A Data-Driven Analysis of Pandemic Educational Impacts and the Integration of Large Language Models in Global Learning Environments

Puti Xu

Thesis submitted for the fulfilment of the requirements for the degree of Doctor of Philosophy

Institute for Sustainable and Liveable Cities
Victoria University, Australia

June 2025

Abstract

Both COVID-19 and the rise of large language models (LLMs) have significantly influenced global education systems. The COVID-19 pandemic, along with its associated policies, profoundly shaped educational practices. The rapid surge in COVID-19 cases and deaths forced numerous countries to implement lockdown measures, resulting in the transformation of traditional face-to-face learning modes into remote or online learning systems. This shift has substantially impacted various dimensions of education, including academic performance and attendance rates across multiple demographic groups worldwide. In parallel with these policy-driven changes, LLMs have been rapidly developed and adopted across multiple disciplines. These models are now widely utilized as examiners and tutors, influencing both students and educators, and presenting both positive opportunities and notable challenges.

Although previous studies have examined the pandemic's impact on academic performance and attendance rates, comprehensive comparative analyses that address variations across multiple countries, different phases of the pandemic, and diverse ethnic groups remain scarce. Furthermore, most research on LLMs has been limited to a single disciplinary context, primarily within the medical field. Addressing these research gaps, this thesis investigates the educational impacts of the COVID-19 pandemic—including its effects on global academic performance and attendance rates among various ethnic groups—while also examining the multifaceted roles of LLMs in education from the perspectives of both students and educators.

The first part of the thesis compares academic performance across different phases of the COVID-19 pandemic—pre-pandemic, during the pandemic, and

post-pandemic—using standardized tests such as TOEFL and GMAT. The subsequent chapters analyze the effects of COVID-19 case numbers and vaccination rates on childcare and school attendance among multiple ethnic groups in New Zealand, including Māori, European/Pākehā, Asian, and Pacific populations. The third research focus explores the diverse roles LLMs play in educational development across various disciplines, such as medicine, programming, engineering, and language studies, where LLMs function as both examiners and tutors. This section also systematically assesses the pandemic's influence on education from both student and educator perspectives.

The thesis contributes to existing knowledge in several aspects. First, it highlights both the differences and similarities in standardized test performance across countries, demonstrating that the most significant changes occurred at the onset of the pandemic, and that the patterns of change varied depending on test difficulty, home-edition test requirements, and participant characteristics on a global scale. Additionally, through Spearman correlation and Mean Absolute Percentage Error(MAPE) results derived from multiple machine learning regression analyses, the findings illustrate that while the pandemic greatly influenced attendance rates, the years with the most significant impacts varied. Moreover, the ethnic groups most affected differed between childcare and school attendance rates. Regarding large language models, existing studies reveal that the latest versions of GPT have achieved substantial breakthroughs in examination accuracy across various disciplines; however, these accuracy rates are inconsistent across fields. Both students and educators also report concerns regarding overreliance on these models, as well as ethical issues such as cheating and plagiarism.

Declaration of Authenticity

"I, Puti Xu, declare that the PhD thesis entitled [A Data-Driven Analysis of Pandemic Educational Impacts and the Integration of Large Language Models in Global Learning Environments] is no more than 80,000 words in length, including quotes and exclusive of tables, figures, appendices, bibliography, references, and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work."

"I have conducted my research in alignment with the Australian Code for the Responsible Conduct of Research and Victoria University's Higher Degree by Research [redacted]

Signature [redacted] Date: 11/10/2025

Acknowledgements

I would like to express my sincere appreciation to all those who have provided guidance, support, and assistance throughout the course of this doctoral research.

First and foremost, I extend my profound thanks to my principal supervisor, Professor Hua Wang, for his steadfast guidance, insightful critique, and continual encouragement throughout this study. Professor Wang offered not only valuable perspectives on the logical structure and coherence of the thesis but also practical advice on enhancing academic writing, improving clarity, and identifying appropriate reference materials and target journals. His comprehensive supervision regarding ethical approvals and thesis requirements, consistently delivered during our weekly meetings, has played a pivotal role in shaping the quality and integrity of this work.

I am also especially appreciative of the academic mentorship and dedication of my associate supervisor, Dr. Jiao Yin. From the inception of this research, Dr. Jiao Yin provided discerning suggestions regarding the conceptual framework and overall organization of the thesis, which were instrumental in refining the research direction and developing a systematic work plan. Our regular discussions served as a structured forum for progress evaluation, detailed feedback, and precise goal-setting at each stage of this research. Her expertise in data selection, methodological considerations, and analytical approaches significantly enhanced the depth and rigor of the study. Moreover, her thoughtful oversight of chapter development and timely provision of suggestions and reminders—even during periods of absence from meetings—ensured that the research advanced in an organized and effective manner.

I am sincerely thankful to Mr.Wei Hong for his technical expertise and methodological advice. His guidance on database selection and analytical techniques—widely applied in both information technology and social science research—contributed greatly to the robustness of the study's methodological design. His assistance with coding, as well as the design and visualization of figures and charts, markedly improved the clarity and communicative power of the thesis. His valuable input on enhancing the readability and coherence of the written work is also greatly appreciated.

I would further like to acknowledge the panel members who provided critical evaluations and insightful recommendations during the confirmation of candidature and thesis submission processes. Their thoughtful feedback identified important areas for refinement, thereby contributing to the improved clarity, consistency, and academic merit of this thesis.

Lastly, I am deeply appreciative of the enduring support and encouragement of my family. In particular, my parents have played a vital role in sustaining my motivation and focus, offering both practical assistance in keeping track of key deadlines and moral support that has underpinned the successful completion of this research endeavor.

Declarations of Authenticity and Authorship Contribution in the Thesis with Publication Format

Published papers:

- P. Xu, W. Hong, J. Yin, K. Wang, and Y. Zhang, "Analyzing the Impact of COVID-19 on Education: A Comparative Study Based on TOEFL Test Results," in *Proceedings of the International Conference on Health Information Science*, 2023, pp. 175–184.
- P. Xu, J. Yin, J. Cao, and Y. Zhang, "The Impact of COVID-19 on Global GMAT Test-Taking Patterns: A Statistical Analysis," in *Proceedings of the International Conference on Health Information Science*, 2025, pp. 303–312.
- P. Xu, W. Hong, J. Yin, H. Wang, and Y. Zhang, "Assessing the impact of COVID-19 on early learning services: A study based on attendance rate in New Zealand," *New Zealand Journal of Educational Studies*, pp. 1–22, 2025.

Under consideration:

- P. Xu, J. Yin, H. Wang, J. Cao, and Y. Zhang, "Disparities in School Attendance During COVID-19: The Impact of Pandemic Factors on Ethnic and Māori Groups in New Zealand," preprint, Apr. 30, 2025. [Online]. Available: https://osf.io/preprints/edarxiv/jhn47_v1

- P. Xu, J. Yin, H. Wang, Y. Miao, and Y. Zhang, "A Systematic Review of Large Language Models in Education: Dual Roles, User Perspectives, and Cross-Disciplinary Impacts," preprint, Apr. 30, 2025. [Online]. Available: https://doi.org/10.35542/osf.io/fk5wa_v1

Contents

Abstract **i**

Declaration of Authenticity **iii**

Acknowledgements **iv**

Declarations of Authenticity and Authorship Contribution in the Thesis with Publication Format **vi**

List of Figures **xiv**

List of Tables **xvi**

1 Introduction **1**

1.1 The COVID-19 Pandemic and Its Impact on Education 1

1.1.1 Overview of the COVID-19 Pandemic 1

1.1.2 COVID-19 and Education Changes 2

1.2 Large Language Models(LLMs)-Foundations and Educational Relevance 4

1.2.1 Evolution of LLMs 4

1.2.2 Global Landscape of LLMs Applications 4

1.2.3 Educational Potentials and Challenges of LLMs 5

1.3 Research Objectives 6

1.4 Research Significance 7

1.5 Thesis Structure 8

2 Literature Review **9**
2.1 Impacts of COVID-19 on Education 9
2.2 LLMs in Educational Contexts 10

3 Preliminary **12**
3.1 Data Comparison Method 12
3.1.1 Parametric Comparsion 13
3.1.1.1 One-Sample t-Test 13
3.1.1.2 Two-Samples t-Test 14
3.1.1.3 More Than Three Variables 16
3.1.2 Non-parametric Comparison 17
3.2 Correlation Analysis 18
3.2.0.1 Pearson Correlation Analysis 18
3.2.0.2 Partial Correlation Analysis 19
3.2.0.3 Spearman Correlation Analysis 19
3.3 Regression Models 20
3.3.1 Linear Regression 20
3.3.2 Support Vector Machine 21
3.3.3 K-Nearest Neighbours 22
3.3.4 Guassian Process Regression 23
3.3.5 Partial Least Squares 24
3.3.6 Multi-Layer Perceptron 24
3.3.7 Decision Tree Regression 25
3.4 Evaluation Metrics 26
3.5 Databases 28
3.5.1 COVID-19 Databases 28
3.5.2 Academic Performance Metrics 29
3.5.2.1 Test Results Databases 29
3.5.2.2 Attendance Rate Databases 31

4 Analyzing the Impact of COVID-19 on Education: A Comparative Study Based on TOEFL Test Results 32
4.1 Introduction 32
4.2 Literature Review 33
4.3 Methodology 34
4.3.1 Data Collection and Pre-processing 34
4.3.2 Research Design 35
4.4 Results 36
4.4.0.1 Total Scores in the World and Different Regions . 36
4.4.0.2 Section Scores Worldwide 38
4.4.0.3 Performance Breakdown for Sections and Regions 40
4.5 Discussion 41
4.6 Implications 42
4.7 Conclusion 42

5 The Impact of COVID-19 on Global GMAT Test-Taking Patterns: A Statistical Analysis 44
5.1 Introduction 44
5.2 Literature Review 45
5.3 Research Design 46
5.3.1 Materials 47
5.3.2 Data Analysis Methods 48
5.3.2.1 Wilcoxon Signed-Rank Test 48
5.3.2.2 Spearman Correlation Coefficient 49
5.4 Experiment Results 50
5.4.1 Data Analysis and Reprocessing 50
5.4.2 COVID-19 Impact on the Number of GMAT Examinees . 51
5.4.2.1 Significance Analysis of Changes in the Number of Examinees 51
5.4.2.2 Correlation Analysis for the Impacts of COVID-19 on Number of Examinees 54
5.4.3 COVID Impact on Mean Total Scores (MTS) 56
5.4.3.1 Significance Analysis of Changes in MTS 56

5.4.3.2 Correlation Analysis for the COVID-19 Pandemic and Changes in MTS 57
5.4.4 Discussion 58
5.5 Conclusion 60

6 Assessing the Impacts of COVID-19 on Early Learning Services: A Study Based on Attendance Rate in New Zealand 61
6.1 Introduction 61
6.2 Literature Review 64
6.3 Materials 65
6.3.1 Data Collection 65
6.3.2 Data Processing and Visualization 67
6.3.2.1 New Cases and Attendance Rates in New Zealand 67
6.3.2.2 New Vaccinations and Attendance Rates in New Zealand 70
6.4 Methodology 71
6.4.1 Data Analysis Method 71
6.4.1.1 Spearman Correlation Coefficient 72
6.4.1.2 Models for Regression Analysis 72
6.4.1.3 Evaluation Metrics 73
6.5 Experimental Results 74
6.5.1 The Impact of the COVID-19 Cases 74
6.5.1.1 The Correlation Between New Cases and Attendance Rates 74
6.5.1.2 Regression Analysis of Case Effects on Attendance Rates 77
6.5.2 The Impact of COVID-19 Vaccinations 79
6.5.2.1 Correlation Between New Vaccinations and Attendance Rates 79
6.5.2.2 Regression Analysis of Vaccination Effects on Attendance Rates 79
6.5.3 Discussion 82
6.6 Conclusion 83

7 Disparities in School Attendance During COVID-19: The Impact of Pandemic Factors on Ethnic and Māori Groups in New Zealand 85
7.1 Introduction 85
7.2 Literature Review 87
7.3 Materials 88
7.3.1 Data Collection 88
7.3.2 Preliminary Data Visualization and Research Design . . . 90
7.3.2.1 New Cases and Attendance Rates by Ethnicity . 90
7.3.2.2 New Vaccinations and Attendance Rates by Ethnicity 92
7.3.2.3 New Cases and Attendance Rates in Māori Schools 94
7.3.2.4 New Vaccinations and Attendance Rates in Māori Schools 95
7.4 Methodology 98
7.4.1 Correlation Analysis 98
7.4.2 Regression Analysis 98
7.5 Experimental Results 101
7.5.1 Variable Selection 101
7.5.2 Effects of the COVID-19 Pandemic on Ethnicity Attendance Rates 103
7.5.2.1 Impacts of COVID-19 Cases on Ethnicity Attendance Rates 103
7.5.2.2 Impacts of COVID-19 Vaccinations on Ethnicity Attendance Rates 104
7.5.3 Effects of the COVID-19 Pandemic on Attendance Rates in Māori Schools 107
7.5.3.1 Impacts of COVID-19 Cases on Māori School Attendance Rates 107
7.5.3.2 Impacts of COVID-19 Vaccinations on Māori School Attendance Rates 109
7.5.4 Discussion 110
7.6 Conclusion 113

8 Synthesizing the Educational Impact of COVID-19 115
8.1 The Impact of COVID-19 Pandemic on Academic Performance . . 115
8.2 The Influence of COVID-19 on Attendacne Rates 117
8.3 Implications and Future Directions 119

9 A Systematic Review of LLMs in Education: Dual Roles, User Perspectives, and Cross-Disciplinary Impacts 120
9.1 Introduction 120
9.2 Literature Selection and Screening 122
9.2.1 Search Strategy 122
9.2.2 Statistics of Selected Papers 123
9.3 Functional Applications of LLMs in Education 127
9.3.1 LLMs as Learners: Evaluating Academic Performance . . . 127
9.3.1.1 Performance of LLMs in Medical Education . . . 127
9.3.1.2 Performance of LLMs in Other Academic Domains 133
9.4 LLMs as Instructors: Supporting Teaching and Assessment 135
9.5 Perceived Educational Impacts of LLMs 140
9.5.1 Impacts from the Student Perspective 140
9.5.1.1 Perceived Benefits for Students 140
9.5.1.2 Perceived Drawbacks for Students 142
9.5.2 Impacts from the Instructor Perspective 143
9.5.2.1 Perceived Benefits for Instructors 143
9.5.2.2 Perceived Drawbacks for Instructors 146
9.6 Discussion 148
9.7 Conclusion 150

10 Conclusion and Prospect 151

List of References 153

List of Figures

4.1 Mean Difference of Total Scores for Each Group across Regions . 37
4.2 Mean Difference for Each Group Globally (Section Grade Based) . 39

5.1 Z-Values of GMAT Examinee Numbers 54
5.2 Z-Values of MTS 57

6.1 New Cases and Attendance Rates in 2021 68
6.2 New Cases and Attendance Rates in 2022 69
6.3 New Vaccinations and Attendance Rates in 2021 70
6.4 New Vaccinations and Attendance Rates in 2022 71
6.5 MAPEs of Attendance Rates Based on New Cases and Total Cases for 2021 and 2022 78
6.6 MAPEs of Attendance Rates Based on Vaccination Variables for 2021 and 2022 81

7.1 New Cases and School Attendance Rates by Ethnicity in 2021 . . 91
7.2 New Cases and School Attendance Rates by Ethnicity in 2022 . . 92
7.3 New Vaccinations and School Attendance Rates by Ethnicity in 2021 93
7.4 New Vaccinations and School Attendance Rates by Ethnicity in 2022 94
7.5 New Cases and School Attendance Rates across Different Māori Schools in 2021 95
7.6 New Cases and School Attendance Rates across Different Māori Schools in 2022 96
7.7 New Vaccinations and Attendance Rates across Different Māori Schools in 2021 97

7.8 New Vaccinations and Attendance Rates across Different Māori Schools in 2022 97
7.9 MAPEs of Ethnicity-Specific Attendance Rate Predictions Based on Case Variables for Different Regression Models in 2021 and 2022 105
7.10 MAPEs of Ethnicity-Specific Attendance Rates Predictions Based on Vaccination Variables for Different Regression Models in 2021 and 2022 107
7.11 MAPEs of Māori School Attendance Rate Predictions Based on Case Variables for Different Regression Models in 2021 and 2022 . 109
7.12 MAPEs of Māori School Attendance Rates Predictions Based on Vaccination variables for Different Regression Models in 2021 and 2022 111

9.1 Flow Diagram of the Article Selection Process for the Systematic Review 124
9.2 Year-Wise Distribution of Selected Articles 125
9.3 Geographic Distribution of Reviewed Articles by Country 125
9.4 Classification of Reviewed Articles by Type 126
9.5 Conceptual Structure of the Review: Functional Roles of LLMs and Their Perceived Educational Impacts 126

List of Tables

4.1 Group Mean Difference across Regions (Total Scores Based) . . . 37
4.2 Group Mean Difference Globally (Section Grade Based) 38
4.3 Group Mean Difference across Regions (Section Grade Based) . . 39

5.1 Z-Values of GMAT Examinee Numbers 53
5.2 Variable List: Correlation Analysis of Changes in Examinee Numbers 55
5.3 Correlation Analysis of COVID-19 Impact on Examinee Numbers 55
5.4 Changes in MTS 57
5.5 Variable List for Correlation Analysis of MTS Differences 58
5.6 Correlation Analysis of COVID-19 Impacts on MTS 58

6.1 Variable List for the Correlation and Regression Analysis 75
6.2 The Correlation Coefficient Between New Cases and Attendance Rates 76
6.3 MAPE of the Attendance Rates Based on Total Cases and New Cases 77
6.4 Correlation Coefficient Between New Vaccinations and Attendance Rates 80
6.5 MAPE Values of Attendance Rates Based on Vaccination Variables 80

7.1 Correlation Between New COVID-19 Cases and School Attendance Rates by Ethnicity for 2021 and 2022 103
7.2 MAPE of Ethnicity-Specific Attendance Rates Predictions Based on Case Variables for Different Regression Models in 2021 and 2022 104
7.3 Correlation Between New COVID-19 Vaccinations and School Attendance Rates by Ethnicity for 2021 and 2022 106

7.4 MAPEs of Ethnicity-Specific Attendance Rates Predictions Based on Vaccination Variables for Different Regression Models in 2021 and 2022 106
7.5 Correlation Between New Cases and Māori School Attendance Rates for 2021 and 2022 108
7.6 MAPEs of Māori School Attendance Rates Predictions Based on Case Variables for Different Regression Models in 2021 and 2022 108
7.7 Correlation Between New Vaccinations and Māori School Attendance Rates for 2021 and 2022 110
7.8 MAPEs of Māori School Attendance Rates Predictions Based on Vaccination Variables for Different Regression Models in 2021 and 2022 110

9.1 LLMs as Learners in Medicine 128
9.2 LLMs as Learners in Other Subjects 134
9.3 LLMs as Instructors 135
9.4 Positive Influence of LLMs on Academic Skills of Students 141
9.5 Negative Influence of LLMs on Academic Skills of Students 142
9.6 Positive Influence of LLMs from the Perspective of Instructors . . 143
9.7 Negative Influence of LLMs from the Perspective of Instructors . . 146

Chapter 1

Introduction

In modern society, many elements are challenging education systems. Two typical elements are the education policies and the corresponding techniques. Education policies are closely associated with pandemics in different countries, and a typical example is the COVID-19 pandemic, which has significantly altered education policies worldwide. For the techniques, a large number of researchers have begun to explore the roles of LLMs in education, greatly challenging the current development of education.

1.1 The COVID-19 Pandemic and Its Impact on Education

1.1.1 Overview of the COVID-19 Pandemic

COVID-19 has had a profound impact on various aspects of the world since its emergence in 2020. Initially, the virus was identified in 2019, and nearly all countries were affected by the pandemic in 2020 [1]. In response to the global outbreak, it was officially declared a pandemic in the first quarter of 2020 [2]. Multiple countries, including China, Japan, Australia, Germany, and France, were severely impacted by the pandemic [3]. In May 2023, the World Health Organization (WHO) announced the official end of the global pandemic phase of COVID-19 [4]. Therefore, the pandemic lasted for more than three years globally.

