difficulty category for both the pretest and posttest, resulting in a total of 24 questions. While the number of questions in previous studies ranged from 4 to 300, 12 questions per phase better-respected participants' time and provided sufficient data. See Appendix A for a copy of the distributed questionnaire.

Ethical Considerations

All participants provided a signature for informed consent prior to beginning the questionnaire. For participants under 16, parents or guardians also provided a signature (see Appendix B for details). All questions on the questionnaire were also optional and partici-

pants could exit at any time. If participants withdrew, their answers were not be submitted and the researcher does not have access to their answers. Participants were also reminded that they can withdraw from the research, skip questions, and that their participation is completely voluntary.

Moreover, as part of the internal calibration process, the questionnaire asks participants to identify one reason their answer might be wrong and one reason their answer might be right. Due to its association with self-esteem, asking for one reason someone might be wrong may pose psychological or emotional risks. To mitigate this risk, a sentence warning participants of upcoming questions about confidence was added before the internal calibration process. This research was approved by the Internal Review Board (IRB) at the researcher's institution.

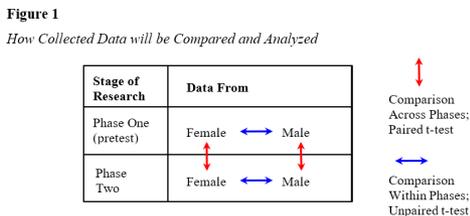
Data Analysis

The research generates two sets of numerical data. The first is the pretest measure from Phase One and the other is the posttest measure from Phase Two. Having two sets of data makes it possible to compare gender overconfidence both within phases and across phases (see Figure 1). Comparing within phases means examining two groups, which reveals gender differences in confidence and answers the first inquiry. Comparing across phases means examining two results from a group with there being two groups of interest: male and female students. This comparison reveals the effectiveness of internal calibration for each gender, answering the second inquiry.

To analyze the data, each participant's overconfidence level (mean confidence - mean accuracy) in both phases were calculated. To compare within phases, two unpaired t-tests were conducted. One unpaired t-test compared gender differences in the pretest and the other t-test, gender differences in the posttest (see Figure 1). The null hypothesis is that there is no difference across the mean confidence of male and female participants, with the alternative hypothesis being that there is a difference in mean confidence. The hypothesis was tested at a significance level of 0.05. To compare across phases, two paired t-tests were conducted, one comparing pretest and posttest confidence of male students and the second, pretest

and posttest confidence of female students (see Figure 1). For both paired t-tests, the null hypothesis is that there is no difference in the pretest and posttest measures, while the alternative hypothesis is that there is a difference in the measures. The hypothesis was tested at a significance level of 0.05. When a difference was found, secondary analyses was conducted with two additional paired t-tests for the data set with the difference (male or female) to specify the direction of the difference. The alternate hypotheses – that the pretest measure is greater than the posttest measure and vice versa – was tested at a significance level of 0.05. A 95% confidence interval was also conducted to investigate the size of the difference.

Figure 1
How Collected Data will be Compared and Analyzed



Results and Discussion

There were in total 80 responses to the questionnaire. Responses that were incomplete or were completed by unqualified participants were eliminated, resulting in 40 responses from male participants and 30 responses from female participants. Male participants had an average pretest overconfidence level of -0.060 and an average posttest overconfidence level of -0.151 (see Table 1). Female participants had an average pretest overconfidence level of -0.083 and an average posttest overconfidence level of -0.093 (see Table 2). Negative confidence values indicate that both female and male participants are underconfident. Additionally, comparing the pretest and posttest scores reveal that male participants began more confident than female participants, but became less confident than female participants following internal calibration. It is important to note that internal calibration did not necessarily make participants better calibrated. Generally, participants became more underconfident in the posttest, but they were also underconfident in the pretest. This internal calibration method thus caused most participants to become less confident and more miscalibrated.

Following analysis, it was found that there was no evidence of a statistically significant difference between male and female overconfidence in the pretest measure, $t(54) = -0.666, p = 0.508 > 0.05$. Similarly, there was no statistically significant difference between male and female overconfidence for the posttest measure, $t(65) = 1.625, p = 0.1089 > 0.05$. On the other hand, comparison across phases showed that internal calibration was not effective for female subjects, $t(29) = 0.826, p = 0.415 > 0.05$. However, for male participants, internal calibration was effective, $t(39) = 3.185, p = 0.0014 < 0.05$. Thus, two additional paired t-tests were conducted for the male pretest and posttest measures in order to evaluate the direction of the difference. The tests revealed that there is a statistically significant decrease in overconfidence level from pretest to posttest, $t(39) = 3.185, p = 0.0014 < 0.05$. The direction of difference is also the sign of the difference in means, which is positive. A positive value means that the posttest measure is smaller than the pretest measure, revealing a decrease in overconfidence level from the pretest to the posttest. A 95% confidence interval conducted following the t-test to see the size of

this decrease found that confidence reduced by 0.041 to 0.19. To ensure statistical robustness, one final unpaired t-test that examines whether there was a statistically significant difference between the differences of the pretest and posttest for each gender was conducted (i.e. (Female_pre - Female_post) compared to (Male_pre - Male_post)). The results indicate that the difference is insignificant at a 5% level but significant at a 10% level, $t(67) = 1.745$, $p = 0.0856 < 0.10$, which helps support the claim that the difference was present for male but not for female participants. A summary of the results of the statistical analysis can be found in Table 3.

Despite the overall results, there were instances where participants became better calibrated after the internal calibration exercise. For instance, in Table 1, Male Participant 3 was underconfident by a measure of 0.0417 in the pretest but became nearly perfectly calibrated in the posttest following internal calibration. Similarly, in Table 2, Female Participant 8 was overconfident by a measure of 0.0375 and following internal calibration, was overconfident by a measure of only 0.0075. It is also noteworthy that after internal calibration, certain participants were more confident. For example, Female Participant 15 was perfectly calibrated in the pretest but became overconfident by a measure of 0.1. Another interesting observation is that certain questions were more difficult for participants. The questions, "What is the name of the unit of measure that refers to a six-foot depth of water?" and "In which city is the US Naval Academy located?" had only 50% of the participants answer correctly (which is close to the percentage of people who answered the questions correctly in Nelson & Narens (1980)'s research). Other questions were much easier; the questions "What is the last name of Batman's secret identity in the Batman comics?" and "What animal runs the fastest?" only had 2 participants and 1 participant answer incorrectly, respectively. One final interesting aspect to note is that most participants used extreme values to communicate their confidence, using values such as 100%, 99%, and 0%. Moreover, when writing out reasons for their chosen answer, many participants wrote out that they had guessed and then made a confidence assessment of 0%, even though there is a 50% chance that their chosen answer is correct.