The long-term pandemic has resulted in extensive health-related consequences, including both physical and psychological issues. First, COVID-19 led to a high number of deaths and infections globally due to its high transmission rate. When the pandemic was officially declared in March 2020, the number of confirmed cases exceeded 120,000 [2], with similar mortality levels observed in many other countries. Moreover, a wide range of long-term symptoms has been identified among recovered patients. According to a medical research study, the COVID-19 symptoms include fatigue, cognitive impairment, and numerous other serious conditions [5], affecting a substantial proportion of individuals who contracted the virus.

In addition, the pandemic has also triggered widespread psychological distress among families, primarily due to lockdown policies. These policies often required students to remain at home during extended periods of school closure. Numerous studies have indicated that some students, particularly children, experienced feelings of isolation, which in turn contributed to elevated levels of anxiety and depression [6]. In several countries, suicide rates also increased during this time [6]. Furthermore, fear of the disease itself was intensified by the pandemic context [6].

1.1.2 COVID-19 and Education Changes

Due to the severity of the pandemic, numerous countries—including the U.S., Australia, Germany, France, the Netherlands, China, Thailand, Cuba, India, and Morocco implemented lockdown policies [7]. These lockdown policies aimed to restrict human mobility and reduce face-to-face interactions among residents [7]. Domestic lockdowns confined residents to their dwellings, while international lockdowns involved the closure of national borders, resulting in significant reductions in international flights and other cross-border public transportation services [7].

One of the most significant policies affecting education was that a large number of educational institutions transitioned from their traditional teaching mode, face-to-face instruction, to online learning. Before the COVID-19 pandemic, the techniques were adopted by some universities [8], but the COVID-19 pandemic strengthened the trends of online studies. During the pandemic, under the strict lockdown measures in multiple countries, nearly all educational institutions, including early childhood services, primary schools, secondary schools, universities, and various other institutions, were forced to close [9]. In response, a vast number of institutions replaced traditional in-person teaching with online learning. Students were required to complete most assessments, presentations, and examinations through online platforms [9]. Various digital tools and platforms, including Zoom and other technologies, were adopted by schools to adapt to the substantial shifts within the education sector [9].

Due to its great influence, many scholars in the education sector have conducted extensive research on its effects on different aspects of education. Some studies have found that academic performance in higher education was negatively affected by pandemic-induced anxiety and fear [10], while others have shown that students' time spent on academic activities significantly declined [11].

Despite these findings, most existing studies have focused on individual countries at a single point in time, lacking broader analysis from a global perspective and failing to consider differences among various ethnic groups. As a result, it is unclear whether the pandemic's impact on education is consistent across all ethnicities and countries worldwide or whether the pattern of impact varies across different stages of the COVID-19 pandemic. Therefore, the scope of the research must be expanded.

Therefore, these are the two research questions in the thesis aiming to address the issues: 1. How has the COVID-19 pandemic influenced academic performance across multiple countries at different stages of the outbreak?

2. How have school attendance patterns among various ethnic groups within a single country been impacted during the different phases of the COVID-19 pandemic?

1.2 Large Language Models(LLMs)-Foundations and Educational Relevance

1.2.1 Evolution of LLMs

As an important branch of artificial intelligence (AI), LLMs have significantly evolved over the past century [12]. These models are defined as deep learning systems designed for language processing tasks [13]. Specifically, many experts describe them as a type of AI system able to generate texts based on extensive training [12].

There are several distinct phases of the development of LLMs. Initially, they were introduced as tools for natural language processing (NLP) [12]. In the 1950s, the foundational concepts of NLP were proposed and began to be implemented [14]. However, by the 1970s, rule-based systems faced significant challenges in generating coherent text due to the complexity and variability of natural languages, as well as the rapid advancement of computing technologies and the internet worldwide [12, 14]. As a result, researchers developed statistical models that did not rely on predefined linguistic rules [12], marking a breakthrough in NLP [14]. In the 2010s, LLMs experienced substantial growth [12]. For example, the GPT series, first introduced in 2018, has undergone multiple iterations in less than a decade and has been widely adopted across many countries [12].

There are various types of LLMs. One of the most prominent is the GPT series. Since its initial release in 2018 [12], the model has seen rapid advancements. GPT-2 demonstrated the ability to perform diverse tasks without task-specific training, while GPT-3 expanded significantly in scale [12]. In addition to the GPT series, other notable LLMs include LLaMA, PaLM, and several others [15].

1.2.2 Global Landscape of LLMs Applications

In modern society, LLMs have been widely adopted across the globe.

LLMs have been implemented in numerous industries across various countries. For example, in China, LLMs have been utilized to analyze the distribution of specific types of crime [16, 17, 18, 19]. Additionally, in India and some other countries, they have been applied in the construction industry for tasks such as text and image generation, engineering design, and querying construction engineering datasets [20, 21, 22, 23].

Beyond China and India, they have been widely adopted in multiple sectors in the United States (U.S.). Many medical students in the U.S. have incorporated GPT series into their academic routines, adopting it for support in areas such as diagnosis, treatment planning, and clinical decision-making [24, 25, 26, 27]. Furthermore, LLMs have also been applied in policy development and reform in the United States [28], indicating a growing presence of such technologies in public administration and governance [29, 30, 31, 32].

In Australia, LLMs have similarly been applied across a range of domains. Some experts have leveraged LLMs for cybersecurity purposes [33, 34, 35, 36]. Additionally, GPT has begun to be implemented in teaching and learning practices at various Australian universities [37]. LLMs were adopted to assess different aspects of remote work during the lockdown period, including job advertisements, employment offerings, health-related tasks, and task coordination [38, 39, 40, 41].

1.2.3 Educational Potentials and Challenges of LLMs

Due to the widespread application of LLMs, it is evident that these models have been extensively adopted in the education sector.

Numerous LLMs have been implemented for various educational purposes. The GPT series, for instance, has been utilized in numerous examinations and assignments across disciplines such as medicine, programming, mathematics, and physics [42, 43, 44]. Additionally, they have also been employed for coding correction and grammar checking [45]. Additionally, other LLMs have also been integrated into educational settings [46]. These models support functions such as personalized learning, course preparation, prompt engineering, and various other applications.

Given the rapid expansion of LLMs across multiple industries and their relatively early stage of development, there is ongoing debate regarding their positive and negative impacts. Many studies highlight the beneficial effects of LLMs on academic activities. For example, some researchers find that LLMs provide more constructive feedback and task-specific suggestions compared to traditional methods [45], while others report that the efficiency of feedback delivery by LLMs surpasses that of human reviewers [47]. Conversely, some studies suggest potential drawbacks. Critics argue that reliance on LLMs may lead to reduced motivation for independent study [48], and concerns have been raised about increased incidences of academic dishonesty and ethical issues among students [48].

Despite the growing body of research, two major gaps remain in the existing literature. First, most studies focus on specific disciplines, primarily medicine [49], thereby limiting the understanding of LLMs' impact across the broader educational spectrum. Second, the majority of research examines either educators' or students' perspectives exclusively, lacking a comprehensive comparative analysis of attitudes toward LLMs from both groups [49, 50].

To address these gaps, the proposed research question is: "What roles do LLMs play in education and how do they influence students and educators?"

1.3 Research Objectives

Due to the existing gaps in understanding how the COVID-19 pandemic and LLMs impact education, this thesis aims to address these issues from multiple perspectives.

The research first addresses the gaps concerning the influence of the pandemic across different countries and phases. Specifically, statistical data from various countries worldwide is analyzed using consistent standards to determine whether the impacts vary between nations. Furthermore, the effects on different ethnic groups within multicultural countries are examined to explore whether there is a consistent pattern of the pandemic's influence on education across diverse ethnicities.

Regarding the gaps related to LLMs and education, several corresponding approaches will be undertaken. First, the roles of LLMs in different academic subjects are investigated. Additionally, the various functions of LLMs in educational contexts are explored. Finally, a comparative analysis is conducted to assess the influence of LLMs on education from both the students' and educators' perspectives.

1.4 Research Significance

The research on the influences of the pandemic and LLMs on education is very important for multiple reasons.

First, the thesis informs future research by providing insights into pandemics, highlighting commonalities and differences across various phases from multiple perspectives. For example, educators and the corresponding educational institutions need to establish clear policies and mechanisms for international standardized tests, based on the observed changes in academic performance across different phases of the COVID-19 pandemic, to mitigate the pandemic's impact and ensure fairness in multiple international assessments.

Additionally, it offers a comprehensive overview of the multiple roles of LLMs in education, shedding light on common applications and critical reflections regarding their adoption. When aligned with international framework standards, this research guides the responsible adoption of LLMs in educational institutions and related IT sectors.

Moreover, examining the impact on multiple ethnicities provides valuable guidance for policy development. Governments and policymakers require a thorough understanding of the similarities and differences in how various ethnic groups are affected, enabling them to formulate more effective and equitable policies in response to future pandemics.

Furthermore, the research offers recommendations for improving LLM-related technologies. By analyzing both the positive effects and negative effects of these models, the findings will support developers in advancing more effective and practical applications. The inclusion of perspectives from both students and educators will also help LLM developers address ethical concerns and implement measures to mitigate moral issues associated with their adoption.

1.5 Thesis Structure

There are several chapters in the thesis. Chapter 2 presents a literature review that systematically analyzes the existing gaps in current research studies, and Chapter 3 outlines the foundational knowledge relevant to the thesis. The literature review analyzes existing research findings and identifies corresponding gaps, while the preliminary section outlines various methodologies and criteria for selecting the database. Chapters 4 and 5 examine changes in TOEFL and GMAT test performances over multiple years. Chapters 6 and 7 explore how the COVID-19 pandemic impacts the school attendance rates in New Zealand. Chapter 8 explains several roles of multiple LLMs in education and the positive and negative influences of these models. Finally, the discussion will analyze the key findings of all chapters, while the conclusions will summarize the highlights of the thesis.

Chapter 2

Literature Review

2.1 Impacts of COVID-19 on Education

COVID-19 has significantly impacted student academic achievements across multiple countries and different phases of the pandemic.

First, the pandemic's effects on education have varied across countries and demographic groups. For example, research on Chinese primary and secondary students indicates gender differences in the pandemic's impact on mathematics performance [51]. Similarly, German students aged 18 to 24 experienced pronounced academic disruptions [52], while in South Africa, school attendance rates sharply declined following the outbreak [53].

Second, the COVID-19 pandemic has influenced the education systems and corresponding studies during the pandemic. Studies analyzing learning growth between 2020 and 2021 reveal negative effects on the K-12 education system during the pandemic [54]. Additionally, research emphasizes the need for targeted education policies to enhance the students' study motivations towards having face-to-face studies [10].

Furthermore, multiple dimensions of education have been affected. Numerous studies report declines in academic performance in multiple examinations [55, 56, 57], alongside reduced participation in examinations and other study activities [58, 59].

However, existing research has some limitations. Most studies focus on one or two years and typically examine only one or a few countries, limiting the generalizability of their conclusions. Moreover, analyses often concentrate on a single educational outcome, and the differential effects on diverse ethnic groups remain underexplored.

This thesis aims to address these gaps by broadening the scope to include multiple countries worldwide with available data. It also incorporates analyses of different ethnicities. Furthermore, the thesis covers a longer timeframe—pre-pandemic, pandemic, and post-pandemic periods—and evaluates various educational outcomes, including academic performance, examinee numbers, and attendance rates.

2.2 LLMs in Educational Contexts

Due to the widespread adoption of LLMs, numerous studies have investigated their applications in the education sector. The research is highly significant because education plays a crucial role in shaping individuals' lives. Its value is intrinsic and fundamental [60], which underscores the importance for educators to design clear schemes for both teaching content and instructional methods. Furthermore, educational institutions provide students with opportunities to develop their personal character [60]. Therefore, these institutions need to implement appropriate strategies to mitigate any potential negative influences of large language models [61].

Due to its significance of education, research has examined the role of LLMs across a variety of disciplines. Some studies focus on their use in medical education [62, 63], while others explore applications in programming [64] and engineering [65].

In addition to disciplinary coverage, existing research has explored various functions of LLMs in educational contexts. These include the roles of LLMs as participants in multiple examinations [66, 67], the functions of feedback generation across subjects, and code generation in computer science education [20, 68].

Despite these contributions, many existing studies are limited in scope, often focusing on a single discipline or specific application. In contrast, the perspectives are limited to students or educators, so there are no comprehensive overviews of the influence of LLMs on different disciplines from different perspectives. Therefore, this thesis seeks to comprehensively analyze LLMs in education, examining their impact across multiple subjects and diverse educational functions.

Chapter 3

Preliminary

3.1 Data Comparison Method

The thesis needs to compare the datasets across multiple years, so it is important to comprehensively analyze each data comparison method.

Generally, there are two types of comparison methods, including the parametric method and the non-parametric method [69]. Generally, in theory, the parametric test is applied only when three assumptions are met [69]. The datasets need to follow the normal distribution. In contrast, the homogeneity of variance in all datasets in the tests should be identical [69], which means that in the practical comparison tests or data analysis, there should be no significant differences in the standard deviations of each data group. The third assumption is that each variable should be independent [69].

However, in practice, it is very challenging to ensure that the distribution is normal while the data variance is identical. The sample size of datasets also influences the choices of the statistical analysis [70], because in practice, it is much more challenging to ensure all datasets are exactly normally distributed. The sample size of the one-sample t-test should be at least 20, while for the two-sample t-test, the sample size should be at least 15 [70].

3.1.1 Parametric Comparsion

3.1.1.1 One-Sample t-Test

The one-sample t-test is usually adopted to explore whether there are significant differences between the population mean and an assumed value [71]. In practice, the assumed value may be standardized, while in many cases, a hypothesized value may be proposed [71].

The tests include one-tailed and two-tailed t-tests [72].

The null hypothesis of both the one-tailed t-test and the two-tailed t-test is the same, which is [72]:

$$H_0 : \mu_1 = \mu_0$$

The alternative hypothesis for the two-tailed t-test is [72]:

$$H_1 : \mu_1 \neq \mu_0$$

The alternative hypothesis for the one-tailed t-test is [72]:

$$H_1 : \mu_1 < \mu_0$$

or

$$H_1 : \mu_1 > \mu_0$$

The formula of the t-value is (3.1) [73]:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \tag{3.1}$$

where $\bar{x}$ denotes the mean value in the sample, s represents the standard deviation, and n is the sample size.

Generally, the p-value is adopted to assess whether the researchers should reject the null hypothesis [74]. The p-value needs to be confirmed with the degrees of freedom and the t-value in certain forms. The degree of freedom is also calculated based on certain standards.

If the p-value is less than a certain percentage, then there is sufficient evidence to reject the null hypothesis that the two means are the same [75]. On the contrary, in theory, the researchers should not reject the null hypothesis because there is insufficient evidence [75]. However, even if the null hypothesis is rejected, it does not mean that the two average numbers are equal [75]. Generally, the p-value is compared with 5% in most research studies [76]. Although there are certain arguments on the adoption of p-values and corresponding levels, 5% is still adopted by most researchers. For example, in multiple medical research studies, the p-value is usually compared with 5% [76]. However, in some research studies on a higher accuracy level, it is necessary to adopt a stricter level. In a large number of precise research studies, including the neural network or certain clinical practice, a much lower p-value may be adopted [77]. Also, in some other studies, such as social sciences, 10% has been adopted for the comparison [78].

3.1.1.2 Two-Samples t-Test

Paired-Samples t-Test The paired-samples t-test is a statistical tool for the comparison of two sets of data to explore the existence of significant differences between them [79]. The aim is to explore whether the mean differences between the two datasets are equal to zero or not in the population. However, one important assumption is that the two datasets need to be associated with each other [79]. Therefore, the test is based on the mean difference for each corresponding pair, and the mean difference needs to follow the normal distribution [79].

These are the hypotheses for the paired-samples t-test [80]: The null hypothesis is:

$$H_0 : \mu_1 - \mu_2 = 0$$

The alternative hypothesis of the two-tailed t-test is:

$$H_1 : \mu_1 - \mu_2 \neq 0$$

The alternative hypothesis of the one-tailed t-test is:

$$H_1 : \mu_1 - \mu_2 < 0$$

or

$$H_1 : \mu_1 - \mu_2 > 0$$

The formula of the t-statistic to explore the p-value is listed as Equation (3.2)[81].

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \tag{3.2}$$

where n represents the number of samples in a group, $\bar{d}$ refers to the average mean differences, while s_d represents the standard deviation of the mean differences.

The results will also be compared with the p-value, a reference statistic for whether to reject the null hypothesis or not [81]. Also, the positive results indicate that the mean of the population in the second group tends to be larger than that of the first group [81]. On the contrary, the mean in the second group tends to be lower than that of the first group.

However, there are some limitations of this data comparison method. The significant outliers in the differences will greatly influence the accuracy of the results [79]. There should be no extreme values in the mean differences. For example, if most of the mean differences are between 20 and 30, but one mean difference is 1000, then it is not appropriate to apply the method. Also, the paired-samples t-test is limited to the continuous scale data, including the interval and ratio level [79].

Generally, in practice, the test is adopted for the comparison of the means of certain groups under specific experimental conditions [82]. The examples include the comparison of student performance of the same student groups in different courses, the body results of the same residents before and after certain medical examinations or other experimental tests, and the comparison of certain aspects of twins [82].

Independent-Samples t-Test The independent-samples t-test is another important data comparison test for two samples, which is different from the paired-samples t-test to some extent [83].

Similar to the paired-samples t-test, the datasets are estimated on a continuous scale [82]. However, the two groups under this analysis method are not dependent on each other [83]. Some key examples of the test include comparing test scores between two random groups in the same courses and the number of articles submitted by two different types of professors in two different years in the same university [82]. Additionally, in the paired-samples t-test, the assumption is that only the corresponding mean differences based on each data point should follow a normal distribution. In contrast, for the independent-samples t-test, both datasets should follow the normal distribution [81].

The null hypothesis is [84]: $H_0 : \mu_1 = \mu_2$

The alternative hypothesis is: $H_1 : \mu_1 \neq \mu_2$,

where μ_1 represents the mean in the first sample group and μ_2 represents the mean in the second sample group.

T-statistic is also adopted in the analysis to explore the corresponding p-value for the hypothesis test. The formula for the t-statistic is (3.3) [84]:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \tag{3.3}$$

where n is the sample size, and the results will also be compared with the p-value, and the logic is similar to the paired-samples t-test.

There are also some limitations of the independent-samples t-test in practical applications. First, it is also limited to data of a continuous scale [85]. Another key limitation is that the samples must be randomly selected. However, it is arguable how to randomly select samples and whether the samples are representative of the population [85].

3.1.1.3 More Than Three Variables

In practice, some comparison tests need to include more than two groups, while the one-way ANOVA test is a typical one among the comparison tests for three or more variables.

The one-way ANOVA test is the extension of the independent-samples t-test, and it is usually adopted for three or more groups [86].In the test, researchers assume that the number of groups for comparison is k. These are the hypotheses for the one-way ANOVA test [86]: The null hypothesis is:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

$$H_1 : \text{At least one group mean differs from the others.}$$

In the test, the researchers will also adopt the p-value for comparison to make decisions on whether to reject the null hypothesis. In the test, the F-statistic is adopted to get the corresponding p-values and explain the data comparison for the research, which is shown as (3.4) [87].

$$F = \frac{MSB}{MSW} = \frac{\dfrac{\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2}{k-1}}{\dfrac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}{N-k}} \tag{3.4}$$

where MSB refers to Mean Square Between Groups, MSW represents Mean Square Within Groups, n refers to the sample size, k refers to the total group number, i refers to the indexing of each group under estimation, and j refers to the indexing of the number of the data point in the i group.

The one-way ANOVA test has been applied in many sectors because of its characteristics. One of the most popular ones is nursing research, especially testing phenomena [88]. Additionally, it has also been adopted in a large number of machine learning studies, especially in situations where there is a high dimensionality issue [89], where the one-way ANOVA test is adopted to filter out the features.

3.1.2 Non-parametric Comparison

Compared with the parametric method, the non-parametric one is not based on the assumption that the outcome variables in certain datasets follow a normal distribution [90]. However, in practice, it is very challenging to have a real normal distribution. Therefore, multiple comparison tests are set up for the non-parametric comparison, while the Wilcoxon signed-rank test is a typical example.