GENDER DIFFERENCES IN OVERCONFIDENCE ACROSS TORONTO ADOLESCENTS

Participant	Phase 1			Phase 2		
	Accuracy	Confidence	Overconfidence	Accuracy	Confidence	Overconfidence
1	0.8333	0.57833	-0.25497	0.8333	0.2542	-0.5791
2	0.41667	0.208333	-0.208337	0.58333	0.2941667	-0.2891633
3	0.5	0.458333	-0.041667	0.5	0.5	0
4	0.75	0.6667	-0.0833	0.6667	0.641667	-0.025033
5	0.41667	0.70427	0.2876	0.8333	0.7491667	-0.0841333
6	0.6667	0.63333	-0.03337	0.8333	0.61667	-0.21663
7	0.6667	0.9	0.2333	1	0.91667	-0.08333
8	0.6667	0.591667	-0.075033	0.8333	0.6083	-0.225
9	0.58333	0.4541667	-0.1291633	0.75	0.5775	-0.1725
10	1	0.754167	-0.245833	1	0.84583	-0.15417
11	0.6667	0.595833	-0.070867	0.66667	0.575	-0.09167
12	0.8333	0.725	-0.1083	0.83333	0.625	-0.20833
13	0.6667	0.654167	-0.012533	0.91667	0.71667	-0.2
14	1	1	0	1	0.94167	-0.05833
15	0.75	0.6292	-0.1208	0.6667	0.5625	-0.1042
16	0.8333	0.84167	0.00837	0.8333	0.71667	-0.11663
17	0.75	0.6792	-0.0708	0.8333	0.5125	-0.3208
18	0.91667	0.74916	-0.16751	0.8333	0.65	-0.1833
19	0.8333	0.8333	0	0.75	0.8792	0.1292
20	0.75	0.875	0.125	0.75	0.8175	0.0675
21	1	0.9667	-0.0333	0.91667	0.9083	-0.00837
22	0.5833	0.7342	-0.1509	0.6667	0.84167	0.17497
23	0.91667	0.8625	0.05417	0.75	0.89167	0.14167
24	0.8333	0.75	0.0833	0.8333	0.7125	-0.1208
25	0.5833	0.44167	-0.14163	0.8333	0.54583	-0.28747
26	0.8333	0.8625	0.0292	0.8333	0.9	0.0667
27	0.91667	0.89167	-0.025	1	0.7083	-0.2917
28	0.8333	0.775	-0.0583	1	0.5583	-0.4417
29	0.8333	0.6708	-0.1625	0.91667	0.64167	-0.275
30	0.75	0.4833	-0.2667	1	0.625	-0.375
31	0.91667	0.729167	-0.187503	0.75	0.775	0.025
32	0.833	0.9375	0.1045	1	0.84167	-0.15833
33	0.75	0.68667	-0.06333	1	0.6083	-0.3917
34	0.9167	0.95	0.0333	0.9167	0.6833	-0.2334
35	0.9167	0.94167	0.02497	0.9167	0.8333	-0.0834
36	0.58333	0.4333	-0.15003	0.6667	0.3333	-0.3334
37	0.8333	0.7833	-0.05	0.7272	0.7272	0
38	0.8333	0.525	-0.3083	0.75	0.525	-0.225
39	1	0.925	-0.075	0.91667	0.74167	-0.175
40	1	0.91667	-0.08333	0.8333	0.70833	-0.12497
Average			-0.05986			-0.15081

Table 1 Accuracy, Confidence, and Overconfidence Level of Male Participants for Pretest and Posttest

GENDER DIFFERENCES IN OVERCONFIDENCE ACROSS TORONTO ADOLESCENTS

Table 2
Accuracy, Confidence, and Overconfidence Level of Female Participants for Pretest and Posttest

Participant	Phase 1			Phase 2		
	Accuracy	Confidence	Overconfidence	Accuracy	Confidence	Overconfidence
1	0.75	0.7075	-0.0425	0.8333	0.61083	-0.22247
2	0.5833	0.791667	0.208367	0.6667	0.70833	0.04163
3	0.6667	0.5575	-0.1092	0.916667	0.641667	-0.275
4	0.916667	0.575	-0.341667	1	0.67967	-0.32033
5	0.6667	0.43167	-0.23503	0.8333	0.355	-0.4783
6	0.8333	0.7825	-0.0508	0.916667	0.80833	-0.108337
7	0.8333	0.8875	0.0542	0.75	0.9092	0.1592
8	0.75	0.7875	0.0375	0.6667	0.6742	0.0075
9	0.91677	0.66667	-0.2501	0.916667	0.674167	-0.2425
10	0.75	0.59667	-0.15333	0.6667	0.51667	-0.15003
11	0.91667	0.74917	-0.1675	0.75	0.725	-0.025
12	0.5	0.6708	0.1708	0.6667	0.591667	-0.075033
13	0.58333	0.69167	0.10834	0.91667	0.795	-0.12167
14	0.8333	0.6175	-0.2158	0.6667	0.5125	-0.1542
15	0.75	0.75	0	0.5	0.6	0.1
16	0.75	0.925833	0.175833	0.8333	0.7825	-0.0508
17	0.6667	0.57083	-0.09587	0.75	0.5375	-0.2125
18	0.6667	0.60833	-0.05837	0.75	0.7333	-0.0167
19	1	0.885	-0.115	0.91667	0.93083	0.01416
20	0.5	0.46667	-0.03333	0.5833	0.50583	-0.07747
21	0.91667	0.9383	0.02163	0.8333	0.8625	0.0292
22	0.6667	0.59167	-0.07503	0.5	0.6	0.1
23	0.5833	0.6667	0.0834	0.667	0.6216	-0.0454
24	1	0.80833	-0.19167	0.5833	0.7742	0.1909
25	0.8333	0.6292	-0.2041	0.8333	0.67083	-0.16247
26	0.75	0.725	-0.025	0.91667	0.875	-0.04167
27	0.91667	0.7625	-0.15417	0.91667	0.7475	-0.16917
28	0.8333	0.70833	-0.12497	0.91667	0.73913	-0.17754
29	0.75	0.29	-0.46	0.75	0.5717	-0.1783
30	0.5833	0.3417	-0.2416	0.5833	0.4433	-0.14
Average			-0.08283			-0.09341

Table 3
Results of Statistical Analysis

Type	Subject of Comparison	Statistical Test	Alternate Hypothesis	Significance Level / Confidence Interval	Result (P-Value, interval)
Comparison Within Phases	Pretest	Unpaired t-test	$\mu_1 \neq \mu_2$	5%	
	Posttest	Unpaired t-test	$\mu_1 \neq \mu_2$	5%	
Comparison Across Phases	Female pretest and posttest	Paired t-test	$\mu_1 \neq \mu_2$	5%	
	Male pretest and posttest	Paired t-test	$\mu > \mu_o$	5%	
		Confidence Interval		95%	0.041 to 0.19
Robustness Check	Female_pre - Female_post	Unpaired t-test	$\mu_1 \neq \mu_2$	10%	0.0856
	Male_pre - Male_post				

Discussion

This research attempted to answer two inquiries: whether previous findings in the gender and confidence literature apply in the adolescent age group and whether the effectiveness of internal calibration varies across adolescent males and females. The results indicate three central findings. Both male and female adolescents were underconfident, there is no statistically significant difference between male and female overconfidence, and internal calibration was shown to be effective for male but not female participants.

Unlike previous literature in the field, both male and female adolescents were found to be underconfident, not overconfident. There are several potential explanations. Adolescence is a stage of growth, and therefore many participants may be in the process of gaining confidence, and thus are not very confident yet. Alternatively, the hard-easy effect, where there is overconfidence with hard items and underconfidence

with easy items, may have occurred. The questions may have been too easy for the high school level, subsequently causing participants to be underconfident. There was also no statistically significant difference in confidence across gender in either the pretest or the posttest. Such findings are in line with Deaves et al. (2008), Lichtenstein & Fischhoff (1981), Kaustia & Perttula (2012), and Lichtenstein et al. (1982), who observe no gender difference in confidence. The lack of difference may be because this study used general knowledge questions, a neutral task that does not provoke gender differences in confidence, as suggested by Beyer & Bowden (1997). Beyer & Bowden (1997) also find that confidence level is impacted by the type of assessment of confidence, and thus, another explanation for the observed data may be that this questionnaire asked participants to assess item-specific rather than general confidence. Lundeberg et al. (1992) also note that there is a gender difference in how women and men react to item-specific and general confidence. This gender difference, however, was not observed in

this study. Meanwhile, it is found that the effectiveness of internal calibration varies across male and female adolescents. Internal calibration was found to reduce confidence levels in male participants, but not female participants. Additionally, calibration did not necessarily make participants better calibrated, but rather less confident. Due to limited research on whether internal calibration differed across genders, it is unclear why this phenomenon was observed.

Due to difficulty obtaining data, recruited participants were from both co-ed and single-sex schools. However, the type of education may impact the confidence of participants. Subjects were also not stratified by class or grade due to limitations in data collection. Both variables could have impacted the validity of the results, as class and grade may affect both confidence and knowledge level. Additionally, while there were around the same number of participants as in previous studies, a greater sample size would have been more representative of the population in question. Another limitation to note is the phenomenon of regression to the mean, which states that if the measure from one sample is extreme, the following measure from the same sample will likely be closer to the population mean. Applied to this study, this phenomenon suggests that the statistically significant difference between the pretest and posttest measure for male participants may have been caused by regression towards the mean rather than the internal calibration treatment. Furthermore, the researcher's background could have led to a biased interpretation of the data. Because of the researcher's gender (female) and background, there is a tendency to spot more gender inequality. Lastly, due to the social desirability bias, participants may have been tempted to google answers or fill out the form together with another participant. Both of these events, if they occurred, would have hindered the validity of the data.

Conclusion

This study contributes to the field of calibration research. Adolescence is an understudied age group with regards to gender, calibration, and confidence, and there are also disagreements among researchers regarding the relationship between confidence and gender. This research fills in such gaps for it tests not

only confidence level but also the effectiveness of internal calibration across gender and age. This study has provided more evidence and data to better understand the relationship between confidence and gender. There is a lack of research on how and if genders react to internal calibration differently; research rarely focuses on testing internal calibration across gender. Another contribution of this study is thus the finding that the effectiveness of internal calibration varies across male and female adolescents.

This study was more exploratory, and further research can focus on the explanatory, answering several key questions. First, why are adolescents underconfident, and why is no gender difference observed despite previous literature stating that adolescence is the age group where differences emerge? Does this indicate that societal ideas have changed and greater emphasis is placed on gender equality? Furthermore, why did internal calibration result in more miscalibration in this study, and why was internal calibration effective for male but not female participants? A possible method to investigate these questions is to code and analyze the reasoning that participants wrote out during internal calibration to help us understand the causes of underconfidence and overconfidence. Themes and their frequencies across genders can be examined, and this qualitative aspect will enable explorations of why certain observations occurred, rather than just what occurred. Future research should also examine whether the results produced in this paper can be replicated, as the adolescent age group remains understudied. Moreover, further research can expand the data set and stratify participants by class, education, and grade. Testing whether these factors influence confidence levels can prompt further questions and help researchers answer how social structures, in addition to human behaviours, impact confidence.

Internal calibration methods that correct overconfidence have wide applications. Miscalibration can have severe implications in the real world, especially in fields like medicine and finance where professionals are expected to make consequential probability judgements. For instance, miscalibrated finance professionals can put their clients and wealth and the economy in general at risk and miscalibrated medical professionals can put their patients' lives at risk. Thus, testing a method that can effectively reduce miscali-

bration is useful in the real world. While the method tested in this research did not necessarily assist participants in becoming better calibrated, it is nonetheless a step towards a better understanding of internal calibration; this study investigated how the effectiveness of internal calibration varies across groups, and knowing for which groups a certain calibration method is effective for will allow us to hone and adapt those methods to increase their effectiveness and applicability to the real world. Decision-making is prevalent in our lives, and continued research on calibration and confidence can greatly help us make better decisions and avoid losses and large risks.