Some researchers define the Wilcoxon signed-rank test as the extension of the paired-samples t-test, while the assumption is that the distributions of the differences need to be symmetrical [91].

Generally, it is often applied for the comparison of the median between two datasets. All of the non-zero differences in the two groups are ranked from the lowest to the highest based on the absolute values. The sum of the ranking numbers of the positive differences is calculated, while the sum of the corresponding negative ones is also calculated. W^+ and W^- represent the sum of the positive ones and negative ones [91]. Then we need to select the lower ones to calculate the p-values. After getting the p-values, they will be compared with 5% or other thresholds to check whether to reject the null hypothesis or not [92].

3.2 Correlation Analysis

Correlation analysis is adopted to analyze the possibility and strength of association between two variables [93]. Generally, three common methods include Pearson correlation, Partial correlation, and Spearman rank correlation [93]. However, it just refers to the relevance between different variables instead of the explanation of the reasons for the changes.

3.2.0.1 Pearson Correlation Analysis

It is one of the most common methods in research studies. The assumption is that the datasets follow a bivariate normal distribution, so the method is usually adopted to analyze the jointly normally distributed data [94].

The corresponding calculation formula is shown as [94]:

$$\mathrm{r} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3.5}$$

where n represents the number of data points, $\bar{x}$ and $\bar{y}$ represent the mean value of the two datasets, respectively, and i refers to the index of the corresponding data points.

The correlation coefficients illustrate the directions and the strengths of the two variables, ranging from -1 to 1. The negative result refers to a negative correlation, while the positive result refers to a positive correlation. If the correlation coefficient is -1 or 1, it indicates that all of the observed values can be described on a straight line [95]. However, in practice, the correlation coefficient values are not up to the ideal level. If the absolute value is less than 0.40, then it is considered as weak; if the absolute value is between 0.40 and 0.69, then it is considered as moderate correlation; if the absolute value is no less than 0.70, then it is considered as strong correlation [95].

The application of the Pearson correlation analysis is under exploration. There has been a large number of corresponding research studies, including food industry research [96] and surgery research [97]. However, in some situations, some experts argue that despite not being normally distributed, the Pearson correlation analysis can still be adopted as a complementary tool, and the typical example is the correlation analyses in medical science [98].

3.2.0.2 Partial Correlation Analysis

Partial correlation is another important correlation analysis method. It is to explore the association between two variables by adjusting for the disturbance of other remaining independent variables [99].

The formula of the partial correlation coefficient is shown as (3.6) [100]:

$$r_{XY\cdot Z} = \frac{r_{XY} - r_{XZ}r_{YZ}}{\sqrt{(1 - r_{XZ}^2)(1 - r_{YZ}^2)}} \tag{3.6}$$

where X, Y, and Z represent the three independent variables under the analysis.

The partial correlation coefficient has been applied in various sectors. Among these, a typical example is its adoption in the financial market [101]. Specifically, researchers have employed this method in stock market studies to examine whether changes in the price of one stock influence the trends of another stock [101]. Since multiple factors in financial markets interact with each other, partial correlation is adopted to control for the influence of a third key variable, allowing for a clearer focus on the direct relationship between two specific stocks.

3.2.0.3 Spearman Correlation Analysis

The Spearman correlation coefficient is another method adopted in multiple research studies. Its applicability is broader than the other two correlation analysis methods in certain scenarios. For example, it can be applied to datasets that include some extreme values, as it is estimated based on rankings instead of actual values [102].

The analysis method is based on ranking data within each group [102]. In both groups, the values of each variable are ranked from the lowest to and highest [102].

The formula for Spearman correlation coeffecient is illustrated as Equation(3.7) [102]:

$$r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \tag{3.7}$$

where n is the sample size and d_i represents the corresponding differences between data points.

The method has also been applied to a large number of applications. For example, the analysis of the consumer preferences [103]. Moreover, it has been applied in some analyses in the transportation industry. For example, some researchers have adopted the method to analyze the reasons for the post-failure of city buses [104].

3.3 Regression Models

Artificial intelligence is booming in modern society, and machine learning methods are one of the most important components. There are a large number of corresponding techniques, including supervised learning, unsupervised learning, and semi-supervised learning [105].

Generally, the supervised learning method is usually adopted to address the classification and regression problems [105]. The classification means categorizing the output variables into several groups [105]. The key examples include scanning the emails to select the trashed ones or not, and categorizing certain animals as reptiles or mammals. Regression is the exploration of the relationship of several existing datasets [105]. Generally, it is applied for data predictions, explanations, causal inferences, and a large number of other aspects.

3.3.1 Linear Regression

Linear regression has been very commonly adopted for the analysis of the association among multiple variables. The method includes simple linear regression and multiple linear regression [106]. The simple linear regression aims to explore whether there is a linear relationship between one independent variable and one dependent variable [106]. The multiple linear regression (MLR) is adopted to explore the relationship between multiple independent variables and one dependent variable, which is to find the most appropriate weights of each predictor, the independent variables, for the predictions [106].

However, despite its popularity, there are several assumptions. Firstly, it is assumed that there should be a linear relationship, while the data points in all of the datasets should be independent, which means that one observation should not impact the values of the other observation [107]. In addition, the prediction errors between the predicted ones and the actual ones should be nearly identical, which means that there should not be large fluctuations among all of the errors, while the errors should follow a normal distribution. Another important assumption is that no linear relationship exists between any two independent variables [107]. However, in practice, it is very challenging to ensure that the five assumptions are exactly right. Therefore, in practice, researchers may also adopt the method even if the statistics do not align with some assumptions.

There are numerous applications of the two types of linear regression models in different disciplines. For medical science, they have been adopted in the association between evaluating certain behaviors and diseases, while for the finance sector, they have been estimated for the investment of certain assets, and some experts analyze different types of environmental science issues with the method [108].

3.3.2 Support Vector Machine

Support Vector Machine (SVM) is another widely adopted machine learning method. Some scholars proposed the concept in the 1970s [109]. Originally developed for classification tasks, SVM was subsequently extended to regression problems, resulting in the formulation of Support Vector Regression (SVR) [109].

SVR is also designed to construct a regression model, but the training process to get an optimal result is different from the linear ones, and the formula is illustrated as (3.8) [110].

$$f(x) = \mathbf{w}^T\mathbf{x} + b \tag{3.8}$$

However, the loss functions of the two regression models are greatly different from each other. Generally, the linear regression model takes all of the losses into consideration, but not all of the differences are included under the regression analysis of SVM [110]. Only the sensitive loss functions are considered.

Because of its suitability for high-dimensional spaces, it has been adopted for some predictions in different research studies. For example, some researchers have adopted it for short-term forecasting COVID-19 trend based on the existing database [111]. Also, some other researchers have explored fruit quality attributes with the support vector machine models and other methods [112]. Additionally, some researchers have found that the predictions were highly accurate in predicting porosity in reservoir evaluation [113].

However, researchers also point out that there are some limitations of SVM in regression analyses. First, it is not very suitable for handling large datasets due to its high computational complexity [110]. In addition, because of the lack of probabilistic output, its reliability may be limited in various scenarios [110].

3.3.3 K-Nearest Neighbours

The history of the application of the k-nearest neighbours(KNN) method is very long. It was originally proposed in 1951 and extended to other applications such as data mining [114].

There are multiple steps of the regression model, while the key principle of KNN is its likelihood of similarity [114]. The first step is to select K. K refers to the nearest neighbor number for a new record [115]. It needs to compare all of the data points to explore the optimal value of K, which is greatly associated with having accurate model predictions [114]. The commonly applied method for the determination of K is experimentation and cross-validation, while in some domains, some experts may adopt a predetermined value [115]. The reason for the strict process is that a lower K may result in overfitting, which reflects some extreme values instead of the overall patterns [114], while a higher K may result in underfitting [114]. The second step is to calculate the distance from all data points based on a new instance, while Euclidean is one of the most common metrics [114]. After calculating the distance, the researchers will select several points according to the K-value, while the last step is to calculate the final value, and in the regression model, the final value is usually the mean or median value [114].

However, despite its simplicity in implementation, the KNN method has certain limitations in practice. One reason is that large-scale databases negatively affect computational efficiency due to increased calculation time [114]. In addition, in high-dimensional problems, the model often fails to achieve optimal performance [114].

Because of its simplicity, it has been adopted in some key areas. For example, it has been applied to disease prediction [116]. Also, in the energy industry, some experts applied the method for the prediction of the solar radiation levels [117]. However, because of its limitations, in many situations, some experts have enhanced its applications with other corresponding methods to improve the prediction accuracy rates [117].

3.3.4 Guassian Process Regression

Gaussian Process Regression (GPR) is an important non-parametric method. It is simple to implement, and its flexibility makes it a powerful tool for various applications [118], as it is based on Bayesian principles.

The logic of GPR differs from that of traditional regression models. The key concept is that, based on the existing data points, the kernel of GPR represents a large number of potential regression functions. Each of these functions is assigned a probability, reflecting different possible regression models. By combining all these models, GPR produces a weighted average of the predictions [118].

Because of its characteristics, there are a large number of applications in multiple industries. In China, it has been applied to time series prediction, and it has also been applied for control system design and identifying dynamic system models [119].

However, one of the key limitations is the limitation of the data size. The computation is very expensive, while if the dataset is sufficiently large, data storage and processing are also important issues to be taken into consideration [118].

3.3.5 Partial Least Squares

Partial Least Squares (PLS) is a set of modeling methods that establish relationships between multiple observed variables and latent variables [120]. It is based on the assumption that a small number of latent variables drive the system and generate the observed variables [120]. The underlying principle of PLS is to transform multiple independent variables into latent variables based on their weights and other factors, thereby reducing feature dimensionality and addressing multicollinearity issues in regression models. This process is repeated several times to identify the most appropriate set of latent variables [120]. Therefore, PLS is suitable for models with many features and relatively small sample sizes.

PLS has been applied across various industries and academic disciplines. For example, it has been adopted in environmental protection efforts, including improvements in plastic waste management [121]. In addition, some researchers have employed this model to analyze public attitudes toward social restrictions in Indonesia during the lockdown period [122].

However, the model also has certain limitations. One key disadvantage of PLS regression is that the parameter estimates of some PLS methods are not consistent, and when the sample size is large, the model tends to exhibit bias [120].

3.3.6 Multi-Layer Perceptron

The Multi-Layer Perceptron (MLP) is a type of neural network composed of an input layer, one or more hidden layers, and an output layer [123]. The input layer receives the original data, while the hidden layers perform weighted calculations and apply activation functions. The final output is produced by the output layer [123].

This method offers several advantages. First, it is relatively simple for researchers to implement because it does not require extensive prior knowledge [124]. Additionally, the model generates results directly through the learning process and can capture both linear and nonlinear relationships [124].

Due to these advantages, MLP has been widely applied across various fields. For example, it has been utilized in diagnostic systems [125] and for optimizing microcellular radio environments [126].

However, the method also has limitations. First, there are no established standards for determining the number of layers in the model, necessitating repeated experimentation [124]. Moreover, the model requires retraining when new datasets are introduced in addition to the original data, which increases computational complexity [124].

3.3.7 Decision Tree Regression

It is another common non-parametric data analysis method. The name of decision tree regression (DTR) originates from its corresponding branching structure, and it has been adopted for data analyses since the 1960s, including classification tree and regression tree [127].

The advantage of DTR lies in a large number of aspects. To begin with, the implementation process is very easy to operate, which simply splits the data statistics into several groups [127]. In addition, in practice, there are some skewed data points in the original datasets; DTR can reduce the impacts of the extreme values in the datasets [127]. Moreover, because it simply splits the datasets into the subgroups, which enables the participants to have an enhanced understanding of the final results, and it is more coherent for the researchers to have a systematic interpretation [127].

The logic of DTR is very straightforward. Based on certain criteria, the datasets are divided into two leaf nodes, and then both of the subgroups are continually split into different groups, and the predictions in both of the subgroups are the average value of the dependent variables in each subgroup [127]. Therefore, the prediction values are different in each group.

It can be applied in a large number of sectors. For example, some experts have attempted to utilize it to optimize the business process [128]. In addition, it has also been adopted for the predictions of in-house mortality for medical areas by some experts [128]. Moreover, it has also been adopted in the travel industry [128].

3.4 Evaluation Metrics

For each regression model, there are likely some differences between the results generated by the models and the actual ones. The differences among regression models, based on observed sample values, exhibit substantial variation.

Therefore, researchers need to select the most appropriate evaluation metrics for comparison to explore whether the estimated values based on certain models are accurate. Each evaluation metric has its advantages and disadvantages. Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), and Root Mean Square Error (RMSE) are the four most common metrics for evaluating whether there are any differences between the estimated values with regression models and the ones in the original datasets.

MAE is an important measure of the errors. It is an indicator to explore whether the differences between the estimated values and the actual ones are significant or not [129]. It has been generally applied for model evaluation, which has been adopted in many disciplines, including data mining, biosciences, geosciences, and other subjects [130]. The formula of MAE is illustrated as Equation(3.9)[129].

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{3.9}$$

where n is the sample size, y_i refers to the actual value of certain independent variables and $\hat{y}_i$ represents the predicted values from regression models. If MAE is equal to 0, it means that the results are perfect in practice.

As shown in the formula 3.9, the key advantage of MAE is its directness. The differences are intuitive because the absolute errors are accumulated, which is not influenced by the weight or other factors [129].

However, the key disadvantage of MAE is that it does not capture the magnitude of changes, as it is scale-dependent. In practice, different datasets have varying scales. Although the MAPEs may not differ significantly, the proportion of changes can vary significantly.

MAPE is another important forecast accuracy metric. The calculation of MAPE is (3.10) [131].

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|, \tag{3.10}$$

where n represents the sample size, y_i represents the actual values and $\hat{y}_i$ denotes the predicted values.

The formula 6.2 illustrates that one of the key advantages of the metric is the illustration of the proportion of the extent of the change levels. Multiple datasets have their scales. The influence of the same differences between the estimated values and observed ones varies in different groups, with varying ranges. The weight of the changes is under analysis and illustrates the corresponding changes for different datasets.

However, one of the most significant disadvantages of it is the accuracy value. In some scenarios, the actual values are minimal, so the MAPE values are quite large [131]. When the observed values are closer to zero, it may provide infinite or undefined values [131].

MSE is another important evaluation metric.

The formula of the MSE is demonstrated as Equation (3.11)[132]:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3.11}$$

where n represents the sample size, y_i represents the actual values and $\hat{y}_i$ denotes the predicted values.

As depicted in the formula 3.11, the metric estimates the mean differences between each actual value and the corresponding predicted ones. The average change is illustrated, which shows the overall differences between the predicted values and the observed values [132].

However, one key disadvantage is that some differences may be magnified. The reason is that the formula 3.11 amplifies the differences. Therefore, it is likely that if there are some extreme differences in values, then results may overstate the actual differences.

RMSE is just the square root of MSE, and the formula is shown as Equation(3.12) [129]:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3.12}$$

where n represents the sample size, y_i represents the actual values and $\hat{y}_i$ denotes the predicted values.

3.5 Databases

Several chapters in the thesis focus on COVID-19 and education. Therefore, it is important to select the most accurate COVID-19 and education datasets.

3.5.1 COVID-19 Databases

The thesis is mainly focused on the impacts of COVID-19 and corresponding issues, so it is essential to select the most appropriate COVID-19 datasets. Throughout the periods of the pandemic, a large number of COVID-19 datasets have been collected. Some datasets reflect the cases and numbers of the pandemic, while others reflect the vaccinations throughout the historical period. It is important to categorize the different databases and conduct a comprehensive review of multiple COVID-19 databases with multiple criteria.

First, it is important to select databases that are collected by authoritative organizations, including governments or other non-profit public organizations. The aim is to ensure the reliability of the databases, which can more accurately reflect the impacts of the pandemic. Generally, governments usually have more resources to allocate for the systematic summary of statistics, so it is important to adopt governmental databases.

In addition, it is necessary to select comprehensive databases with multiple criteria. One important metric is to select databases spanning multiple years, which reflect the changes of the pandemic throughout different periods. Another important metric is to select international databases, because generally, the data collection methods for identical databases are consistent across multiple countries. If the COVID-19 databases are from multiple countries, the data collection methods are very likely to be inconsistent, making it impossible to compare the varying influences across countries. Moreover, it is also important to select databases that cover multiple aspects of the COVID-19 pandemic. The factors include the daily cases, the accumulated cases, the vaccinations taken, the people fully vaccinated, the emergency level, and other relevant ones. Therefore, the databases selected also need to include multiple aspects of the COVID-19 pandemic.

In addition, the selection of databases with daily statistics is also essential. The reason is that daily databases can reflect the changes in COVID-19 trends within each country, while also allowing comprehensive comparisons with the corresponding policies.

3.5.2 Academic Performance Metrics

Different academic performance datasets are also collected by national and international organizations.

There are two important aspects of the evaluation of academic performance. One is the test scores. The test score is one of the general evaluation methods of different educational institutions to evaluate the study achievements in certain subjects. The results tend to be more objective, and it is more appropriate for the researchers to have a comprehensive analysis of different performances based on the quantitative results, so the criteria are appropriate to measure the study abilities among students, especially during the COVID-19 pandemic. Another important aspect is the attendance rate in different educational institutions, including different types of primary schools, secondary schools, and childcare services. The attendance rate is another important metric in education, as it reflects student participation throughout the year, which directly illustrates the motivation towards studies to some extent. It is suitable for the analysis of the influence COVID-19 pandemic on education, because the results can be analyzed with the trends of the changes of the pandemic.

3.5.2.1 Test Results Databases

It is important to select international databases for international examinations or tests. Different countries implement their own education systems and cultural backgrounds, so even the comparison of the same subjects can be biased due to significant differences in difficulty levels, examination formats, and grading criteria. Therefore, tests or examinations with consistent standards need to be selected for corresponding analysis, which are generally international tests.

Also, the international databases should span multiple years with consistent standards. This is necessary to enable analysis of pandemic-related factors and academic performance across different phases of the pandemic, allowing identification of consistent patterns over time.

Globally, two types of examinations have consistent standards. One type is international English tests, including IELTS (International English Language Testing System), TOEFL(Test of English as a Foreign Language), and GMAT (Graduate Management Admission Test), which are adopted in a large number of countries with identical examination formats. Generally, these tests are required for university admissions or visa applications.

Another type of examination is international standardized tests across different countries, including the Programme for International Student Assessment (PISA) [133], the Trends in International Mathematics and Science Study (TIMSS) [134], the Progress in International Reading Literacy Study (PIRLS) [135], and the Programme for the International Assessment of Adult Competencies (PIAAC) [136]. PISA is applied to measure the abilities of reading, mathematics, and science for 15-year-old students [133]. TIMSS is another important examination that explores students' achievements in mathematics and science at the fourth and eighth grades [134]. PIRLS evaluates the reading comprehension skills of Year 4 students [135]. PIAAC is adopted to assess information processing skills, including cognitive and workplace skills, of adults [136].

However, there are limitations to these skill evaluation tests. The testing cycle is three years for PISA [133], four years for TIMSS [134], five years for PIRLS [135], and more than five years for PIAAC [136].

Therefore, given the time intervals and the completeness of the databases across multiple periods, the global English tests with yearly data have been selected.

3.5.2.2 Attendance Rate Databases

There are several standards for selecting attendance rate databases. First, the data should be collected consistently over at least two years to capture different phases for the comprehensive comparison. Moreover, the attendance rates should be recorded daily to reflect trends in the impact of the pandemic and associated lockdown policies. In addition, datasets that include different categories of students or participants, such as by ethnicity or regions, are also important, as they help to reveal the various impacts on multiple demographic groups. Therefore, if possible, the datasets of a multicultural country with online records of daily attendance rates should be selected.

Chapter 4

Analyzing the Impact of COVID-19 on Education: A Comparative Study Based on TOEFL Test Results

4.1 Introduction

In response to the COVID-19 pandemic, numerous educational institutions in different countries had to implement home-based teaching and tests [137]. Due to the wide adoption of online learning, researchers have begun to investigate the impact of the pandemic on academic performance. Extensive research studies have explored the impact of the pandemic on education sectors across some subjects and regions [138, 139, 140]. However, one significant shortcoming is that most studies focused only on one or several subjects and countries, lacking the evolving impact during different stages of the pandemic.

To address gaps in the previous studies, the research question in this chapter is: "How does academic performance across regions change before the COVID-19 pandemic, at the peak of the COVID-19 pandemic, and after the peak of the COVID-19 pandemic?" The research question aims to explore how the educational outcomes across regions differ before and during the COVID-19 pandemic.

Therefore, the chapter contributes to future academic studies and practice from two perspectives:

- Firstly, this chapter statistically shows that the COVID-19 pandemic and corresponding educational modes bring benefits to the education sector through an analysis of the TOEFL test worldwide.
- Secondly, it provides brief insights into the dynamic impact of pandemics through the pandemic and possible explanations for the findings based on the analysis and previous research studies.

The rest of this chapter is arranged as follows: Section 4.2 identifies the key gaps from previous studies. Section 4.3 presents the research method and data source. The findings from the data analysis are displayed and discussed in Section 4.4 and 4.5, respectively. Finally, Section 4.6 illustrates the implications while Section 4.7 concludes the chapter with future directions.

4.2 Literature Review

Some experts researched the impact of the pandemic on the academic performance of students in primary and secondary schools. A study on fourth-grade primary school students in Germany revealed a drastic decline in mean reading achievement between 2016 and 2021 [141]. However, a study of 100 primary students in Spain indicated that more than 40% of the respondents reported no significant change in their academic performance during the pandemic [142]. Additionally, an academic study on senior secondary students revealed a significant decrease in academic performance after the COVID-19 pandemic began [143]. Some scholars also found that, compared with the pre-pandemic mathematics achievements among over 368 secondary students in Spain, there was a significant decrease after the implementation of lockdown policies [144].

Moreover, extensive research studies were conducted on the influence of the pandemic on study performance in universities. An investigation of 200 U.S. public higher education institutions showed that GPA(Grade Point Average) slightly increased after the implementation of online studies [145]. However, findings based on research at Nueva Ecija University of Science and Technology suggested that academic performance did not show significant changes after the widespread adoption of online education [146].

Despite the comprehensive findings, a major missing part is the comparative studies worldwide. In general, the current research is lacking in the analysis of cultural diversity and education system differences in multiple countries and regions. In addition, most studies lack a comprehensive analysis of the dynamics of pandemic effects before and during the outbreak. Therefore, it is necessary and important to conduct a comprehensive comparison over the period and across the regions.

4.3 Methodology

4.3.1 Data Collection and Pre-processing

In this chapter, a quantitative methodology is applied to explore the changes in academic performance, which is applied in multiple research studies. For example, it is applied in a research study on an e-health system during the period of pandemic [147], where statistics are presented concisely with graphs and charts. Therefore, it is appropriate to apply the quantitative methodology to conduct research with concise and direct comparisons across different periods and regions.

The data source is selected regarding two important criteria, availability and consistent standardization [148, 149]. Generally, public databases from international institutions can be accessed through online sources, and in many cases, the data is collected with consistent criteria. For example, public databases are utilized in the research of a certain classification approach of analyzing alcoholic Electroencephalogram signals [150].

Therefore, the datasets of the TOEFL test are selected for data analysis. TOEFL is a globally recognized test with four sections, including Reading, Listening, Speaking, and Writing, allowing for in-depth exploration of different subsections. The TOEFL data is sourced from the reports[1] of the Education Testing

[1] https://www.ets.org/toefl/teachers-advisors-agents/ibt/scores.html

Service (ETS)[1], a non-profit organization certified by the international institutions, which ensures the reliability of data because of consistent sampling and calculation criteria. Moreover, the reports sourced from ETS cover a substantial time span and a wide range of countries, which is appropriate for comparative studies both in time and space.

On the other hand, it is crucial to select the appropriate periods for the research [151, 152]. The outbreak of COVID-19 was in 2020 [141], and multiple education institutions gradually eased their restriction in 2021 [153]. For the pre-pandemic period, the years 2017, 2018, and 2019 are selected. In the chapter, the annual results from 2017 to 2021 will be reorganized into four groups for comparison purposes, which are Group 1 (2018-2017), Group 2 (2019-2018), Group 3 (2020-2019), and Group 4 (2021-2020).

However, the consistency of comparison should be maintained. If one country is missing any years in the dataset, it will be excluded from the analysis. Additionally, the databases of COVID-19 cases and deaths are introduced to serve as an indicator of pandemic severity. Therefore, countries that are not on the COVID-19 report[2] are also excluded for data consistency concerns.

In addition, it is necessary to assess whether the sample size is appropriate. In total, 147 countries or regions are selected. Among them, 34 are selected out of 51 in Africa, while 25 are selected out of 47 in America. In Asia, 23 are selected out of 35, while 43 are selected in Europe out of 53. In the Middle East, all of the 19 are selected. However, only 3 countries or regions satisfy the data selection criteria in the Pacific out of 23, which is not sufficient to represent the academic performance of this grand region. As a result, Pacific countries or regions will be included in the overall worldwide analysis, while the Pacific grand region itself will be excluded from the grand region analysis.

4.3.2 Research Design

The paired-samples t-test is applied for the comparison between the two years within each group. In general, the paired-samples t-test compares the means of

[1]https://www.etsglobal.org/fr/en/content/who-we-are
[2]https://ourworldindata.org/covid-vaccinations?country=OWID-WRL

the two measurements from the same units [79], which is appropriate to compare the differences in the mean scores between two years within each group of the same countries. Hypotheses are formulated to compare whether the results are equal or not between two consecutive years.

Therefore, the hypotheses for the paired-samples t-test aim to test whether the mean difference between the two years within the group is equal to zero or not [80]:

H_0: m_1 - m_2=0 (The mean difference of the populations within the group is equal to 0.)

H_1: m_1 - $m_2 \neq 0$ (The mean difference of the populations within the group is not equal to 0.)

4.4 Results

4.4.0.1 Total Scores in the World and Different Regions

As shown in Table 4.1, star marks are used to indicate the significance level of the results, which represent a p-value less than 5% with one star and a p-value less than 1% with two stars. As Table 4.1 shows, none of the p-values in Group 1 reach the preset significance level (5%), indicating that in all regions, the average total scores do not exhibit significant changes between 2018 and 2017. However, half of the p-values in Group 2 pass the 5% significance threshold, which suggests considerable differences in total scores worldwide. Specifically, in comparison with the result in 2018, the average total scores in Asia and the Middle East were significantly higher in 2019 by 0.565 and 1 point, respectively. At the same time, the test results from Africa, America, and Europe do not show significant differences for Group 2 (2019-2018). As for Group 3 (2020-2019), the significance levels of all regions are less than 0.05 or 0.01, and positive mean differences are observed, suggesting that compared with 2019, the means of the total scores were significantly higher in 2020 across all regions. In Group 4, the significance level indicates there are no significant differences in the mean scores in the five regions.

We visualize the test score mean difference for all the groups and regions in Fig. 4.1, and it clearly shows the significant increase in total TOEFL scores

Table 4.1 Group Mean Difference across Regions (Total Scores Based)

Group	World	Africa	America	Asia	Europe	Mid East
G1 (2018-2017)	-0.075	-0.441	-0.320	0.304	-0.023	0.000
G2 (2019-2018)	0.395*	0.353	-0.120	0.565*	0.349	1.000*
G3 (2020-2019)	2.367**	2.706**	2.731**	2.087**	2.119**	2.368**
G4 (2021-2020)	-0.041	-0.735	-0.320	0.261	0.535	0.000

*p<0.05, **p<0.01

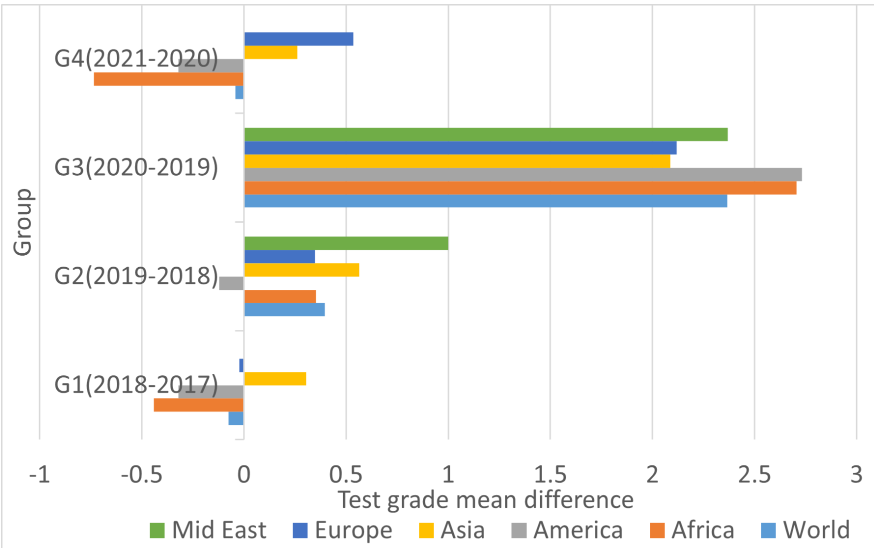


Fig. 4.1. Mean Difference of Total Scores for Each Group across Regions

Table 4.2 Group Mean Difference Globally (Section Grade Based)

Group	Reading	Listening	Speaking	Writing
G1 (2018-2017)	-0.238**	0.088	0.000	-0.170**
G2 (2019-2018)	0.286**	0.259**	0.095*	-0.041
G3 (2020-2019)	0.769**	0.714**	0.116*	0.653**
G4 (2021-2020)	0.075	0.014	-0.014	-0.109

*$p<0.05$,**$p<0.01$

worldwide between the years 2020 and 2019.

4.4.0.2 Section Scores Worldwide

The results in Table 4.2 reveal significant differences in the academic performance changes among different test sections worldwide in the four groups. As the first row in Table 4.2 shows, there exists a slight performance drop in the Reading and Writing test when comparing the result in 2018 with the result in 2017, and the significance level for this conclusion is as high as 0.01. However, in Group 2 and Group 3, the significance levels of Reading, Listening, and Speaking, and the mean differences indicate significant increases in these test sections. Compared with the results in Group 2, the mean difference values are considerably higher in Group 3 ($0.769 > 0.286$, $0.714 > 0.259$, $0.116 > 0.095$), indicating a significant improvement in performance in 2020. Surprisingly, in Group 4, there are no significant differences for the four test sections of TOEFL worldwide since no star is marked. Therefore, the mean differences value and significance level show that the increase in mean scores of the four test sections between 2020 and 2019 is the most significant, while no significant changes are observed between 2021 and 2020.

We draw the bar chart of mean difference for the four test sections across all the groups in Fig.4.2, and the surge of test performance for all four sections could be clearly observed in Group 3, which compares the results between 2020 and 2019.

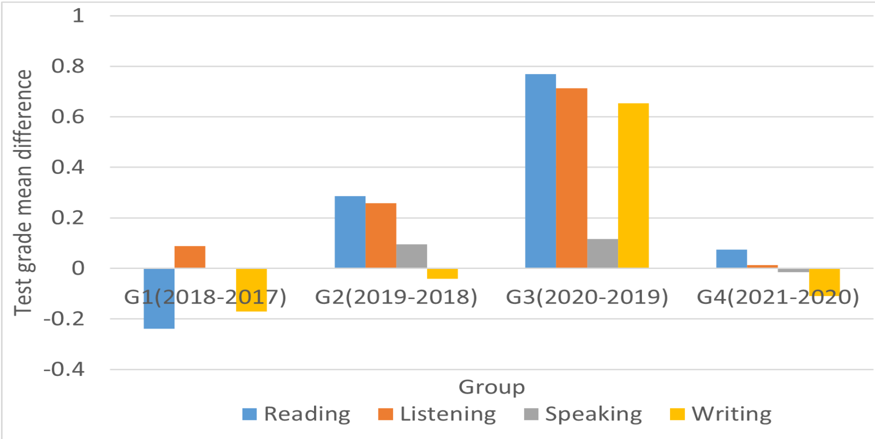


Fig. 4.2. Mean Difference for Each Group Globally (Section Grade Based)

Table 4.3 Group Mean Difference across Regions (Section Grade Based)

Region	Group	Reading	Listening	Speaking	Writing
Africa	G1 (2018-2017)	-0.412**	0.118	-0.088	-0.206
	G2 (2019-2018)	0.059	0.412*	0.176	-0.088
	G3 (2020-2019)	0.735**	0.735**	0.412**	0.588**
	G4 (2021-2020)	0.029	-0.206	-0.176	-0.206
America	G1 (2018-2017)	-0.280*	0.000	0.120	-0.160*
	G2 (2019-2018)	0.120	-0.080	0.000	0.000
	G3 (2020-2019)	0.808**	0.808**	0.115	0.769**
	G4 (2021-2020)	-0.080	0.040	-0.120	-0.200
Asia	G1 (2018-2017)	0.000	0.130	-0.130	-0.261
	G2 (2019-2018)	0.261	0.261*	0.304*	-0.130
	G3 (2020-2019)	0.739**	0.696**	0.000	0.565**
	G4 (2021-2020)	0.348*	0.261*	-0.087	0.043
Europe	G1 (2018-2017)	-0.256*	0.023	0.023	-0.070
	G2 (2019-2018)	0.442**	0.186	-0.047	-0.047
	G3 (2020-2019)	0.786**	0.714**	0.024**	0.643**
	G4 (2021-2020)	0.070	0.070	0.326**	-0.047
Mid East	G1 (2018-2017)	-0.263	0.105	0.105	-0.263
	G2 (2019-2018)	0.579**	0.526**	0.105	0.105
	G3 (2020-2019)	0.842**	0.789**	-0.105	0.737**
	G4 (2021-2020)	0.105	-0.158	-0.105	-0.053

*p<0.05,**p<0.01

4.4.0.3 Performance Breakdown for Sections and Regions

To further investigate the dynamics of how the score changes across the four sections and in all regions, we break down the results based on both regions and test sections in Table 4.3. Different patterns of mean differences and significance levels are observed across the regions and test sections. Specifically, the results in Group 3 (2020-2019) show the most significant increase in test scores across different regions. Within this Group, the average scores in Reading, Listening, and Writing enjoyed an overwhelming and significant increase when comparing the results of 2020 with the results of 2019. However, except for Africa, the results suggest that the scores in Speaking do not have a significant change.

In Group 1 (2018-2017), despite a few significant changes in some regions in terms of some sections, the majority of the regions did not see a significant change in most of the test sections. In detail, only the p-values and mean differences in Reading in Africa, America, and Europe, and Writing in America indicate a significant decrease.

Additionally, in Group 4, there only exist significant increases in Reading and Listening scores in Asia, and Speaking scores in Europe, while the test scores of other sections across the rest regions exhibit no significant changes between the years 2021 and 2020.

The results in Group 2 (2019-2018) are the most divided among the four groups. On one end, the reading score in Africa, America, and the Middle East witnessed a significant increase in 2019, while the Listening score in Asia and the Middle East saw a significant increase. On the other hand, the test score for Speaking shows a significant increase in Asia, while the score for Writing does not exhibit a significant increase in the Middle East.

Therefore, out of all the complicated change patterns across the regions and test sections, the overall increase in test scores in Group 3 (2020-2019) is the most significant and obvious.

4.5 Discussion

The findings from the previous part offer insights into how different stages of the COVID-19 pandemic have factored into the changes in TOEFL test results. At the initial stage of the pandemic, the increase in total scores in the world and most regions is much more significant than the improvement before the COVID-19 pandemic. Similarly, the mean differences in the four test sections are also more significant in the early stages of the pandemic. However, after reaching the peak of the COVID-19 pandemic, the results show that the upward trend of substantial increase across the test sections and regions diminished.

For this phenomenon, our best guess is that the increase in mean scores during the pandemic is very likely due to the widespread application of online learning and testing worldwide. On the one hand, online courses facilitate flexible interaction and timely revision, which enhances the daily study processes. An academic study among secondary students shows that online learning during the COVID-19 pandemic might stimulate the creativity of students [144]. On the other hand, the flexibility of the online home testing of language tests might help students achieve better results[1].

The flexibility of online learning might also help explain the decline in the increase in test results globally. A study in 2022 [137] claims that online learning increases the flexibility of the materials and the attendance rate in digital lectures is significantly higher than the traditional ones, thus boosting academic performance. When transitioning from online courses to face-to-face learning after the pandemic, the attendance rate might decrease in certain institutions, which might result in a performance drop. Another academic study [153] among some Chinese international students also finds that the abrupt and substantial reduction of digital learning disrupts the 'pandemic-style' study process, which, to some extent, brings negative effects on daily studies due to the lack of flexibility.

[1]https://ischoolconnect.com/blog/coronavirus-impact-on-tests-like-gre-gmat-ielts-toefl/

4.6 Implications

The chapter provides an overview of the academic performance changes worldwide with consistent TOEFL standards. During the pandemic, it is very common for an educational institution would set up online courses, and this study has provided theoretical implications that countermeasures such as online learning during the pandemic may present opportunities for improving study outcomes. Therefore, the adoption of a new education mode, with some limitations needed, may enable students to achieve better academic performance on a global scale.

Another important implication of this chapter is the dynamic impact of the COVID-19 pandemic on a global scale. Previously, many research studies only investigated specific periods during the COVID-19 pandemic in limited countries or regions. However, it is possible that there were similar patterns of changes in test results before the COVID-19 pandemic. Moreover, very few studies have investigated whether the impacts of the pandemic have differed along different stages. The study provides great insights into different stages of the pandemic, suggesting directions for future studies in terms of the dynamic COVID-19 impact on education.

4.7 Conclusion

In conclusion, the chapter examines the changes in academic performance during different periods of the COVID-19 pandemic through a comprehensive analysis of TOEFL test results worldwide. The findings indicate there has been a significant increase in TOEFL scores across different test sections after the COVID-19 pandemic broke out globally. However, the performance improvement stopped after the pandemic reached its peak. The study contributes to the understanding of the similarities and differences among different learning areas globally, which provides insights into the changing impacts of the pandemic during different periods.

However, some limitations need to be addressed in future studies. Firstly, the current databases available only include integer values, which might reduce the accuracy of the studies. Secondly, the scope of this chapter is limited to the results of the TOEFL test, neglecting other language proficiency tests. Thirdly, the indicator for academic performance is quite limited in this chapter since only the TOEFL test score is analyzed, and other indicators, such as attendance rate, could be investigated in the future. Therefore, future studies should encompass a broad range of language studies, various indicators of the study effects, and multiple periods to reveal the dynamic impact of the pandemic on education.

Chapter 5

The Impact of COVID-19 on Global GMAT Test-Taking Patterns: A Statistical Analysis

5.1 Introduction

The COVID-19 pandemic has had a profound global impact since its discovery in 2019, spreading rapidly to nearly every country worldwide [1]. Among the significantly affected sectors, education, particularly examination procedures, underwent substantial transformations. Despite the pre-existing popularity of e-learning [154], the pandemic accelerated its adoption globally [155]. This shift to online testing required students to take exams remotely from their homes [156], leading to various implications.

Firstly, this shift resulted in psychological effects, such as decreased motivation and increased stress levels among students [157]. Secondly, it notably impacted academic performance, particularly during the initial stages of the pandemic [158]. Additionally, these changes influenced the demographic of examinees, with some exams being canceled due to pandemic-related challenges [159].

Despite existing studies, significant gaps persist in the literature, notably the absence of comprehensive comparative analyses of examinations with consistent global standards, such as those for exams like GMAT. Furthermore, research often overlooks how the association between COVID-19 and examinations varies across different years. To address these gaps and explore global testing changes during the pandemic, this chapter aims to answer the research question: "How does the COVID-19 pandemic influence GMAT tests globally, including examinee numbers and average scores?" This chapter contributes in two main aspects: systematically illustrating similarities and differences in COVID-19 impact on examinee numbers and academic performance across continents during different pandemic periods, and demonstrating correlations between new COVID-19 cases/deaths and examinee fluctuations.

The chapter is organized into several sections for clarity and coherence. Section 2, the Literature Review, critically examines existing studies, analyzes their findings, and identifies research gaps. Section 3, the Research Design, meticulously details data collection and analysis methods. Section 4 presents experimental results, offering insights derived from datasets, summarizing key findings, discussing chapter limitations, and providing a critical assessment. Finally, Section 5, the Conclusion, consolidates research outcomes and explores future research directions.