References

- Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, 116(1), 261–292. <https://doi.org/10.1162/003355301556400>
- Beyer, S., & Bowden, E. M. (1997). Gender differences in Self-perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin*, 23(2), 157–172. <https://doi.org/10.1177/0146167297232005>
- Boekaerts, M., & Rozendaal, J. S. (2010). Using multiple calibration indices in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning and Instruction*, 20(5), 372–382. <https://doi.org/10.1016/j.learninstruc.2009.03.002>
- Bol, L., & Hacker, D. J. (2012). Calibration Research: Where Do We Go from Here? *Frontiers in Psychology*, 3, 1–6. doi:10.3389/fpsyg.2012.00229
- Brenner, L. A., Koehler, D. J., Liberman, V., & Tversky, A. (1996). Overconfidence in probability and frequency judgements: A critical examination. *Organizational Behavioral and Human Decision Processes*, 65(3), 212–219. doi:10.1006/obhd.1996.0021
- Deaves, R., Lüders, E., & Luo, G. Y. (2008). An experimental test of the impact of overconfidence and gender on trading activity. *Review of Finance*, 13(3), 555–575. <https://doi.org/10.1093/rof/rfn023>
- Fitzsimmons, T. W., Yates, M. S., & Callan, V. J. (2021). Lean in? the role of single sex schools in the gendering of confidence in high school adolescents. *Australian Journal of Career Development*, 30(2), 139–149. <https://doi.org/10.1177/10384162211012045>
- Gutierrez, A. P., & Price, A. F. (2016). Calibration between undergraduate students' prediction of and actual performance: The role of gender and performance attributions. *The Journal of Experimental Education*, 85(3), 486–500. <https://doi.org/10.1080/00220973.2016.1180278>
- Jakobsson, N. (2012). Gender and confidence: Are women underconfident? *Applied Economics Letters*, 19(11), 1057–1059. <https://doi.org/10.1080/13504851.2011.613745>
- Kaustia, M., & Perttula, M. (2012). Overconfidence and debiasing in the financial industry. *Review of Behavioural Finance*, 4(1), 46–62. doi:10.1108/19405971211261100
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning & Memory*, 6(2), 107–118. doi:10.1037/0278-7393.6.2.107
- Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior & Human Performance*, 20(2), 159–183. [https://doi.org/10.1016/0030-5073\(77\)90001-0](https://doi.org/10.1016/0030-5073(77)90001-0)
- Lichtenstein, S., & Fischhoff, B. (1981). The Effects of Gender and Instructions on Calibration. *Perceptronics*, 1–20.
- Lichtenstein, S., Fischhoff, B., & Phillips, L. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under Uncertainty: Heuristics and Biases* (pp. 306–334). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511809477.023
- Lundeberg, M. A., Fox, P. W., & LeCount, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology*, 86(1), 114–121. <https://doi.org/10.1037/0022-0663.86.1.114>
- Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior*, 19(3), 338–368. [https://doi.org/10.1016/s0022-5371\(80\)90266-2](https://doi.org/10.1016/s0022-5371(80)90266-2)
- Odean, T. (1998). Volume, volatility, price, and profit when all traders are above average. *The Journal of Finance*, 53(6), 1887–1934. <https://doi.org/10.1111/0022-1082.00078>
- Oskamp, S. (1962). The relationship of clinical experience and training methods to several criteria of clinical prediction. *Psychological Monographs: General and Applied*, 76(28), 1–27. <https://doi.org/10.1037/h0093849>
- Robins, R., Trzesniewski, K. H., Tracy, J., Gosling, S., & Potter, J. (2002). Global self-esteem across the life span. *Psychology and Aging*, 17(3), 423–434. <https://doi.org/10.1037/0882-7974.17.3.423>
- Sarsons, H., & Xu, G. (2021). Confidence men? Evidence

on confidence and gender among top economists.
AEA Papers and Proceedings, 111, 65–68. <https://doi.org/10.1257/pandp.20211086>

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124–1131. <http://www.jstor.org/stable/1738360>

Appendix

Appendix A

The Distributed Google Form Questionnaire

Phase 1

Instructions

The following section contains twelve multiple-choice general knowledge questions. For each question, there are two options to choose from. Please choose the option you believe to be correct. Then, following each question, you will be asked to assess your confidence in your answer to the question. Please enter your confidence level in percentage (i.e., 0% - 100%) and avoid any decimal values (e.g., 1.5%)

PLEASE DO NOT GOOGLE ANSWERS TO QUESTIONS. Your answers will remain anonymous and the researcher has no means of identifying who you are.

Your participation is completely voluntary. You are free to refuse to answer any questions. You are also under no obligation to participate and are free to withdraw at any time without consequence. Your decision to withdraw will not influence your relationship with the researcher in any way. If you want to withdraw now, simply close this tab. If you choose to withdraw, your questionnaire answers will not be submitted and the researcher will not have access to your answers. However, if you choose to withdraw after completing the questionnaire, the researcher will not be able to locate and remove your data due to anonymization.

Questions

What is the name of the poker hand in which all of the cards are of the same suit?

Full house

Flush

What is the name of the legendary one-eyed giant in Greek mythology?

Minotaur

Cyclops

What is the name of the city in Italy that is known for its canals?

Florence

Venice

What is the name of the unit of measure that refers

GENDER DIFFERENCES IN OVERCONFIDENCE ACROSS TORONTO ADOLESCENTS

to a six-foot depth of water?

- Furlong
- Fathom

What is the name of the ocean that is located between Africa and Australia?

- Indian Ocean
- Atlantic Ocean

What is the name of the chapel whose ceiling was painted by Michelangelo?

- St. Paul's chapel
- Sistine chapel

Of which country is Buenos Aires the capital?

- Argentina
- Ecuador

What is the name of the furry animal that attacks cobra snakes?

- Badger
- Mongoose

In which city is the US Naval Academy located?

- Annapolis
- Baltimore

What is the name of the Roman emperor who fled while Rome burned?

- Nero
- Augustus

What is the last name of the author of the book "1984"?

- Orwell
- Dickens

What is the name of the collar bone?

- Clavicle
- Scapula

After each question, participants are asked:

How confident are you in your previous answer?

Please enter your confidence level in percentage (i.e., 0% - 100%) and avoid any decimal values (e.g., 1.5%).

Phase 2

Instructions

WARNING: This section contains questions related to confidence and self-esteem. If you think this could be triggering for you, please exit out of this questionnaire by closing the tab.

Similar to the last section, this section contains twelve multiple-choice general knowledge questions. For each question, there are two options to choose from. Please choose the option you believe to be correct. Next, you will be asked to list one reason your answer could be wrong, and one reason it could be right. Then, you will be asked to assess your confidence in your answer to the question. Please enter your confidence level in percentage (i.e., 0% - 100%) and avoid any decimal values (e.g., 1.5%).

PLEASE DO NOT GOOGLE ANSWERS TO QUESTIONS. Your answers will remain anonymous and the researcher has no means of identifying who you are.

Your participation is completely voluntary. You are free to refuse to answer any questions. You are also under no obligation to participate and are free to withdraw at any time without consequence. Your decision to withdraw will not influence your relationship with the researcher in any way. If you want to withdraw now, simply close this tab. If you choose to withdraw, your questionnaire answers will not be submitted and the researcher will not have access to your answers. However, if you choose to withdraw after completing the questionnaire, the researcher will not be able to locate and remove your data due to anonymization.

Questions

What is the last name of the actor who played Rhett Butler in "Gone with the Wind"?

- Gable
- Chaplin

What is the name of the small Japanese stove used for outdoor cooking?

- Hibachi
- Kamado

What is the name for the astronomical bodies that enter the Earth's atmosphere?

GENDER DIFFERENCES IN OVERCONFIDENCE ACROSS TORONTO ADOLESCENTS

Asteroids
Meteors

What animal runs the fastest?
Cheetah
Cougar

What is the name of the automobile instrument that measures mileage?
Ammeter
Odometer

What is the unit of electrical power that refers to a current of one ampere at one volt?
Watt
Joule

What is the last name of the author who wrote "The Old Man and The Sea"?
Twain
Hemmingway

What is the last name of the woman who began the profession of nursing?
Bell
Nightingale

What is the last name of Batman's secret identity in the Batman comics?
Wayne
Stark

What is the last name of the man who invented the telegraph?
Wright
Morse

What is the capital of New York?
Albany
Rochester

Which country was the first to use gunpowder?
Britain
China

After each question, participants are asked:
Please write down (a) one reason supporting your chosen answer and (b) one reason contradicting it.