5.2 Literature Review

Numerous scholars have conducted extensive investigations into the reduction of examinations and tests across various countries, a trend anticipated to decrease the number of examinees. Research highlights a significant decline in examination numbers in Norway, where schools eliminated all final exams in response to the COVID-19 pandemic [160]. Similarly, many A-level exams were canceled in England due to the pandemic's impact [159]. Furthermore, a global study reported that entrance examinations were canceled in over 100 countries since the onset of the COVID-19 pandemic [161].

In addition to the reduction in examinations potentially influencing the number of examinees, researchers have explored the pandemic's effects on academic progress and student engagement. For instance, a study in Afghanistan revealed that many students perceived the COVID-19 pandemic as detrimental to their educational advancement [162]. Similarly, research from the Netherlands indicated that students encountered significant challenges in maintaining academic progress while adapting to remote learning environments [58]. Further research has underscored considerable learning setbacks among children due to COVID-19 lockdowns [59]. Additionally, universities in over 100 countries were compelled to cancel their entrance exams [161].

The pandemic's impact on academic performance has also been a critical study area. Research on the Italian university system indicated a decrease in exam pass rates during the pandemic compared to the pre-pandemic period [55]. Similarly, studies from Ghanaian colleges of education demonstrated a decline in academic performance amid the COVID-19 pandemic [163]. Additionally, a research study across multiple universities in China highlighted how anxiety and depression during the pandemic adversely affected academic outcomes [164].

While existing research consistently underscores the profound and detrimental effects of the COVID-19 pandemic on academic progress and performance, a significant limitation is its narrow geographical focus, often concentrating exclusively on specific countries or regions. This restricted scope fails to capture the diverse educational disruptions experienced globally. Furthermore, many studies have not conducted comparative analyses across multiple years of the pandemic, which is crucial for understanding temporal variations in its impact. Therefore, there is a compelling need for a more comprehensive and globally inclusive analysis spanning multiple years to provide a more accurate assessment of the effects of the pandemic on examinations such as the GMAT test.

5.3 Research Design

This chapter adopts a quantitative methodology to examine the relationship between the COVID-19 pandemic, the number of examinees, and their performance scores. This methodological approach is selected for its objectivity and precision in analyzing quantitative datasets and is widely recognized in scholarly research for its rigor. For instance, a quantitative approach was utilized in a study investigating the impacts of the pandemic on e-learning [165], and similarly in research concerning information dissemination within virtual universities [8].

5.3.1 Materials

Careful selection of appropriate datasets is crucial for the success of the research. Utilizing public databases from authoritative sources will ensure robustness and reliability. Given the global scope of our research, it is essential to adopt a universally recognized test accepted across countries. Furthermore, consistency in COVID-19 statistics within chosen datasets is paramount.

The GMAT test has been chosen for several compelling reasons. The GMAT serves as a widely recognized admission criterion in over 1,700 schools worldwide [166]. Its standardized test sections and scoring protocols remain consistent across countries, allowing for rigorous comparisons across diverse nations and regions [166]. This uniformity facilitates comprehensive and reliable international assessments.

For this chapter, the "Profile of GMAT Testing 2023"[1] report has been chosen as the primary dataset for two crucial reasons. First, the data is compiled by the Graduate Management Admission Council (GMAC), a reputable global non-profit organization, ensuring the authority and reliability of the dataset. Second, this report includes comprehensive data on GMAT scores and the number of test-takers spanning from 2019 to 2022, encompassing over 120 countries worldwide. This extensive dataset supports a thorough and robust analysis across multiple countries and years, enhanced by gender-based categorizations that provide nuanced insights into global trends. Moreover, consistent standards ensure the objectivity of the research. In the profile, the number of examinees is represented by the "number of exams taken", which reflects the valid total score reports, so those who took the test twice will be counted as two persons.

In addition to the GMAT dataset, selecting an appropriate COVID-19 dataset is crucial for this analysis. The COVID-19 dataset sourced from "Our World in Data" adheres to stringent data collection standards and includes consistently gathered information up to March 8, 2023 [167]. There are two main reasons for selecting this database. Firstly, this dataset encompasses various metrics such as confirmed cases and fatalities, providing a comprehensive overview of

[1] https://www.gmac.com/market-intelligence-and-research/research-library/gmat-test-taker-data/profile-of-gmat-testing-residence-ty2019-ty2023

the pandemic's impact across countries. These robust metrics are essential for examining the relationship between the COVID-19 pandemic and the academic performance of GMAT examinees, ensuring a thorough and rigorous analysis. Secondly, the consistency in data collection standards up until March 8, 2023, satisfies the research requirements across various years. This ensures the analysis is accurate and relevant, and supports the validity of the research findings.

5.3.2 Data Analysis Methods

5.3.2.1 Wilcoxon Signed-Rank Test

The Wilcoxon signed-rank test is utilized to compare annual data to determine whether there are significant differences. There are several reasons for selecting the Wilcoxon signed-rank test. Firstly, the pairs in a group consist of corresponding data points; specifically, the same countries are represented in the pairs in a group. Therefore, the pairs in a group are related to each other, necessitating a comparison test for related samples. Moreover, most data in our datasets do not follow a normal distribution, which does not align with the assumptions of the paired-samples t-test [91]. In this scenario, the non-parametric Wilcoxon signed-rank test is appropriate for datasets regardless of their distributional properties [91]. The Wilcoxon signed-rank test is particularly suitable for such paired data, as it is designed to handle related samples, such as the same subjects under different conditions [91].

It is crucial to establish hypotheses for the Wilcoxon signed-rank test. Typically, these hypotheses compare the medians of paired samples to assess whether their distributions are equivalent. In this context, the hypotheses are listed (5.1) and (5.2), where m denotes the median of a pair [92].

$$H_0 : m_1 - m_2 = 0. \tag{5.1}$$

$$H_1 : m_1 - m_2 \neq 0. \tag{5.2}$$

Specifically, H_0 represents the median difference of the population between the pairs, which is equal to 0, and H_1 denotes the median difference that is not equal to 0.

To explore whether to reject the null hypothesis, the key is to calculate the W-statistic, which is the sum of ranks. Initially, the differences between the pairs are calculated, and then the absolute value $|d|$ will be ranked in ascending order [168]. When the difference is 0, it is not taken into account [169]. There will be two ranks, T^+ and T^-. For all the positive differences, T^+ is the sum of these ranks. For all negative differences, T^- is the sum of these ranks. After calculating T^+ and T^-, the lower one is selected as the W-statistic [92]. After obtaining the W-statistic in the Wilcoxon signed-rank test, the z-value is calculated to determine the corresponding p-value, and the p-value is then used to decide whether to reject the null hypothesis [92]. If N is larger than 20, then the z-value is calculated as (5.3) [170],

$$z = \frac{W - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}}, \tag{5.3}$$

where N represents the sample size within each pair in the group.

5.3.2.2 Spearman Correlation Coefficient

In this chapter, the Spearman correlation coefficient will be utilized to examine the associations among the COVID-19 pandemic, variations in the number of test-takers, and fluctuations in mean total test scores.

There are several compelling reasons for utilizing the Spearman correlation coefficient. Firstly, it is well-suited for assessing non-linear relationships [171]. Given the pandemic's potential to directly influence both the quantity of test-takers and their academic performance, assuming a linear relationship between these variables would be unrealistic. The Spearman correlation coefficient addresses this by evaluating the strength and direction of associations based on rankings rather than raw values [94], making it particularly suitable for this analysis. Another crucial rationale is that the datasets involved exhibit primarily non-normal distributions. Therefore, the Spearman correlation coefficient is appropriate as it does not require normality assumptions [94].

The formula for the Spearman correlation coefficient is expressed as (5.4) [102]:

$$r = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \tag{5.4}$$

where d_i represents the differences between ranks after sorting the variables, and n denotes the dataset size.

5.4 Experiment Results

5.4.1 Data Analysis and Reprocessing

Based on the dataset, this chapter has selected five distinct years to represent different phases of the COVID-19 pandemic for the Wilcoxon signed-rank test. TY2019 refers to the period from July 1, 2018, to June 30, 2019, predating the COVID-19 pandemic [172]. TY2020 spans from July 1, 2019, to June 30, 2020, encompassing the year of the COVID-19 pandemic outbreak [172]. TY2021 covers the period from July 1, 2020, to June 30, 2021, representing a year during the global impact of COVID-19 [173]. TY2022 extends from July 1, 2021, to June 30, 2022, highlighting another year affected by the global spread of new virus variants [173]. TY2023 includes the timeframe from July 1, 2022, to June 30, 2023, marking the period following the relaxation of pandemic restrictions in many regions [174].

To ensure consistency in the data analysis, it is crucial to select countries based on the completeness of their data in both the GMAT and COVID-19 datasets. A country will only be included in the analysis if all necessary variables are available in both datasets. This selection criterion ensures that the data is comprehensive and meets the requirements for thorough and accurate analysis. Therefore, 110 countries and regions are selected for the worldwide analysis. Also, the number of countries and regions chosen for each continent should objectively reflect each continent's situation. Around 15 countries and regions are selected in Africa, while around 34 countries and regions are selected in Asia. In Europe, around

35 countries and regions are chosen. In North America, around 13 countries and regions are selected, while in South America, around 9 countries and regions are chosen. However, Oceania was omitted from the data analysis because data were only available for 3 out of its 14 countries and regions. Nevertheless, the total number of countries in Oceania will be accounted for in the overall tally.

5.4.2 COVID-19 Impact on the Number of GMAT Examinees

5.4.2.1 Significance Analysis of Changes in the Number of Examinees

The comparison is conducted in four groups: G1 (TY2020 and TY2019), G2 (TY2021 and TY2020), G3 (TY2022 and TY2021), and G4 (TY2023 and TY2022). Each group of data will be analyzed, covering all continents as well as a worldwide perspective. Following the division of years into these four groups, the hypotheses for the total number of examinees for each group are listed as (5.5) and (5.6), where m_1^{TN} represent the median of the total number of examinees for the next test year, and m_1^{TN} represent the median of the total number of examinees for the previous test year.

$$H_0^{TN}: m_1^{TN} - m_2^{TN} = 0 \tag{5.5}$$

$$H_1^{TN}: m_1^{TN} - m_2^{TN} \neq 0 \tag{5.6}$$

Similarly, the hypotheses for the number of male examinees of each group are listed as (5.7) and (5.8).

$$H_0^{MN}: m_1^{MN} - m_2^{MN} = 0 \tag{5.7}$$

$$H_1^{MN}: m_1^{MN} - m_2^{MN} \neq 0 \tag{5.8}$$

Further, the hypotheses for the number of female examinees of each group are listed as (5.9) and (5.10).

$$H_0^{FN}: m_1^{FN} - m_2^{FN} = 0 \tag{5.9}$$

$$H_1^{FN}: m_1^{FN} - m_2^{FN} \neq 0 \tag{5.10}$$

A significance level of 0.05 is chosen to test these hypotheses, aligning with common practice in statistical analysis where the p-value is typically evaluated against this threshold [74]. A p-value below 0.05 indicates sufficient evidence to reject the null hypothesis [74]. Moreover, if the null hypothesis is rejected, the z-value provides further clarification: a positive z-value indicates that the statistics in test year 2 surpass those in test year 1, while a negative z-value suggests the opposite [170]. Therefore, by employing a significance level of 0.05, this study ensures a rigorous and standardized approach to assessing median differences between the two test years, with the z-value aiding in interpreting the direction and magnitude of these differences.

Table 5.1 and Fig.5.1 present z-values derived from comparisons between two consecutive test years within each group across continents. In the table, "*" denotes p-values less than 0.05, and "**" indicates p-values less than 0.01, following the established conventions of the Wilcoxon signed-rank test for interpreting statistical significance levels. These standards are applied to all tables in the chapter.

As depicted in Table 5.1 and Fig.5.1, a consistent trend emerges in the total number of GMAT examinees globally across the study periods. Within G1, statistical analysis reveals a significant decrease in the total number of examinees in TY2020 compared to TY2019 across all continents—Africa, Asia, Europe, North America, and South America—as well as globally, with all computed p-values being below 0.05, and notably, all p-values being less than 0.01. The negative z-values observed in G1 (-2.666 to -7.913) underscore a pronounced reduction in overall GMAT examinee numbers following the onset of the COVID-19 pandemic. In subsequent groups, changes in the total number of examinees were statistically non-significant across most continents, indicating no significant alterations in total examinee numbers after TY2020.

In addition to the findings on the total number of examinees within each continent, the trends in the number of male and female examinees mirrored the overall changes. In G1, all p-values are below 0.05, and all z-values are negative.

This indicates a statistically significant decrease in the number of both male and female examinees across all continents. This consistency underscores uniformity in gender-specific trends, aligning with the observed changes in the total number of examinees.

There are some differences in the changes observed among different continents after TY2020. In Europe and globally, the total number of examinees, as well as the number of male and female examinees, significantly decreases across all four groups. However, in other continents, significant decreases in these numbers are observed only in G4. For G2 and G3 in these continents, there are nearly no significant changes in the number of examinees.

Table 5.1 Z-Values of GMAT Examinee Numbers

Continent	Group	Total	Male	Female
World	G1	-7.913**	-7.008**	-7.928**
	G2	-2.058*	-1.476	-2.098*
	G3	-5.269**	-5.523**	-4.215**
	G4	-3.728**	-3.107**	-4.135**
Africa	G1	-3.077**	-3.181**	-2.862**
	G2	-0.210	-0.114	0.000
	G3	-0.471	-0.939	-0.142
	G4	-0.995	-0.377	-1.480
Asia	G1	-4.416**	-3.878**	-4.334**
	G2	-0.268	-1.069	-0.384
	G3	-1.282	-1.701	-0.881
	G4	-3.243**	-3.078**	-3.385**
Europe	G1	-4.481**	-3.770**	-4.508**
	G2	-4.097**	-3.740**	-3.513**
	G3	-5.501**	-5.032**	-4.800**
	G4	-0.856	-0.008	-1.109
North America	G1	-3.181**	-2.797**	-3.062**
	G2	-0.769	-0.196	-0.918
	G3	-2.385*	-2.551*	-1.140
	G4	-1.470	-1.061	-1.458
South America	G1	-2.666**	-2.429*	-2.552*
	G2	-1.007	-0.830	-0.491
	G3	-1.245	-1.542	-0.474
	G4	-1.125	-1.482	-1.364

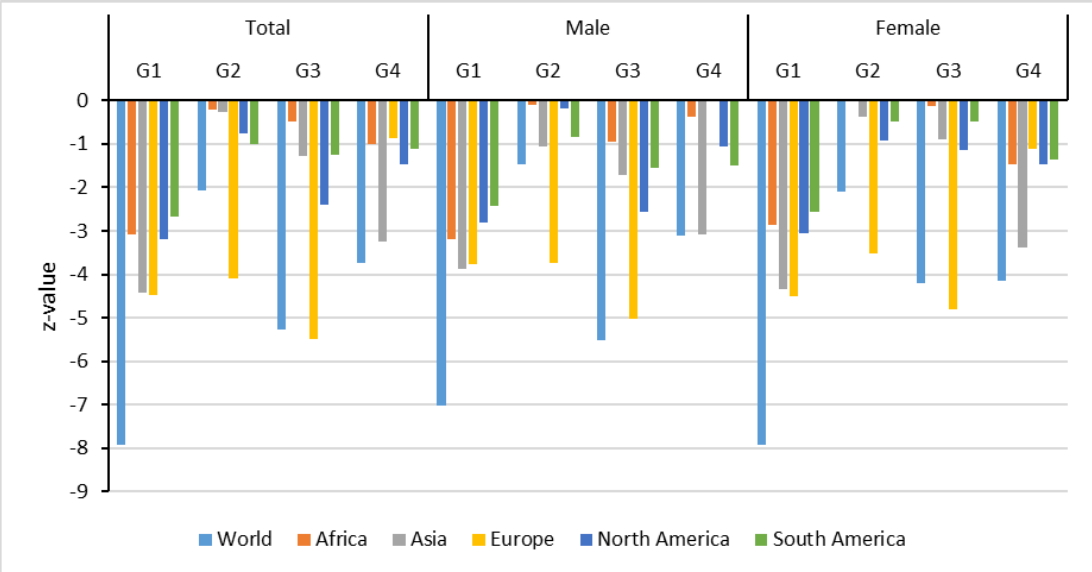


Fig. 5.1 *Z*-Values of GMAT Examinee Numbers

5.4.2.2 Correlation Analysis for the Impacts of COVID-19 on Number of Examinees

The analysis aims to identify significant correlations between COVID-19 factors, such as new deaths and new cases, and changes in the number of GMAT examinees. The analysis is conducted on three distinct groups. For each group, the total number of new cases and new deaths in each respective test year are selected as the COVID-19 variables. In each test year, the correlation analysis will examine the relationship between changes in the number of GMAT examinees and the COVID-19 variables. Table 5.2 summarizes the variables included in the correlation analysis.

Table 5.3 presents the correlation between the COVID-19 variables and the yearly GMAT examinee. As shown in Table 5.3, G1 consistently exhibits strong correlation coefficients. Across this cohort, there are predominantly negative correlations between the COVID-19 variables and the yearly GMAT examinee change, including gender breakdowns between TY2020 and TY2019. The correlation coefficients' absolute values are notably high, ranging from 0.441 to 0.867, indicating a substantial influence between COVID-19 variables and changes in

Table 5.2 Variable List: Correlation Analysis of Changes in Examinee Numbers

Group	COVID-19 Variables		Yearly GMAT Examinee Changes
G1	New cases in TY2020	New deaths in TY2020	Total
			Male
			Female
G2	New cases in TY2021	New deaths in TY2021	Total
			Male
			Female
G3	New cases in TY2022	New deaths in TY2022	Total
			Male
			Female

GMAT examinee numbers. Specifically, as COVID-19 cases and deaths increased in TY2020, the total number of examinees in TY2020 was generally lower than in TY2019 across all continents, resulting in negative correlations. The findings suggest that higher new cases and deaths were associated with fewer GMAT test participants globally, across different continents and genders.

Table 5.3 Correlation Analysis of COVID-19 Impact on Examinee Numbers

Continent	**COVID-19**	**G1**			**G2**			**G3**		
		Total	Male	Female	Total	Male	Female	Total	Male	Female
World	Cases	-0.611**	-0.616**	-0.498**	0.261**	0.290**	0.195*	-0.497**	-0.500**	-0.474**
	Deaths	-0.577**	-0.566**	-0.485**	0.211*	0.282**	0.121	-0.236*	-0.275**	-0.204*
Africa	Cases	-0.755**	-0.765**	-0.667**	0.186	0.213	0.064	-0.371	-0.469	-0.373
	Deaths	-0.733**	-0.720**	-0.730**	0.236	0.324	0.081	-0.457	-0.578*	-0.353
Asia	Cases	-0.564**	-0.617**	-0.441**	0.150	0.112	0.210	0.026	0.025	0.105
	Deaths	-0.550**	-0.592**	-0.486**	0.020	0.028	0.092	0.332	0.315	0.380*
Europe	Cases	-0.604**	-0.591**	-0.453**	0.434**	0.524**	0.325*	-0.851**	-0.784**	-0.870**
	Deaths	-0.674**	-0.612**	-0.522**	0.358*	0.469**	0.229	-0.502**	-0.454**	-0.520**
North America	Cases	-0.828**	-0.691**	-0.697**	-0.396	-0.014	-0.414	-0.465	-0.663**	-0.393
	Deaths	-0.770**	-0.724**	-0.576*	-0.335	0.173	-0.442	-0.324	-0.646**	-0.305
South America	Cases	-0.867**	-0.933**	-0.731*	0.183	0.217	-0.383	-0.460	-0.611	-0.350
	Deaths	-0.767**	-0.917**	-0.555	0.350	0.383	-0.283	-0.544	-0.669*	-0.350

Moreover, regarding gender breakdown, although there are significant correlation coefficients between COVID-19 variables and examinee changes for both male and female examinees in G1, there are differences in the values. Generally, the correlation coefficients' absolute values for male examinees are larger than those for female examinees, indicating a stronger correlation between COVID-19 variables and male examinee changes.