How confident are you in your previous answer?
Please enter your confidence level in percentage (i.e., 0% - 100%) and avoid any decimal values (e.g., 1.5%).

Demographic Information

What is your gender?

Female

Male

Other/prefer not to say

What is your grade?

9

10

11

12

Appendix B

Informed Consent Form

Project title

Unfounded Confidence? Gender differences in overconfidence across Toronto adolescents

Study investigators

Principal Investigator: -----

Study Supervisor: -----

Invitation to participate

You are being invited to participate in a research study on gender differences in overconfidence and the reaction to internal calibration (a process to decrease overconfidence) across Toronto adolescents. Choosing whether or not to participate is entirely your choice. If you decide not to participate, there will be no negative impacts on your relationship with the researcher. The information provided in this form tells you about what is involved in the research, what you will be asked to do, and any potential risks or benefits. Please read this form carefully, take all the time you need, and ask any questions you may have.

Consent is an ongoing process. During the research study, we will tell you about any significant finding that could affect your willingness to continue to participate in this study.

Purpose of study

Adolescence has been noted to be a period when gender differences in self-esteem emerge. This study therefore tests whether there are gender differences in

GENDER DIFFERENCES IN OVERCONFIDENCE ACROSS TORONTO ADOLESCENTS

confidence level at adolescence. Furthermore, because overconfidence can have severe implications in the real world, such as when professionals are expected to make consequential probability judgments, testing a method that can effectively reduce miscalibration has useful implications.

What you will be asked to do

As a participant, you will fill out an online questionnaire that should take no more than twenty minutes to complete.

In this questionnaire, you will first answer twelve multiple-choice general knowledge questions and then assess your confidence in your answer for each question. The aim is to see your overconfidence level and whether your accuracy matches your confidence assessments. Then, you will answer twelve more general knowledge questions, and after each question, you will complete an internal calibration exercise and assess your confidence in your answer.

The researcher will then calculate overconfidence levels and group the data by gender in order to test whether there is a gender difference in confidence both before and after the internal calibration process.

Who can take part in the study?

The researcher hopes to have around fifty participants. Participants must be a Toronto adolescent that is between the ages of 14 and 18, inclusive. The researcher hopes to have similar numbers of female and male students.

Possible risks and benefits

Risks: The risks associated with the study are minimal. The harms or discomforts are no greater than what an average person would expect to experience in everyday life. However, some sections of the questionnaire will ask you to list reasons why you might be wrong. If you feel uncomfortable, you are free to refuse to answer any questions and/or withdraw at any time. Furthermore, by agreeing to participate in this research you are not giving up or waiving any legal rights in the event that you are harmed during the research.

Benefits: There is no guarantee that you will benefit directly from participating in this study. However, your participation will advance knowledge about gender differences in overconfidence among adolescents, which may indirectly benefit you in the future.

Privacy and confidentiality

In the questionnaire, you will only be asked to state

your gender and grade. Your personal data used for sampling purposes (e.g., your school, grade, age) will be stored in an electronic file protected by a password. Only the researchers involved in the study will have access to this file.

Reporting of results

Although the project outcomes will be determined by the research findings, possible research products will include: articles in academic journals and a report for the College Board. The researcher will aggregate the data and only report group results, and therefore, you will not be identified in any way in the reports. If you wish to be informed of the results of the research, please indicate this on the signature page below.

Withdrawing from the study

Your participation is completely voluntary. You are under no obligation to participate and are free to withdraw at any time without consequence. Your decision to withdraw will not influence your relationship with the researcher in any way. If you choose to withdraw while completing your questionnaire, your questionnaire answers will not be submitted and the researcher will not have access to your answers. However, if you choose to withdraw after completing the questionnaire, the researcher will not be able to locate and remove your data due to anonymization.

Conflicts of interest

None of the researchers have any conflicts of interest in this study.

Questions and contact information

If you have any questions about the study, please contact: -----

If you have any questions about your rights as a research participant, please contact: -----

Signature Page

Project title: Unfounded Confidence? Gender differences in overconfidence across Toronto adolescents

Lead researcher: -----

Statement of consent

By signing this form, I agree that:

The study has been explained to me

All my questions have been answered

Possible harm and discomforts and possible benefits (if any) of this study have been explained to me

I have been told that my personal information will be kept confidential

In addition, I understand that:

I have the right not to participate and the right to

GENDER DIFFERENCES IN OVERCONFIDENCE ACROSS TORONTO ADOLESCENTS

stop at any time

- I may refuse to participate without consequence
- I have a choice of not answering specific questions
- I am free now, and in the future, to ask any questions about the study

No information that would identify me will be released or printed without asking me first I will receive a signed copy of this consent form

 Participant Name
 Participant Signature Date

Please provide an email address below if you would like to be sent a summary of the study results.

Email address: _____

If you are under 16, please ask your parent/guardian to sign here: _____

Signature of the person obtaining assent

By signing this form, I attest that:

I have explained the study to the prospective participant

I answered all of their questions

I provided a copy of this assent form to the participant

The participant seemed to understand the assent form and agreed to participate

Name:

Signature:

Date:

Contributors

Alexander Barry is a Junior at Newark Charter School. He is interested in Mechanical Engineering and Materials Science.

Brendon Frankel is a senior at Hampton High School with a keen interest in pathogen-related science and research. He plans to expand upon the implications of the presented research, focusing on antibiotic resistant bacteria, as well as explore the possibilities of immunologic or viral research in the future.

Uma Kamath is currently a 11th grader at the Obra D. Tompkins High School in Katy, Texas. Uma has been conducting independent research since she was in middle school. The research project she completed in her 7th grade was awarded with an Honorable Mention in Mathematics Category at the Science and Engineering Fair of Houston. Uma holds leadership positions or participates in the Science Fair Scholars Club, TEDx Club, Future Business Leaders of America Club, National Honor Society, and Artists for Healing Club. In addition to enjoying Science and Math classes, Uma is an accomplished artist. Her high school artwork has been selected for the Houston Rodeo School Art Auction twice in two years. Uma enjoys helping others and believes art has a lot of power to heal others. She founded the Artists for Healing Club in her high school and enjoys working with her fellow Club members to complete service projects. As a recognition for her service at the start of the COVID-19 pandemic, she was featured in the Houston Chronicle in June 2020. Uma has been conducting art classes at various senior living homes, which led her to venture into this research project. Most of all, though, Uma loves spending time with her loved ones and being able to make memories with them, especially through travel across the world.

Dhiraj Kanneganti is a junior at DuPont Manual HS in Louisville, KY. Dhiraj has always been interested in machine learning and web design, so over the past year, he has been working on a research project that combines both of his passions. In his free time, Dhiraj enjoys watching football, baking, and listening to podcasts.