Conversely, G2 and G3 exhibit less pronounced results. In G2, only differences in total examinee numbers and gender-specific counts correlate with cases and deaths in Europe, with these correlations being positive. This suggests that higher COVID-19 figures in Europe were associated with changes in examinee numbers from TY2020 to TY2021. While notable correlations are observed in other continents within G3, Europe consistently demonstrates the strongest and most significant correlations across different test years.

In summary, correlation coefficients are most consistent across continents in G1, whereas in G2 and G3, correlations are less substantial across most continents globally. This underscores the profound impact of the COVID-19 pandemic on GMAT examinee numbers, particularly within the European context.

5.4.3 COVID Impact on Mean Total Scores (MTS)

5.4.3.1 Significance Analysis of Changes in MTS

The comparison is conducted on four distinct groups: G1 (TY2020 and TY2019), G2 (TY2021 and TY2020), G3 (TY2022 and TY2021), and G4 (TY2023 and TY2022). Each group will undergo a comprehensive analysis to explore various trends within the dataset.

Setting up hypotheses for the data analysis is crucial after dividing the years into these four groups. These hypotheses will guide the examination of trends and changes over time, ensuring a structured approach to understanding the data's evolution across different periods.

The hypotheses for the MTS of each group are listed (5.11) and (5.12).

$$H_0^{MTS} : m_1^{MTS} - m_2^{MTS} = 0 \tag{5.11}$$

$$H_1^{MTS} : m_1^{MTS} - m_2^{MTS} \neq 0 \tag{5.12}$$

Table 5.4 and Fig.5.2 illustrate comparisons of MTS changes among continents for each group.

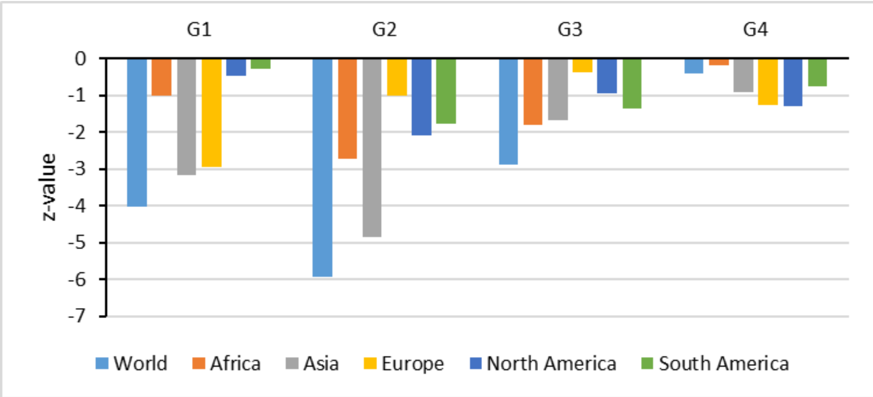


Fig. 5.2 *Z*-Values of MTS

Table 5.4 Changes in MTS

	World	Asia	Europe	North America	South America	Africa
G1	-4.016**	-3.165**	-2.952**	-0.489	-0.296	-1.005
G2	-5.942**	-4.849**	-1.019	-2.098*	-1.779	-2.727**
G3	-2.894**	-1.694	-0.369	-0.942	-1.364	-1.817
G4	-0.425	-0.924	-1.252	-1.293	-0.770	-0.199

As depicted in Table 5.4 and Fig. 5.2, in G1, substantial changes in MTS are observed worldwide, particularly notable in Asia and Europe, where the corresponding p-values are less than 0.05 and the z-values are negative. Similarly, G2 demonstrates a significant decrease worldwide in MTS, especially in Asia, Africa, and North America.

Conversely, G2 and G3 show virtually no significant differences in MTS between the two test years across any continent, as almost all p-values exceed 0.05. This consistency underscores the stability in MTS across continents during these periods.

5.4.3.2 Correlation Analysis for the COVID-19 Pandemic and Changes in MTS

Table 5.5 presents the variables used for the correlation analysis assessing the impact of the COVID-19 pandemic on changes in mean total scores.

Table 5.5 Variable List for Correlation Analysis of MTS Differences

Group	Covid-19 Variables		Yearly MTS Changes
G1	New cases in TY2020	New deaths in TY2020	Between TY2020 and TY2019
G2	New cases in TY2021	New deaths in TY2021	Between TY2021 and TY2020
G3	New cases in TY2022	New deaths in TY2022	Between TY2022 and TY2021

Table 5.6 presents the results of the correlation analysis examining the Spearman correlation coefficient between COVID-19 pandemic variables and changes in MTS in each test year. As shown in Table 5.6, no significant correlation is found between COVID-19 variables and the changes in MTS for all groups, as all p-values exceed 0.05. These findings suggest that there is no direct association between the occurrences of COVID-19 cases and deaths and the changes observed in GMAT scores across all continents.

Table 5.6 Correlation Analysis of COVID-19 Impacts on MTS

Continent	**Correlation Coefficient**	**G1-Differences**	**G2-Differences**	**G3-Differences**
World	Cases	-0.050	-0.040	0.181
	Deaths	-0.070	-0.046	0.056
Africa	Cases	-0.182	-0.177	-0.057
	Deaths	-0.171	-0.209	0.029
Asia	Cases	-0.068	0.183	0.092
	Deaths	-0.032	0.180	-0.077
Europe	Cases	-0.069	-0.007	0.252
	Deaths	-0.147	0.038	0.253
North America	Cases	0.225	0.366	0.000
	Deaths	0.110	0.322	0.019
South America	Cases	-0.117	-0.100	0.159
	Deaths	0.067	0.233	-0.008

5.4.4 Discussion

The experiments show that with the onset of the COVID-19 pandemic, all continents initially witnessed a sharp decline in the total number of examinees. This decrease was predominantly attributed to the sudden outbreak and the implementation of stringent governmental measures [175], leading to widespread exam cancellations. As restrictions gradually eased across many regions and policies reverted to pre-pandemic norms, the number of examinees stabilized.

During the early phases of the pandemic, a global negative correlation emerged between COVID-19 variables and fluctuations in GMAT examinee numbers. However, this correlation weakened after the first half of 2020. The physical and mental challenges posed by the pandemic likely impeded examinees' preparation and participation in the testing process during the initial outbreak [176], with stress levels potentially easing as the pandemic's impact lessened.

Additionally, the study highlights gender differences among examinees within each continent, particularly at the onset of the pandemic. Despite a significant correlation, COVID-19 variables, including total new cases and deaths, showed a stronger association with changes in the number of male examinees. Recent research indicating better academic performance among females in some institutions during the pandemic [177] supports the finding of a lesser negative impact of the pandemic on female test participation.

In addition to its effect on examinee numbers, significant changes in MTS were observed in certain continents during the pandemic's initial two years. These changes likely stemmed from psychological stress and challenges in adapting to new online testing platforms [10]. However, as the pandemic stabilized, these score variations became less pronounced.

Interestingly, no direct correlation was found between new COVID-19 cases/deaths and changes in mean total scores. This contrasts with the observed significant correlation between COVID-19 variables and GMAT examinee numbers, suggesting that factors beyond mere infection rates influenced academic performance, such as psychological stress and digital accessibility challenges [178].

Despite these valuable insights, the chapter is subject to several limitations. While the study incorporates "Other" gender categories in the total counts of examinees and scores, the original dataset lacks specific data for the "Other" gender category. Furthermore, the original data summarizing the number of valid reports may introduce inaccuracies due to potential duplication of examinee numbers.

5.5 Conclusion

In summary, this chapter examines the impact of the COVID-19 pandemic on GMAT examinee numbers and academic performance across different phases of the crisis. The analysis reveals significant fluctuations in the number of examinees and their mean total scores during various periods, highlighting substantial changes in test-taking behavior and academic outcomes amid the pandemic. Additionally, the study identifies a notable correlation between COVID-19 metrics—such as cases and deaths—and the variations in examinee numbers across continents at the onset of the pandemic. This correlation underscores the intricate relationship between external health crises and academic testing behaviors, offering valuable insights into how prospective examinees adapt during global uncertainty.

This chapter concludes by recommending several avenues for future research. It advocates for using more precise databases that incorporate updated criteria such as gender and attendance, aiming to refine our understanding of how socio-demographic factors influence test participation and performance under extraordinary circumstances. Furthermore, expanding the analysis to encompass global tests and a broader array of COVID-19-related variables is essential for enhancing the robustness and generalizability of findings.

By extending beyond regional impacts and incorporating diverse pandemic factors, future studies can better elucidate the nuanced effects of such crises on academic assessments worldwide. These insights are crucial for informing more resilient educational policies and practices, ensuring educational systems can adapt effectively to global health crises.

Chapter 6

Assessing the Impacts of COVID-19 on Early Learning Services: A Study Based on Attendance Rate in New Zealand

6.1 Introduction

COVID-19 resulted in a great pandemic in the majority of countries [179]. Among the sectors affected by the pandemic, educational services witnessed notable setbacks [180]. A common response involved numerous educational institutions transitioning to online teaching [181], which impacted the traditional learning process [57, 165]. Also, the worldwide tests have been affected by the pandemic. For example, the average TOEFL test results among students worldwide have greatly changed at the beginning of the pandemic [158].

In addition to its impact on education, the COVID-19 pandemic significantly influenced children's health outcomes. For instance, the research from England highlighted disruptions in children's daily routines, including irregular diet and sleep patterns [182]. Furthermore, a study in Shanghai, China, identified a notable increase in emotional and behavioral problems among children during the pandemic [183]. These findings emphasize the pandemic's broad and multifaceted effects on children's physical, emotional, and behavioral well-being.

Despite extensive research on the pandemic's effects on global education and children's health outcomes, its detrimental impacts on early childhood education need to be explored. Early childhood education is a critical period in a child's development, as many lifelong structures and functions are shaped during this time [184]. Furthermore, research indicates that a child's early experiences significantly influence their developmental trajectory [185]. Consequently, early childhood education plays a foundational role in shaping a child's future academic success, extending through primary and secondary education and into higher education. However, the COVID-19 pandemic placed immense strain on the operations of childcare service centers, exacerbating challenges in this crucial sector [186].

Another critical reason to prioritize research on early childhood education is the significant challenges it faced during the COVID-19 pandemic. Early learning services, a key component of early childhood education, include education and care services, kindergartens, and play centers[1]. For instance, a recent survey in the United States revealed that the pandemic severely disrupted the regular functioning of these services [187]. Furthermore, additional research highlighted a substantial decline in enrollment at early learning service centers during this period [188]. These findings underscore the pressing need to investigate the pandemic's impact on early learning services to mitigate long-term consequences.

Although experts emphasize the importance of early childhood education and numerous studies have examined the impacts of early learning services, significant gaps remain in understanding how early childcare services have been affected. Existing research lacks a comprehensive analysis of the multifaceted factors related to the COVID-19 pandemic, such as the role of case numbers and vaccination rates in influencing early learning services [189, 190, 191]. Moreover, these studies often fail to investigate the varying effects of the pandemic on different ethnic groups within the same country [189, 190, 191]. Examining these differences is crucial, as ethnic groups within the same country possess distinct traditions and cultural backgrounds, which may lead to diverse experiences and impacts during the same pandemic under the same conditions. Understanding these variations

[1] https://www.educationcounts.govt.nz/__data/assets/pdf_file/0005/203648/COVID-19-Attendance-FAQs.pdf

is essential for developing policies tailored to each group's unique needs and circumstances, ensuring equitable support and effective interventions.

To address these gaps and examine the impacts of the COVID-19 pandemic on childcare services, it is essential to conduct a critical analysis of how the pandemic has influenced early learning services, particularly how it has altered children's vulnerabilities to attending centers during this period. Attendance rates offer a more relevant measure for early learning services. Unlike other performance metrics, attendance rates are not confined to specific educational levels and are summarized at an aggregate level. Consequently, some researchers have proposed attendance rates as a potential indicator for assessing the dynamics of the early learning services sector during the pandemic [192].

The research question is formulated as follows "How does the COVID-19 incidence and vaccination uptake influence the attendance rate of Early Learning Services in New Zealand". Based on the research question, two hypotheses are set up:

1. The COVID-19 incidence negatively influences the early learning services attendance rate.

2. The vaccination uptake positively influences the early learning services attendance rate.

The chapter's contribution primarily encompasses three key aspects. First, the comprehensive data analysis of datasets from different years reveals significant variations in the impact of new pandemic cases over time. Second, the chapter examines multiple variables, including case numbers and vaccination rates, across different ethnicities, highlighting that the pandemic's impact is more pronounced in certain ethnic groups.

The chapter is divided into several sections. Section 2 summarizes the gaps in the existing research studies. Section 3 outlines the data collection methods and analysis techniques adopted in this chapter. Section 4 presents the experimental results and indicates the implications of the findings. Finally, Section 6 concludes the chapter and guides future directions.

6.2 Literature Review

Numerous academic studies have revealed the adverse effects of the COVID-19 pandemic on attendance rates in early learning services. For example, a recent study in North Carolina reported a substantial decline in childcare enrollment during the pandemic [189]. Similarly, research conducted in Australia observed a significant decrease in early childcare attendance following a COVID-19 outbreak in the Footscray district of Victoria [190]. In Turkey, a study on early childhood education documented a notable reduction in classroom attendance, as students faced challenges adapting to new protective measures in schools [191]. While these findings provide evidence of declining attendance rates during the pandemic, they do not fully address the influence of other factors, such as COVID-19 case numbers, deaths, and vaccination rates.

In addition, there are great disparities in the influence of the COVID-19 pandemic on multiple ethnic groups of children in various countries. For example, a comprehensive population-based study in England highlighted substantial disparities in the outcomes of COVID-19 among children aged 0 to 18 from various ethnic groups [193]. Similarly, the Bradford study revealed notable differences in the frequency of physical activities among children of multiple ethnicities [194]. In the United States, studies have documented considerable variations in the incidence of multisystemic inflammatory syndrome among children aged 5 to 13 years in different ethnic groups [195]. Despite these findings, a notable gap exists in the literature, as most studies have not specifically examined the ethnic disparities among preschool-aged children.

The impact of vaccinations on children, alongside the broader effects of the COVID-19 pandemic, remains a contentious issue. A study assessing the mRNA-1273 vaccine for children under six years of age concluded that it is safe, with no significant negative effects on children's daily attendance [196]. In contrast, research exploring parental concerns about children's vaccinations suggests that perceived risks may influence children's attendance patterns [197]. Despite these findings, a notable gap in the literature is the lack of investigation into the direct effects of the pandemic on the daily routines of preschool-aged children. This chapter addresses this gap by examining the pandemic's impact on children from diverse ethnic backgrounds within a single country.

Building on previous findings, it is crucial to consider the diverse research models used to analyze the effects of the pandemic on children and related services. For example, a differences-in-differences model was employed to examine childcare enrollment rates during the COVID-19 pandemic in Northern California [189], while hierarchical multiple linear regression models were used to assess the negative impacts on children in the Netherlands [198]. Additionally, a study on the impact of the pandemic on children from different ethnic backgrounds in England utilized logistic regression analysis [193]. Given the distinct limitations of each regression model, there is potential for bias in the results.

Addressing these gaps is essential for a comprehensive understanding of the relationship between the pandemic, vaccination, and attendance in educational services. This chapter explores multiple pandemic-related factors to fill these gaps. Additionally, analyzing multiple ethnic groups within a single country will highlight both the similarities and disparities among children from different backgrounds. To further enhance the analysis, this chapter utilizes various machine learning methods for regression analysis on the same dataset to identify consistent findings.

6.3 Materials

6.3.1 Data Collection

In this chapter, the attendance rate dataset originates from New Zealand's early learning services reports sourced from the Ministry of Education (the Ministry), New Zealand Government[1]. Metrics for calculating attendance rates remained consistent from October 2020 to September 2023, ensuring continuity. The dataset includes national attendance rates and rates among four ethnicities: European/Pākehā, Māori, Pacific, and Asian, facilitating a comprehensive analysis of pandemic impacts. Throughout periods of consistent metrics, the daily attendance rate for early learning services is calculated as (6.1).

[1] https://www.educationcounts.govt.nz/statistics/attendance-under-covid-19

$$\text{Attendance Rate} = \frac{\text{Number of Children Present}}{\text{Total Expected Attendance}} \tag{6.1}$$

The total expected attendance is calculated as the sum of the children present and absent on a specific day, indicating the total number of children scheduled to attend[1]. The services cover various early childhood education services such as education and care, kindergarten, and play centers. The ethnicity data adopted in this study were sourced from the New Zealand Government Ministry of Education, and the collection and processing of the data are described in their public documentation[2]. A self-identification approach is used for ethnicity classification, allowing students to report multiple ethnicities. Since some individuals may identify with more than one ethnic group, this could influence the accuracy of ethnicity-specific findings. These factors should be taken into account when interpreting the results.

Another limitation is that the Ministry only compiles statistics submitted through the Early Learning Information (ELI) system. As a result, the total number of early learning services included in the data summary and the response rates varied from week to week. This variation in the scope of respondents across weeks may affect the consistency and comparability of the data.

The database includes information on expected and actual attendance rates, along with corresponding percentages. While the Ministry collected student identifiers, such as names, genders, and student IDs, this information has not been disclosed publicly. Furthermore, the Ministry has obtained the necessary approvals to conduct the data collection, ensuring that all procedures comply with relevant laws and regulations.

Apart from early learning services attendance rates in New Zealand, it is important to select reliable COVID-19 cases and vaccination datasets, as these cases are a key reflection of the seriousness of the disease, and vaccination is an important measure to reduce the infection rate. The data are sourced from the "Our World in Data" COVID-19 dataset [167], covering the period from 2020 to 2023. Beyond the time range, the dataset provides comprehensive daily records

[1] https://www.educationcounts.govt.nz/__data/assets/pdf_file/0005/203648/COVID-19-Attendance-FAQs.pdf

[2] https://www.educationcounts.govt.nz/data-services/national/attendance

of various aspects of the COVID-19 pandemic in New Zealand, including "Total Cases," "New Cases," "Total Vaccinations," "People Vaccinated," "People Fully Vaccinated," and "New Vaccinations," thus meeting the data analysis requirements of the chapter.

6.3.2 Data Processing and Visualization

In the initial stages, we focus on comparing the impacts of COVID-19 cases and vaccinations on Early learning service attendance rates for the years 2021 and 2022. While acknowledging the COVID-19 outbreak in New Zealand in 2020[1], we exclude this year from our analysis for several reasons. First, the new attendance rate calculation metrics introduced in September 2020 limited our dataset to only 44 samples spanning a few months, hindering a comprehensive overview of attendance rate fluctuations throughout the year. Second, the initiation of vaccination promotion in New Zealand in February 2021[2] renders vaccination data for 2020 unavailable. In contrast, both 2021 and 2022 provide consistent application of new calculation metrics and the initiation of vaccination promotion from the beginning of this period, ensuring comprehensive and complete records of vaccination data. Therefore, our decision to analyze data from 2021 and 2022 maintains research consistency, while these considerations justify excluding 2020 from our analysis.

In the chapter, different letters are used to represent students from various ethnic groups. "N" denotes the nationwide students, while "A" represents Asian students, and "E/P" stands for European/Pākehā students. "M" represents Māori students, and "P" signifies Pacific students.

6.3.2.1 New Cases and Attendance Rates in New Zealand

As illustrated in Fig. 6.1, the number of new COVID-19 cases in New Zealand in 2021 remained stable from late February to late August, consistently hovering around zero. In September, however, there was a sharp increase, with new cases

[1] https://www.brookings.edu/articles/policy-and-institutional-responses-to-covid-19-new-z
[2] https://www.beehive.govt.nz/release/first-batch-covid-19-vaccine-arrives-nz

spiking from zero to 100. Despite subsequent fluctuations, the overall trend indicated a gradual rise in case numbers following this September spike. In November, the number of new cases further increased, reaching 200.

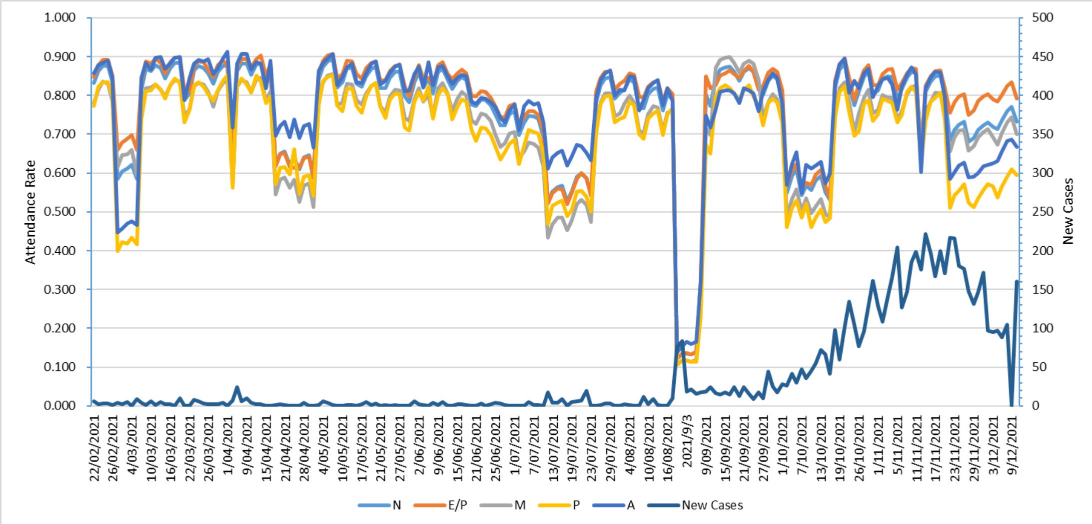


Fig. 6.1 New Cases and Attendance Rates in 2021

In contrast to the clear patterns observed in new cases, attendance rates exhibited greater variability across different ethnic groups, with frequent fluctuations throughout the year. Before September, significant decreases in attendance rates were observed in early March, late April, and late July, though these were followed by recoveries to normal levels. For most other periods, attendance rates stabilized within a range of 0.600 to 0.900. However, in late August, the attendance rate plummeted sharply to 0.100, followed by a rapid spike a few days later. A similar pattern was observed again in October.

The trends in new cases and attendance rates suggest a correlation between spikes in new cases and significant declines in attendance across various ethnic groups. The sharp decrease in attendance rates observed in August and September is likely associated with the substantial rise in new cases during the same period. This underscores the adverse impact of COVID-19 on attendance in New Zealand's early learning services in 2021.

As illustrated in Fig. 6.2, the changes in new cases and attendance rates in 2022 reveal significant fluctuations in new cases during the first nine months of the year, exceeding those observed during the same period in 2021. Between February and April, the number of new cases varied dramatically, ranging from 0 to 40,000. Subsequently, the numbers gradually declined between late April and June. In July, fluctuations re-emerged, but by late September, the case numbers had stabilized.

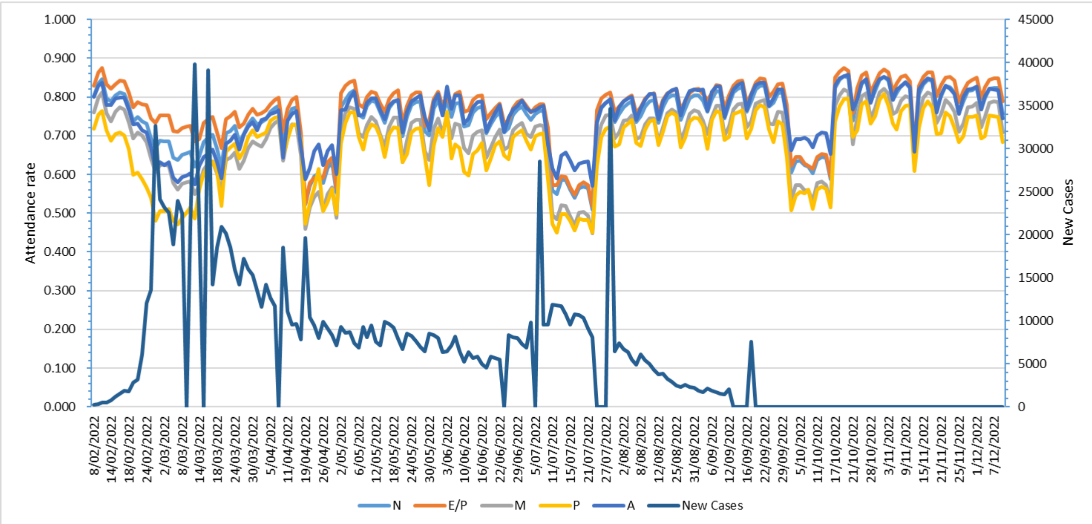


Fig. 6.2 New Cases and Attendance Rates in 2022

Similarly, attendance rates among the four ethnic groups, as well as the national attendance rate, exhibited more pronounced fluctuations in 2022 compared to 2021. While significant changes were observed during the first nine months of the year, October showed comparatively more stable attendance rates. Between February and March, a sharp decrease in the attendance rates of each ethnic group occurred in March, followed by a marked increase. Similar patterns were also observed in late April, early July, and mid-October.

From the graph, it is evident that spikes in new cases coincide with declines in attendance rates, particularly during the peaks in March, April, July, and August. As new cases declined, the fluctuations in attendance rates also diminished later in 2022.

In summary, the data from 2021 and 2022 indicate a potential inverse relationship between new case numbers and attendance rates at specific time points.

6.3.2.2 New Vaccinations and Attendance Rates in New Zealand

In 2021, as illustrated in Fig.6.3, the number of new vaccinations fluctuated, generally showing an upward trend until September, with figures ranging from 0 to 80,000. After peaking in September, a marked decrease was observed, followed by another increase in October. This was succeeded by a subsequent decline after reaching a second peak in November.

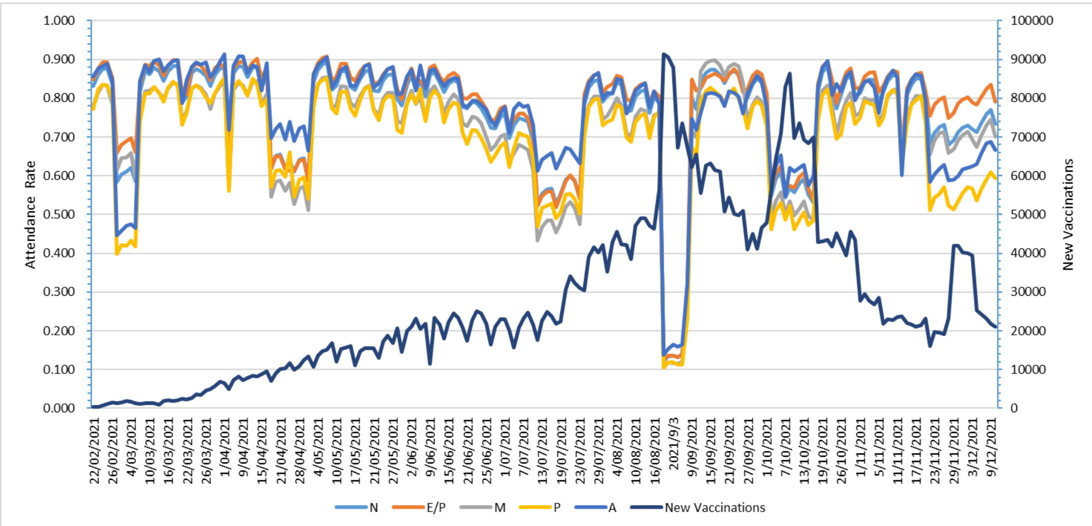


Fig. 6.3 New Vaccinations and Attendance Rates in 2021

The number of new vaccinations fluctuated, generally showing an upward trend until September, with values ranging from 0 to 80,000. After peaking in September, there was a marked decrease, followed by another increase in October. This was subsequently followed by a decline after reaching another peak in November.

In 2022, as depicted in Fig.6.4, the number of new vaccinations exhibited a stable trend for most of the year. However, there was a significant decline in vaccinations before March. Despite some fluctuations in June and July, the number of new vaccinations subsequently stabilized at a consistently low level of around 0.

The graph also illustrates changes in attendance rates. Despite some fluctuations, the attendance rates remained relatively stable, although a notable decrease

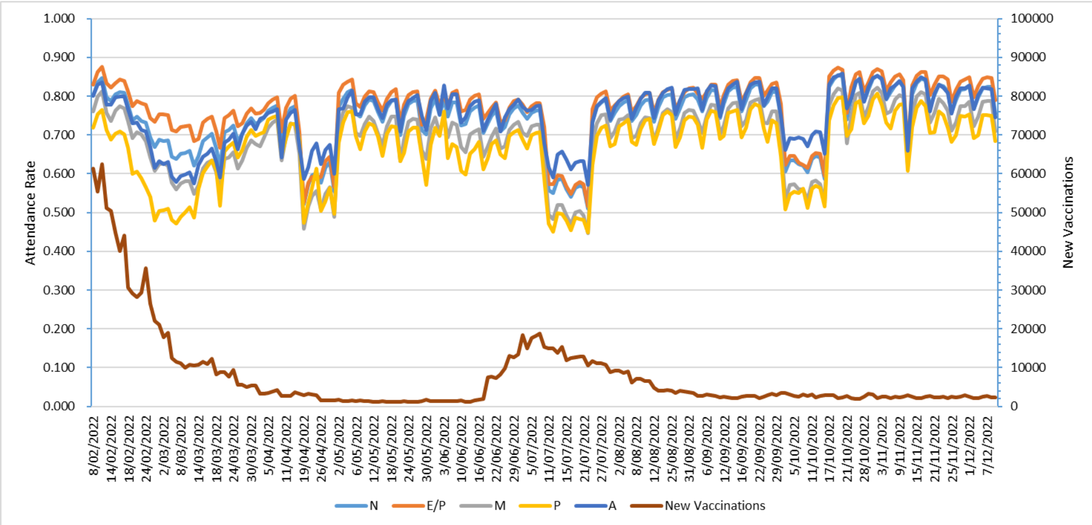


Fig. 6.4 New Vaccinations and Attendance Rates in 2022

was observed across the four ethnic groups when new vaccinations increased in June, followed by another increase in early May. A similar phenomenon occurred in July and December. After July, while still variable, attendance rates generally stabilized.

In summary, there appears to be an inverse trend between new vaccinations and attendance rates throughout 2021 and 2022.

6.4 Methodology

The chapter utilizes a quantitative methodology. This approach offers a more objective depiction based on databases, elucidating the relationships among various variables.

6.4.1 Data Analysis Method

To assess the proposed hypothesis thoroughly, we utilize two statistical methods: correlation coefficient analysis and various machine learning techniques for regression analysis.

6.4.1.1 Spearman Correlation Coefficient

The correlation coefficient is frequently utilized to analyze the strength and direction of the association between two variables [94].

The chapter utilizes the Spearman correlation to investigate the associations among new cases, vaccination doses, and attendance rates. While both Pearson and Spearman correlations measure correlation coefficients, they serve different purposes: Pearson for linear relationships and Spearman for monotonic relationships [94]. We chose the Spearman correlation for two main reasons: First, real-world scenarios often involve non-linear relationships, such as the relationship between new cases and attendance rates in our study. Second, our dataset comprises non-continuous data, such as integer-based new case numbers and vaccination doses, aligning well with Spearman correlation analysis [102]. Thus, Spearman correlation is a more suitable approach for the data analysis.

6.4.1.2 Models for Regression Analysis

Considering the inherent limitations of each machine-learning method, this chapter utilizes a range of models for regression analysis. Through this approach, we expect consistent results across different models to enhance the credibility of our conclusion. Each method has its advantages, but it also has limitations. Overall, the chapter focuses on the impacts of COVID-19 cases and vaccinations, so the influence of demographic factors, including ethnicity distribution and restriction policies, is not listed as a variable. Still, they will be analyzed in the findings.

A brief introduction to adopted models is listed below:

- MLR. It is appropriate for examining the influence of various factors on a single dependent variable. However, cases and vaccinations may not directly affect attendance rates. As a result, predictions of attendance rates based on this method may differ significantly from the actual outcomes.
- SVR. It aims to find a function that approximates the relationship between input features and continuous target values, which is appropriate for the high-dimensional spaces, the multiple independent variables in this chapter

[115]. However, the prediction performance is not efficient if there are great overlaps in targeted datasets [115], which is highly probable in this chapter.

- KNN. The KNN algorithm can be utilized for regression analysis and classification [115]. KNN selects the nearest K neighbors to identify the closest data points [115]. However, despite its simplicity, the KNN algorithm is prone to overfitting or discontinuity in model fitting [117]. In this chapter, the large variance in the range of attendance rates, new cases, and new vaccinations may affect the accuracy of predictions, making it challenging to select an appropriate K value.

- GPR. GPR is a tool for non-parametric analysis, capable of accommodating both linear and non-linear data structures, which enables it to be appropriate for addressing complex regression tasks [118]. The chart illustrates significant differences in the trends of changes for each variable, enabling the Gaussian process to be suitable for predictions based on multiple variables in this chapter.

- PLS. PLS is commonly adopted for comparing different response and explanatory variables [199]. The corresponding regression model is suitable for scenarios involving multiple interrelated predictor variables [199], while some variables may overlap in this chapter. However, the limitation of the method is that it is sensitive to some extreme values[200]. There are some sudden changes in the attendance rates and new cases, which may influence the accuracy of the setup of the model.

- MLP. It is skilled at handling complex datasets and capturing non-linear relationships [201]. However, overfitting is a common issue with this model [202].

6.4.1.3 Evaluation Metrics

In statistical modeling, discrepancies often arise between predictions made by regression models and actual data. Assessing the magnitude of these differences typically relies on specific metrics and smaller errors. Generally, lower errors indicate a better model fit, suggesting a stronger association among the variables.

In this chapter, MAPE is utilized as the principal metric for evaluating prediction accuracy. MAPE facilitates a comprehensive assessment for forecast accuracy [131], thereby establishing a robust and interpretable measure of model performance in alignment with rigorous academic standards. The formula of MAPE is shown as Equation (6.2) [131].

$$\text{MAPE} = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{A_t - F_t}{A_t}\right|, \tag{6.2}$$

where N denotes the sample size of the dataset, A_t represents the actual values and F_t represents the predicted values based on the corresponding independent variables [131].

The formula aims to calculate the proportion of differences between predicted and actual values, which is pertinent for the data analysis process to offer a consistent and objective measure of discrepancies. Metrics such as MAE, MSE, and RMSE may not adequately indicate the importance of changes as they solely compare absolute differences between predictions and actual values without considering data magnitude [203]. In comparison, the percentage difference can more objectively reflect the magnitude of changes.

6.5 Experimental Results

As illustrated in Table 6.1, the collected data are organized as independent variables and dependent variables, some of which are adopted for both correlation analysis and regression analysis. Based on the selection of available data, our analysis focuses exclusively on dates where complete data for all variables are available, yielding 194 dates from 2021 and 207 from 2022 for examination.

6.5.1 The Impact of the COVID-19 Cases

6.5.1.1 The Correlation Between New Cases and Attendance Rates

In statistical analysis, asterisks denote the significance level of a specific value. "*" typically signifies a significance level of less than 0.05, while "**" indicates a

Table 6.1 Variable List for the Correlation and Regression Analysis

Name	Description	Variable Type	Correlation Analysis	Regression Analysis
Total Cases	The accumulated COVID-19 cases on a specific day, including probable cases	Independent	✓	
New Cases	The number of COVID-19 cases on a specific day, including probable cases	Independent	✓	✓
Total Vaccinations	The accumulated number of vaccination dosages on a specific day	Independent	✓	
People Vaccinated	The accumulated number of people who get at least one vaccination dosage on a specific day	Independent	✓	
People Fully Vaccinated	The accumulated number of people who get all dosages recommended by the prescription on a specific day	Independent	✓	
New Vaccinations	The number of the new vaccination dosages on a specific day	Independent	✓	✓
Attendance Rate	N(National), A(Asian), E/P(European/Pākehā), M(Māori), P(Pacific)	Dependent	✓	✓

significance level of less than 0.01. In this study, a significance level below 0.05 suggests the presence of a correlation between the two variables.

Table 6.2 reveals significant disparities in correlation coefficients between new COVID-19 cases and attendance rates across ethnicities in New Zealand. In 2021, prominent negative correlations were observed among individuals of Asian and Pacific ethnicities, with p-values below 5%, indicating statistically significant associations. Conversely, correlations for other ethnicities, including the national average, were negligible. The findings indicate an inverse relationship between new COVID-19 cases and attendance rates, particularly pronounced among Asians and Pacific Islanders.

Table 6.2 The Correlation Coefficient Between New Cases and Attendance Rates

Ethnicity	2021	2022
P	-0.147*	-0.464**
E/P	-0.077	-0.519**
M	-0.058	-0.524**
N	-0.096	-0.525**
A	-0.232**	-0.567**

*p<0.05,**p<0.01

In contrast, in 2022, all ethnicities displayed significant negative correlations between new COVID-19 cases and attendance rates. Table 6.2 shows a consistent pattern with p-values below 5% for all ethnicities and the national average, signifying a uniform negative correlation. The results suggest a heightened association compared to 2021, with a more pronounced decline in attendance rates as the number of new cases escalated across all demographics in 2022.

Despite the coherence of correlation coefficients, slight variations in values can be observed. Notably, the absolute values of correlation coefficients rank from lowest to highest as follows: Asian, National, Māori, European/Pākehā, and Pacific.

Therefore, the findings from Table 6.2 underscore a notable increase in the correlation between new COVID-19 cases and attendance rates throughout New Zealand in 2022 compared to 2021. Specifically, in 2021, correlations between new cases and attendance rates were primarily evident among specific ethnic groups, with correlation coefficients notably lower than those observed in 2022. However, in 2022, all coefficient values are statistically significant, indicating a substantial correlation between new cases and attendance rates across the ethnic groups.

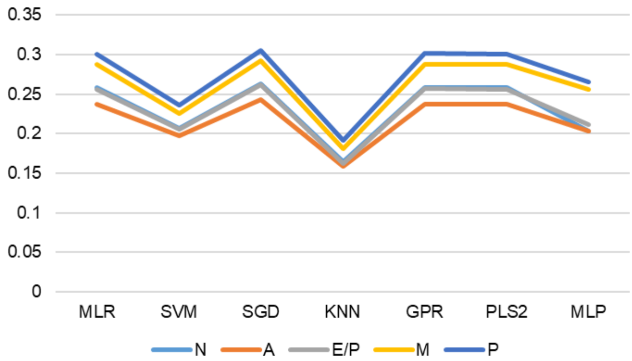
6.5.1.2 Regression Analysis of Case Effects on Attendance Rates

Table 6.3 and Fig.6.5 show that across the seven regression models using new cases and total cases, MAPE values for 2021 exceed those for 2022. In 2021, most MAPEs are above 0.20, while in 2022, all MAPEs are below 0.12. This suggests a significant improvement in attendance rate prediction accuracy for four ethnicities and the national average in New Zealand in 2022 compared to 2021, likely due to better fitting of the dataset by regression models using both “Total Cases” and “New Cases” variables.