Alexandra McWatters is a rising senior at Newark Charter High School in Newark, DE, USA. She is heavily involved in the school's extracurricular offerings, including, but not limited to, Science Olympiad, Technology Student Association, Women in STEM Experience, and Lacrosse. She tied her passion for mathematics and computer programming in her research.

Manisha Palaniappan is a student at Bergen County Technical High School in Teterboro. Her interests include data science research, particularly within the natural language processing sector of artificial intelligence.

Justin Sousa is a high school student at Bergen County Technical High School - Teterboro in the Strategic Asset Management program. He is studying topics such as economics and equity and derivatives trading, and he is working to become Bloomberg Certified. He is interested in economics, education policy, mathematics, and financial analysis.

Breanna Villarreal is an eleventh-grade AP student who is passionate about both investigating and uncovering knowledge regarding the Black Latina community. When Breanna is not spending time completing her research, she can be found on the stage, or participating in service events for her community.

Jessica Yu is a grade 12 student in Toronto, Canada. Most of her work centres around economics and equality, and she has participated in a behavioural economics project at Binghamton University as a summer researcher. Jessica hopes to conduct more research into behavioural economics in the future.

Pool of Consulting Editors of the Journal

Suzanne Conklin
Akbari

BA, MA, MPhil, PhD

Suzanne Conklin Akbari is professor of English and Medieval Studies at the University of Toronto, and was educated at Johns Hopkins and Columbia. Her research focuses on the intersection of English and Comparative Literature with intellectual history and philosophy, ranging from neo-platonism and science in the twelfth century to national identity and religious conflict in the fifteenth century. Akbari's books are on optics and allegory (*Seeing Through the Veil*), European views of Islam and the Orient (*Idols in the East*), and travel literature (*Marco Polo*); she is currently at work on *Small Change: Metaphor and Metamorphosis in Chaucer and Christine de Pizan*. She is volume editor for the *Norton Anthology of World Literature* (Volume B: 100-1500), co-editor of the *Norton Anthology of Western Literature*, and editor of *The Oxford Handbook to Chaucer*. She has begun a new research project called *The Shape of Time*, contrasting the temporal breaks found in medieval chronicle traditions with poetic narrations of the historical past. Akbari is cross-appointed to the following units at the University of Toronto: Centre for Medieval Studies; Centre for Comparative Literature; Centre for Jewish Studies; Department of Near and Middle Eastern Civilizations; Centre for Reformation and Renaissance Studies.

Becky Bakhshaei

BSc, MSc, Dip. Chem
Lab Technology

Becky Bakhshaei is a professional R & D chemist. She earned a master's degree in organic chemistry, an undergraduate degree in biochemical and pharmaceutical chemistry, and a diploma in chemical laboratory technology.

David Baldesarra

B.Sc., M.Eng., P.Eng.

David Baldesarra is a professional engineer with experience in structural failure cause analysis; assessment of property damage related to fire, explosion, vibration, ground movement, wind load, snow load, vehicle impact, fallen tree impact, and water (moisture) infiltration; analysis of building science and building envelope deficiencies/failures; structural design for the new construction and retrofit of buildings and miscellaneous, and structures, including design of structural steel, reinforced concrete and engineered wood, among other areas of expertise.

Barrie Bennett

BPE, MEd, PhD

Barrie Bennett is professor emeritus at the Ontario Institute for Studies in Education at the University of Toronto (OISE/UT). His research work focuses primarily on the design of powerful learning environments for students and teachers through the process of systemic change. He is currently working in districts in three countries on long-term projects related to instructional intelligence and systemic change (Australia, Ireland and Canada). Instructional intelligence involves intersecting the current research on curriculum, assessment, and instruction guided by what is known about how students and teachers learn. That intersection being driven by what is known about change and systemic change. He also assists teachers, schools, and districts with issues related to classroom management and school wide-discipline. Barrie has taught at the elementary and secondary levels, as well as, having worked in group homes, prisons, and security units for juvenile offenders. He has written six books: *Cooperative Learning: Where Heart Meet Mind*; *Classroom Management: A Thinking and Caring Approach*; *Beyond Monet: The Artful Science of Instructional Integration*; *Graphic Intelligence: Playing With Possibilities* and most recently *Power Plays*. Currently he is just finishing a text titled, *Effective Group Work: Beyond Cooperative Learning*.

Anthony Campbell
BA, MA, PhD

Anthony Campbell established Grow for Good Urban Teaching Farm in 2013 as a business model innovation laboratory and learning centre for young entrepreneurs. He spent time working throughout North America, Europe, Australia, Asia and now resides in his hometown of Toronto. Examples of Anthony's work are documented in *The Innovator's Field Guide* (2014), co-authored by David Crosswhite and Peter Skarzynski, as well as multiple Harvard Business School and Corporate Executive Board case studies chronicling the innovation and capability-building efforts of companies such as Samsung, Whirlpool, Best Buy and McDonald's. Previously, Anthony taught Film Studies, Writing and English Literature at The University of Western Ontario.

Jeremy B. Caplan
ScB, PhD

Jeremy Caplan is an Associate Professor in the Psychology Department at the University of Alberta, where he is also the Principal Investigator at the University of Alberta Computational Memory Lab. The lab is focused on human verbal memory behaviour and its basis in cognitive and neural processes. The team takes several approaches towards research, including mathematical modeling, measures of behaviour in the cognitive psychology tradition, and measures of brain activity using electroencephalography (event-related potentials and oscillations) and functional magnetic resonance imaging. He has been a referee for 38 academic journals.

Nathan Chow
BSc, Msc, BEd

Nathan Chow teaches Physical Sciences at Royal St. George's College (RSGC) in Toronto and consults on educational material at Perimeter Institute for Theoretical Physics in Waterloo. He has worked with Ontario's Ministry of Education to develop research-based teaching resources for Physics teachers within the province and across the rest of Canada. His academic research explored whether or not we understand how gravity behaves at intergalactic distances and provided possible explanations for our observations of dark energy. He has presented this research and led teacher training workshops around the world. In addition to AP Physics 1 and C, he teaches the AP Capstone Research course at RSGC.

Hance Clarke
BSc, MSc, MD, PhD,
FRCPC

Dr. Clarke is the director of Pain Services and the medical director of the Pain Research Unit at the Toronto General Hospital. He is appointed to the Institute of Medical Sciences at the University of Toronto and is a graduate of the Royal College Clinician Scientist Program. His research interests include identifying novel acute pain treatments following major surgery, identifying the factors involved in the transition of acute postsurgical pain to chronic pain, studying the genetics of acute and chronic pain after surgery, and identifying risk factors associated with continued opioid use and poor health related quality of life after major surgery as well as the efficacy of hyperbaric medicine. Over the past five years he has authored 47 peer reviewed manuscripts.

- Will Fripp
BA, MA
- Will Fripp is a public affairs and political risk analyst for Canadian and international clients. A B.A. in History and Political Science from Victoria University at the University of Toronto and an M.A. in Intelligence and International Relations from the University of Salford in Manchester, England, he is a historian specializing in intelligence and espionage, and its modern influences. Will anchored www.spiesintheshadows.com, a web based curriculum outlining Canadian foreign intelligence history and its impacts on Canada's national development. An occasional lecturer, Will's writings and review articles appear in peer-reviewed academic journals like *Intelligence and National Security*, and elsewhere.
- Michael Gemar
BSc, BA, PhD
- Michael Gemar received undergraduate degrees in Psychology and Philosophy from Rice University, and a PhD in experimental psychology from the University of Toronto. He has worked as a researcher at the Centre for Addiction and Mental Health, examining the cognitive and neural correlates of mood disorders, and was involved in a landmark study demonstrating the efficacy of mindfulness meditation to prevent depressive relapse. He has co-authored numerous journal articles, and taught for over a decade at U of T. More recently, he has worked in the area of health policy, and is currently at a Canadian non-profit.
- Jennifer Goldberg
BA, BEd, MA
- Jennifer Goldberg holds an M.A. in History from the University of Toronto. Her graduate studies focused on teacher misconduct in 19th century Ontario, and her research is published in *Historical Studies in Education*. She currently leads the English department and teaches at Havergal College, where she has also served as Chair of Teaching and Learning. In this capacity, she has explored the role of feedback in student learning, and has presented on this work at the National Coalition of Girls' Schools and Conference of Independent Teachers of English.
- Margaret S. Herridge
BSc, MSc, MPH, MD,
FRCPC
- Margaret Herridge is a Professor of Medicine and Senior Scientist at the University of Toronto. She is also a senior clinician in Critical Care and Respiratory medicine at University Health Network. Her research focus is on long-term outcomes after critical illness for patients and families and specifically on functional, neuropsychological, healthcare utilization and quality of life metrics. Her graduate studies were in Cell and Molecular Biology at Queen's University where she subsequently obtained her degree in Medicine. After completing her clinical training in Internal Medicine/Respirology and Critical Care at the University of Toronto, she obtained her Master of Public Health in Epidemiology and Statistics from the Harvard School of Public Health.
- Ted Higginbotham
BSc, MSc (Candidate)
- Ted Higginbotham is a graduate student at The University of Toronto and Hospital for Sick Children. His research is focused on further delineating the role of genomic structural variation in autism spectrum disorder and human disease. Ted is a contributing member of the Clinical Genome Resource (ClinGen), an international consortium working to define the clinical relevance of genes for use in precision medicine and translational research.
- Tim Hutton
BA, MLIS
- Tim Hutton is a teacher-librarian at Royal St. George's College. He has a BA in History and American Studies from the University of Toronto and a Masters in Library and Information Science from San Jose State University. At the secondary level, he has taught courses in the social sciences, humanities and communications technology, including a locally designed interdisciplinary course in urban studies.

Ira Jacobs

Dip Phys Ed, MHK,
DrMedSc

Professor Ira Jacobs became dean of the Faculty of Kinesiology & Physical Education at the University of Toronto on July 1, 2010, and was re-appointed to his current second decanal term. Before assuming this role, Jacobs was chair of York University's School of Kinesiology and Health Science from 2007 until 2010, and a federal government scientist from 1982 until 2007.

Jacobs earned his doctorate in clinical physiology from Sweden's Karolinska Institute, where he specialized in skeletal muscle metabolism. For the next 25 years, he did extensive exercise physiology research in Canada's human sciences laboratory, operated by the Department of National Defence. There, Jacobs rose to the position of chief scientist and led a unique international research group that helped to enhance the performance of military special operations units through their research into physiological, nutritional and pharmacological strategies.

He is a past president of the Canadian Society for Exercise Physiology and the Canadian Council of University Physical Education and Kinesiology Administrators. He is a fellow of the American College of Sports Medicine, an international fellow of the US National Academy of Kinesiology, and in 2016, he was named a Fellow of the Canadian Academy of Health Sciences.

Jacobs' research has led to the publication of more than 200 scientific articles, reports and book chapters about his research interests that include the physiological responses to physical exertion in environmental extremes, performance enhancement through pharmacological and nutritional manipulation of metabolism, and exercise pharmacology.

During his term as dean, the Faculty of Kinesiology & Physical Education has been rated as among the top academic programs in the world for kinesiology, physical education, sport and exercise sciences.

John Lambersky

BA, MA, BEd, PhD

John Lambersky is a teacher and head of the Canadian and World Studies department at Royal St. George's College in Toronto, where he leads the AP Capstone program. He has presented his work on teaching practice at the conferences of the International Boys' School Coalition, the National Association of Independent Schools, and the Canadian Accredited Independent Schools. His academic research is focused on school culture as a mechanism for school improvement. His work has been featured in *Leadership and Policy in Schools*, *The Dalhousie Review*, and *The Nashawaak Review*.

Blake Lee-Whiting

BA, MPP

Blake Lee-Whiting is a third year PhD student in the Department of Political Science at the University of Toronto. He received his BA from Queen's University and his MPP from the University of Toronto. He is interested in Canadian politics, public policy, and electoral politics. He is a member of the Policy, Elections, & Representation Lab at the Munk School of Global Affairs & Public Policy where he is currently working on projects related to the health of politicians, electoral success, and electoral candidacy.

Lori Loeb

BA, MA, PhD

Lori Loeb is Associate Professor of Modern British history at the University of Toronto. She has a Masters in Museum Studies and a PhD in History. A specialist in the Victorian period, she is the author of *Consuming Angels: Advertising and Victorian Women*. Generally, she writes about things in nineteenth-century Britain. A past Deputy Chair and Associate Chair (Graduate) of the History Department, she is currently MA Coordinator. She teaches courses in nineteenth and twentieth-century British history, Victorian material culture and the English country house.

- Gaven MacDonald
BSc, BEd
- Gaven MacDonald is a Physics and Mathematics teacher at Havergal College, where he is the faculty advisor for the Robotics Team. He is a member of the school's Blended Learning Team, which focuses on developing methods to combine online education resources with in-person classroom teaching. Gaven has designed physics simulations on the website www.cutequbit.com, that teachers can use to assist with their teaching, or to make individualized student assessments. Gaven also runs an educational YouTube channel which focuses on electronics and programming.
- Jaime Malic
BA (Hons), MA, BEd,
PhD
- Jaime Malic completed her PhD in Educational Leadership and Policy at the Ontario Institute for Studies in Education at the University of Toronto; her research focused on leadership values and practices in independent schools in Ontario. Jaime has fifteen years of experience as an educator in both independent and public schools. She currently teaches both AP Capstone Seminar and AP Research, as well as senior English courses at St. Clement's School. Jaime has served as a Reader for AP Capstone Seminar, written for *Independent Teacher* and *Independent Ideas*, and presented on various topics at the American Educational Research Association's Annual Conference, the Conference of Independent Teachers of English Annual Conference, the Ontario Advanced Placement Administration Conference, and the Advanced Placement Annual Conference.
- William J. McCausland
BASC, MEng, MA, PhD
- William McCausland is an associate professor of economics at the Université de Montréal. His research applies Bayesian statistical methods in two main areas. The first is discrete choice, at the interface of economics and psychology, where researchers study how people make choices from a small menu of available options. The second is time series modelling in economics, which has many applications in macroeconomics and financial economics. His undergraduate studies were in Engineering and he received his Ph.D. degree in economics from the University of Minnesota.
- Matt Mooney
BA Hons., BEd
- Matt Mooney is currently a secondary teacher in the Canadian & World Studies department at Royal St. George's College in Toronto, where he also serves on the Excellence in Teaching and Learning Committee. Matt earned an Honours BA from The University of Toronto, with a double major in History and Geography, and his Bachelor of Education from the Ontario Institute for Studies in Education. He has been teaching in Ontario since 2011 and has experience with curriculum development, such as his work on the Education Committee for Magna Carta Canada. Since 2019, Matt has helped to oversee *The Young Researcher*.
- Gurbir Perhar
BSc, PhD
- Gurbir Perhar received his academic training at the University of Toronto. His Doctoral work focused on mathematically modelling the transmission of highly unsaturated fatty acids in aquatic food chains, culminating in a suite of cybernetic models. His post-doc work saw him consulting for the governments of Canada, USA, South Africa, and China. These days Gurbir runs the Data Intelligence strategy in a multi-billion dollar corporation.

- Kate Schumaker
MSW, PhD
- Kate Schumaker is the Manager of Quality Assurance & Outcome Measurement at the Catholic Children's Aid Society of Toronto, and holds the position of Assistant Professor (status only) at the Factor Inwentash Faculty of Social Work, University of Toronto. She has worked for over 20 years in child welfare and children's mental health, including front-line clinical positions and 10 years producing and implementing child welfare policy for the provincial government. In 2011-12 she worked for the Commission to Promote Sustainable Child Welfare, supporting accountability framework development, including the establishment of a set of standardized performance indicators for the child welfare sector in Ontario. Her areas of practice and research interest include poverty, child neglect, trauma-informed practice, child welfare decision-making, and evidence-informed policy and practice.
- Eva Serhal
BA, MBA, PhD
- Eva Serhal is the Director of Virtual Mental Health and Outreach at the Centre for Addiction and Mental Health in Ontario, Canada and Director of the ECHO Ontario Superhub, a collaboration between CAMH and UHN that provides training and implementation support to new ECHO telementoring projects throughout Canada. Eva completed a PhD in Health Services Research at the University of Toronto, with a focus on outcomes and evaluation in virtual models of healthcare. Eva's current research assesses the implementation, adoption and economic factors of virtual care in Ontario. Eva also has significant experience with leadership and governance; she currently co-chairs the Toronto Telemedicine Collaborative and sits as a board member of the Children's Aid Society of Toronto.
- Michael Simmonds
BPE, M.A., Dip.Ed.,
MEd, EdD
- Michael Simmonds has worked in independent schools for over two decades. He taught science, biology, chemistry, physics, and math before becoming an administrator and Head of School. He earned graduate degrees from both McGill and Columbia universities respectively before receiving his doctorate from the University of British Columbia in Educational Policy & Leadership. His work on accountability synopticism is published in the peer-reviewed, *The International Education Journal: Comparative Perspectives*. He currently works at Havergal College as the VP School Life, Operations & Student Wellness.
- Sydney Stoyan
B.A, M.A., Ph.D.
- Sydney Stoyan holds a B.A. in French Literature from the University of Toronto, and an M.A. and a Ph.D in English Literature from the University of Ottawa. Her doctoral thesis, "The Widow's Might: Law and the Widow in British Fiction, 1689-1792," won the Governor General's Gold Medal for the Arts in 2002. She has since written freelance and worked as an editor for various publications and projects.

Alumni Editor

Andrew Pyper is an analyst at Charles River Associates in Washington DC, working in the Antitrust & Competition Economics Practice; in this role, he produces economic analysis for clients with antitrust-related litigation and regulatory issues. He graduated from the University of Chicago in 2022, where he majored in economics and political science, and graduated from Royal St. George's College in 2018, where he completed the AP Capstone program. His AP Research paper, published in *The Young Researcher*, examined RSGC students' perceptions of the school's implementation of formative assessment. In university, he continued his education work by advising local high school students on the university application process and continued to engage with academia as a data research assistant for a suicide attack research project and as an intern for the Milken Institute, supporting research on building more sustainable capital markets in developing countries.

Guidelines for Contributors

The Young Researcher is a peer-reviewed journal dedicated to publishing the best original research from secondary school students.

The journal's mission is to provide a larger audience for the original academic research of ambitious secondary students, provide a forum for peer-review, and create a community of young researchers. In addition, the journal strives to advance the quality of academic writing in secondary schools.

The Young Researcher is edited by secondary school students working closely with scholars and active researchers at universities and in the community. The journal operates a blind peer-reviewed review process, following those found in academic research journals.

The journal encourages submissions of original research (including relevant replication studies) from a wide range of academic disciplines within the social sciences, humanities, and sciences.

Submission Guidance:

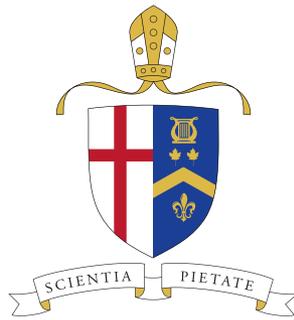
- No more than 5,000 words, excluding references and appendices (in English)
- Articles should have the following sections or equivalent:
 - Introduction
 - Literature Review
 - Method, Process, or Approach
 - Findings or Results
 - Discussion, Analysis, and/or Evaluation
 - Conclusion and Future Directions
 - References
- Papers should be formatted using discipline-appropriate methods (MLA, APA, and Chicago are acceptable).
- Papers should have an abstract (no more than 150 words) and have 4-6 keywords
- All units of measurement should be in metric wherever possible
- All studies involving human participants must have been approved by a Research Ethics Board

This is an open-access journal. By submitting your article, you agree to have it published in the journal. You always retain copyright as the author.

Since the journal publishes original research, by submitting your work you affirm that it is yours, that it is free from issues of academic dishonesty, and that it has not been submitted elsewhere.

To read more about TYR, about our editorial process, or to submit an article, please visit our website: www.TheYoungResearcher.com.

Please direct any questions to TheEditors@TheYoungResearcher.com



RSGC

Royal St. George's College