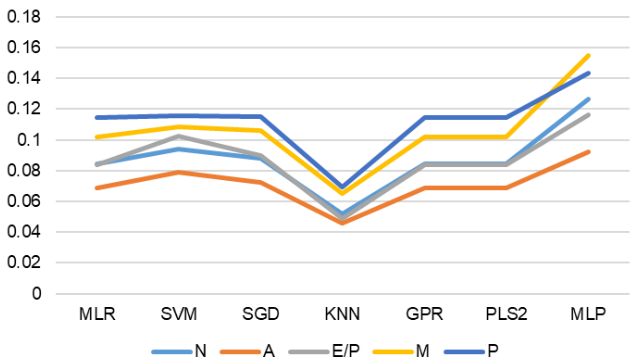
Table 6.3 MAPE of the Attendance Rates Based on Total Cases and New Cases

	N		A		E/P		M		P	
	2021	2022	2021	2022	2021	2022	2021	2022	2021	2022
MLR	0.2587	0.0844	0.2373	0.0684	0.2565	0.0836	0.2879	0.1018	0.3009	0.1144
SVR	0.2072	0.0938	0.1978	0.0790	0.2056	0.1027	0.2258	0.1087	0.2363	0.1158
SGD	0.2634	0.0879	0.2428	0.0723	0.2615	0.0895	0.2921	0.1061	0.3049	0.1153
KNN	0.1651	0.0521	0.1584	0.0461	0.1624	0.0487	0.1813	0.0652	0.1910	0.0692
GPR	0.2588	0.0844	0.2374	0.0684	0.2566	0.0836	0.2879	0.1018	0.3010	0.1144
PLS	0.2587	0.0844	0.2373	0.0684	0.2565	0.0836	0.2879	0.1018	0.3009	0.1144
MLP	0.2031	0.1268	0.2034	0.0922	0.2114	0.1165	0.2561	0.1545	0.2654	0.1432

At the same time, the lowest MAPE values are consistently observed for the Asian group compared to other ethnic groups across all regression models in both 2021 and 2022, which indicates a closer association between COVID-19 case numbers and attendance rates among Asians. For example, in 2021, the MAPE for Asian ethnicity under MLR was 0.2373, which is lower than that of the European/Pākehā (0.2565), Māori (0.2879), and Pacific (0.3009) ethnicities. Meanwhile, the value for Asian ethnicity under SVR is 0.1978, which is lower than those of the European/Pākehā (0.2056), Māori (0.2258), and Pacific (0.2363) ethnicities. In 2022, the MAPE for the Asian ethnicity under MLR was only 0.0684,



6.5(a) MAPEs in 2021



6.5(b) MAPEs in 2022

Fig. 6.5 MAPEs of Attendance Rates Based on New Cases and Total Cases for 2021 and 2022 significantly lower than those of the European/Pākehā (0.0836), Māori (0.1018), and Pacific (0.1144) ethnicities. Similarly, under SVR, the MAPE for the Asian ethnicity was 0.0790, again much lower than the corresponding values for European/Pākehā (0.1027), Māori (0.1087), and Pacific (0.1158) ethnicities. Furthermore, using the MLP model, the MAPE for the Asian ethnicity was 0.0922, notably lower than that of the European/Pākehā (0.1165), Māori (0.1545), and Pacific (0.1432) ethnicities. This result echoes previous correlation analysis that the COVID-19 case in Asian ethnicity correlates with the attendance rate than other ethnicities do in 2021 and 2022.

The MAPE analysis also reveals a significant increase in the impact of COVID-19 cases from 2021 to 2022. In previous correlation analysis, there was minimal correlation between COVID-19 cases and attendance rates across ethnicities in 2021, with a significant increase in 2022. This observed trend is further supported by the MAPE results, which exhibit significantly larger values in 2022 compared to 2021 for each regression model.

6.5.2 The Impact of COVID-19 Vaccinations

6.5.2.1 Correlation Between New Vaccinations and Attendance Rates

Table 6.4 consistently displays significant negative correlation coefficients between new vaccination doses and attendance rates across all ethnicities and the national attendance rate in New Zealand for both 2021 and 2022, with p-values below 5%. This indicates that as new vaccinations increase, attendance rates decrease across all ethnicities and at the national level in both years. The magnitude of the correlation coefficients experiences only slight decreases from 2021 to 2022, with no significant changes in direction or ranking observed. Notably, in both years, the Asian ethnicity exhibits the largest coefficient magnitude, indicating the strongest negative correlation between new vaccination and attendance rate.

6.5.2.2 Regression Analysis of Vaccination Effects on Attendance Rates

As depicted in Table 6.5 and Fig.6.6, MAPEs across all regression models witnessed a significant decrease in 2022 for each ethnic group and the whole country. In 2021, the MAPEs ranged from 0.12 to 0.23; in 2022, all values were less than

Table 6.4 Correlation Coefficient Between New Vaccinations and Attendance Rates

Ethnicity	2021	2022
P	-0.331**	-0.312**
E/P	-0.350**	-0.297**
M	-0.247**	-0.271**
N	-0.319**	-0.303**
A	-0.429**	-0.357**

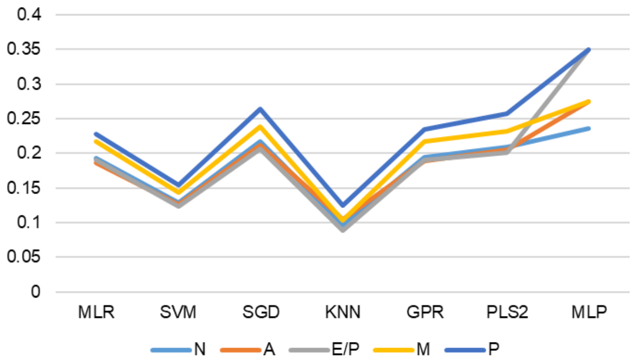
*p<0.05,**p<0.01

0.11. This indicates an improved predictive accuracy in 2022, reflecting enhanced model fitting to the dataset. Consequently, a strengthened association between vaccinations and early learning services attendance rates during 2022 can be inferred.

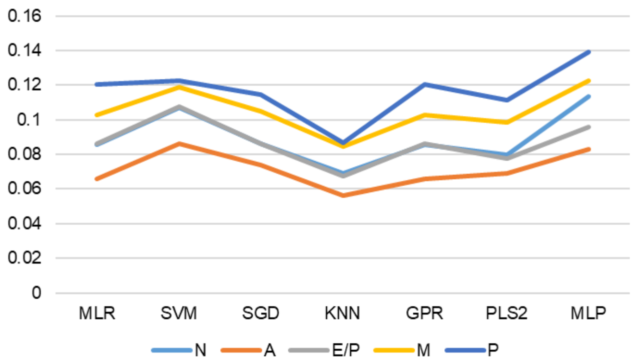
Table 6.5 MAPE Values of Attendance Rates Based on Vaccination Variables

	N		A		E/P		M		P	
	2021	2022	2021	2022	2021	2022	2021	2022	2021	2022
MLR	0.1930	0.0855	0.1867	0.0661	0.1900	0.0860	0.2173	0.1030	0.2280	0.1205
SVR	0.1293	0.1069	0.1258	0.0861	0.1231	0.1077	0.1430	0.1189	0.1544	0.1224
SGD	0.2165	0.0863	0.2119	0.0740	0.2058	0.0861	0.2381	0.1051	0.2642	0.1145
KNN	0.0960	0.0690	0.1043	0.0564	0.0885	0.0672	0.1039	0.0845	0.1254	0.0869
GPR	0.1939	0.0855	0.1890	0.0661	0.1900	0.0860	0.2174	0.1030	0.2339	0.1205
PLS	0.2088	0.0796	0.2049	0.0693	0.2015	0.0778	0.2324	0.0985	0.2575	0.1104
MLP	0.2359	0.1135	0.2752	0.0830	0.3492	0.0958	0.2747	0.1226	0.3499	0.1391

In both 2021 and 2022, consistently low MAPEs were observed for the Asian ethnicity, and in particular, the lowest MAPEs in 2022 were consistently associated with the Asian group, followed by the European/Pākehā, Māori, and Pacific ethnicities, which indicates a closer association between multiple vaccination variables and attendance rates among Asians. In 2022, the MAPE for the Asian ethnicity under MLR was only 0.0661, significantly lower than those of the European/Pākehā (0.0860), Māori (0.1030), and Pacific (0.1205) ethnicities. Similarly, under SVR, the MAPE for the Asian ethnicity was 0.0861, again



6.6(a) MAPEs in 2021



6.6(b) MAPEs in 2022

Fig. 6.6 MAPEs of Attendance Rates Based on Vaccination Variables for 2021 and 2022 much lower than the corresponding values for European/Pākehā (0.1077), Māori (0.1189), and Pacific (0.1224) ethnicities. Furthermore, using the MLP model, the MAPE for the Asian ethnicity was 0.0830, notably lower than that of the European/Pākehā (0.0958), Māori (0.1226), and Pacific (0.1391) ethnicities. This underscores an intensified association between vaccination dosage and attendance rates within this demographic. This trend aligns with correlation coefficient outcomes, where the Asian ethnicity exhibits the largest absolute value, indicating heightened susceptibility to vaccination effects. Overall, these results suggest that children of Asian ethnicity in New Zealand have consistently displayed the most substantial responsiveness to vaccination effect over the two years.

6.5.3 Discussion

The findings reveal a significant shift in the relationship between COVID-19 cases and attendance rates across all ethnic groups in New Zealand in 2022, in contrast to 2021, when this association was limited to specific ethnicities. These results support the hypothesis of a negative correlation between COVID-19 cases and attendance rates in 2022. The discrepancy between the two years can be attributed to the low case count in 2021 and the relaxation of restrictions in 2022. In 2021, restrictions were stringent, with a level 4 alert[1], while in 2022, measures were eased, allowing children more opportunities for face-to-face interactions, while the increased interaction likely contributed to higher infection rates and subsequent absenteeism.

Similarly, a consistent negative correlation was observed between new vaccination rates and attendance across all ethnicities in both 2021 and 2022. Despite the intended benefits of vaccinations, empirical evidence suggests that vaccination may negatively impact attendance rates. One plausible explanation is the occurrence of post-vaccination symptoms, particularly fatigue, which could contribute to increased absenteeism, especially among individuals with immune-related conditions [204].

A significant finding highlights the marked impact of the COVID-19 pandemic on children of Asian ethnicity in New Zealand. Analysis reveals a strong correla-

[1] https://www.educationcounts.govt.nz/statistics/attendance-under-covid-19

tion between both case numbers and new vaccinations and the attendance rates of Asian children in 2021 and 2022. This aligns with existing research suggesting that individuals of Asian descent are more susceptible to COVID-19 infection than other demographics, which explains the heightened impact of cases on Asian children each year [205]. Additionally, limited access to COVID-19 testing among Asians may contribute to this trend. A recent study in the United States highlights disparities in testing accessibility among Asian Americans, which may reflect a similar situation in New Zealand [206]. Furthermore, research on children in England indicates that Asian children were more likely to have COVID-19 hospital admissions compared to White children [193].

However, there are also several limitations of the analyses. Firstly, including children from diverse ethnic backgrounds may introduce variability in the experimental outcomes. Secondly, the exclusive focus on preschool children limits the generalizability of the findings. Future research should include primary and secondary school students to expand the scope of understanding. Additionally, the chapter's narrow focus on New Zealand restricts the applicability of its findings to other countries.

Future research should incorporate a broader range of nations and diverse population samples to enhance external validity. Furthermore, the chapter has primarily explored case factors and vaccination rates, leaving demographic factors, such as gender, and other variables, including restriction policies, unexplored by regression analysis.

6.6 Conclusion

This chapter provides valuable insights into the impact of the COVID-19 pandemic on attendance rates among children from diverse ethnic backgrounds enrolled in early learning services in New Zealand. The analysis demonstrates a significant negative correlation between daily increases in new COVID-19 cases and attendance rates, particularly during periods of case surges and policy changes. Furthermore, the chapter underscores the adverse effect of COVID-19 vaccinations on attendance rates, especially among children of Asian ethnicity. This highlights the need to consider cultural barriers when designing policies and informing the development of targeted interventions to ensure equitable access to education during public health crises for all ethnic groups within a country. It is crucial to account for ethnic differences when addressing the spread of the pandemic and to examine how changes in case and vaccination numbers impact service delivery in these communities.

Future research should expand the scope to include multiple countries and refine the categorization of ethnic groups among children. Extending the investigation to primary and secondary school students will also be crucial for deepening our understanding of the pandemic's effects on diverse populations. Such efforts are essential for informed policymaking and the development of effective intervention strategies.

Chapter 7

Disparities in School Attendance During COVID-19: The Impact of Pandemic Factors on Ethnic and Māori Groups in New Zealand

7.1 Introduction

The COVID-19 pandemic has profoundly impacted regions globally since 2020 [207]. This unprecedented crisis has resulted in extensive fatalities and infections [208], prompting nations worldwide to implement stringent measures such as lockdowns to mitigate transmission risks [7]. Consequently, educational institutions, ranging from primary schools to universities, were compelled to rapidly transition to online learning modalities [9], significantly reshaping the educational landscape.

The influence of the COVID-19 pandemic on education manifests in several dimensions. Studies have indicated significant changes in academic performance among students [57, 158], with findings suggesting both challenges and opportunities associated with online education during the pandemic [165]. Moreover, the mental health impacts on students have been notably adverse, as many suffered from anxiety and depression [6]. Additionally, the pandemic has highlighted significant disparities in school attendance rates, particularly affecting marginalized communities and Indigenous populations such as the Māori in New Zealand [209].

Faced with these significant challenges during the pandemic, the New Zealand government implemented several measures to mitigate the broader impacts of COVID-19. One of the key strategies, consistent with approaches adopted in other countries, was the lockdown policy, with the initial lockdown lasting for two and a half months [210]. Vaccination strategies were also introduced, with double-dose vaccinations made freely available to a large proportion of the population [210]. In addition, for individuals with lower incomes, the government introduced relevant tax relief and financial support policies [210].

Despite extensive research on these aspects, several critical gaps remain. Firstly, the majority of existing studies focus on the general impacts of the pandemic on attendance rates [192, 211], resulting in a notable lack of comprehensive analyses concerning the variability in school attendance patterns throughout the pandemic, particularly regarding ethnic disparities such as those affecting Māori students. However, it is essential to examine the influence of the pandemic on ethnic differences and Māori schools. Ethnicity must be considered in research, as it affects the generalizability of findings and ensures that the benefits of educational policies are equitably extended to diverse ethnic groups [212]. Māori ethnicity represents a key component of this diversity. For instance, according to research conducted in New Zealand, many schools enroll a significant number of Māori students [213]. Māori culture constitutes an integral part of New Zealand's cultural landscape, and given the notable achievements of Māori schools, it is important to investigate their corresponding performance [214]. Furthermore, due to the predominant emphasis on the general effects of the pandemic, existing studies often overlook nuanced analyses of COVID-19-related factors across different periods, such as case numbers and vaccination rates [192, 211], which may exert variable influences over time. Moreover, the analytical approaches employed in current research are frequently confined to correlation analysis or a single regression model [215, 216], which may lead to biased findings and undermine objectivity.

This chapter seeks to address these gaps by investigating the influence of COVID-19 pandemic-related factors on school attendance in New Zealand. It examines the correlation between fluctuations in COVID-19 case numbers and vaccination rates with school attendance patterns, with particular emphasis on understanding these dynamics within Māori educational settings. Therefore, the research question is: "How has the COVID-19 pandemic influenced the attendance rates of different ethnic groups and Māori schools in New Zealand?"

The chapter contributes to the existing literature in two primary aspects. First, it uncovers significant year-over-year differences in the association between various pandemic-related variables and school attendance rates. Second, it demonstrates considerable variability in this association across different ethnicities, with the most pronounced impacts observed among certain ethnic groups and specific types of Māori schools.

This chapter is structured as follows. Section 7.2 critically synthesizes existing research, highlighting key areas for further investigation. Section 7.3 outlines the data collection strategies and examines the trends and changes in different variables within the datasets. Section 7.4 details the analytical approaches utilized, justifying the selection of these methods. Section 7.5 presents the empirical results and discusses their implications. Finally, Section 7.6 summarizes the key findings and suggests avenues for future research.

7.2 Literature Review

The COVID-19 pandemic has significantly influenced student participation in educational institutions, as evidenced by several studies. Research on online classes at Hashemite University revealed a negative impact of the pandemic on student interaction [211]. Another research regarding student attendance rates reported a decline in school attendance during the COVID-19 pandemic [192]. Likewise, research on adult education in the Czech Republic observed a significant drop in attendance following the onset of the pandemic in 2020 [217]. Furthermore, a multinational study across 19 countries highlighted adverse effects on academic performance among students globally [218].

Researchers have also extensively investigated the impacts of vaccinations on students during the COVID-19 pandemic. A study across several Vietnamese universities suggested that fully vaccinated students are less inclined towards online learning [219]. Additionally, one research study in China indicated that the pandemic had minimal impact on physical education courses among students [220].

Furthermore, scholars have investigated the impacts of the pandemic on student engagement using various analytical methods. For instance, a study examining student loneliness utilized Pearson correlation methods and hierarchical regression analyses [215]. Another study focused on secondary school students utilized comparison tests to explore significant differences in engagement during the pandemic from multiple perspectives [221]. Additionally, research on elementary school students' engagement applied Pearson correlation coefficients and a mediation model for its analysis [216].

Despite these insights, several notable gaps persist in existing research. Firstly, insufficient attention has been given to how these effects vary among ethnic groups and school types, particularly concerning the Māori ethnicity. Secondly, a lack of comparative analysis across different pandemic phases hinders a comprehensive understanding of its evolving impacts over time. Variations in cultural, social, and economic factors among ethnic groups may lead to diverse impacts of the pandemic. Additionally, rigorous analytical methods remain scarce, particularly in utilizing regression models to assess the effects of COVID-19-related factors on school attendance rates.

7.3 Materials

7.3.1 Data Collection

The data collection utilizes public databases, which are divided into two parts: the dataset on attendance rates and the COVID-19 dataset.

The public database on attendance rates is sourced from Education Counts [222], an agency under the New Zealand Government. "Schools" in this context refers to both primary and secondary schools. In New Zealand, the compulsory school entry age is between six and 16, although many students begin their schooling after their fifth birthday [223].

This database includes daily attendance rates from most schools in New Zealand, collected from September 2020 to May 2023 using standardized methods.

The data from multiple years must be comparable to facilitate a comprehensive and consistent comparison across various years. However, in 2020, only attendance rate data after May 25, 2020, are included, and in 2023, only data before May are included. As a result, the years 2020 and 2023 are excluded from the analysis, which represents a limitation in the data collection for this study.

The collected attendance rate data in 2021 differs from that in 2022. In 2021, the attendance rate was reported as a single aggregated value. By contrast, in 2022, the data were disaggregated into online and offline attendance rates, although the total number of registered attendees remained unchanged. Both online and offline attendance data, including national-level figures and those disaggregated by ethnic group, were collected using consistent standards. A student was considered off-site if they studied at home at any point during the day. However, if a student attended both at home and at school on the same day, their attendance was recorded as both on-site and off-site [222].

Therefore, the aggregated attendance rate in 2022 is calculated as follows:

$$\text{attendance rate} = \text{online attendance rate} + \text{offline attendance rate}. \tag{7.1}$$

The database includes various student groups, including the national group, Asian group, Māori group, Pacific group, and European/Pākehā group. Additionally, the database categorizes Māori schools by the medium of instruction: those that exclusively use the Māori language, those using a bilingual approach, and those that exclusively use English.

In addition to the New Zealand attendance rates, the COVID-19 database is adopted for data analysis. Data from WHO reports has been utilized in this chapter [167] for several reasons. First, the data collection standards for COVID-19 case variables and vaccination variables were consistent in 2021 and 2022, with new standards introduced in March 2023. Furthermore, the dataset includes a variety of case and vaccination variables, providing comprehensive data for thorough analysis.

Data selection is based on the availability of key variables within the datasets. To enhance the objectivity of the analysis, only dates for which data on attendance rates, new cases, total cases, total vaccinations, people vaccinated, people fully vaccinated, and new vaccination numbers were available were included in the analysis.

To ensure a comprehensive and relevant dataset aligned with the corresponding objectives, this chapter selected attendance rate data by ethnicity from 164 out of 180 dates in 2021 and 166 out of 184 dates in 2022 for analysis. Additionally, attendance rate data from different Māori schools were analyzed using 164 out of 180 dates in 2021 and 179 out of 184 dates in 2022. The selected dates were determined by the availability of COVID-19 case numbers and vaccination data, ensuring that the analysis reflects the temporal context of the pandemic.

7.3.2 Preliminary Data Visualization and Research Design

This section aims to use preliminary data visualization to guide the subsequent research design. Visualizing trends in attendance rates alongside new COVID-19 cases and vaccinations provides a comparative view of how these variables fluctuate together. These visual insights inform the formulation of hypotheses, allowing for a more structured and systematic evaluation of the relationships between school attendance and pandemic-related factors.

7.3.2.1 New Cases and Attendance Rates by Ethnicity

Fig.7.1 illustrates the relationship between new COVID-19 cases and school attendance rates in New Zealand during 2021 and 2022. The graph depicts the changes in new case numbers and attendance rates across different ethnic groups from February to December 2021, specifically N (national attendance rate), M (Māori attendance rate), P (Pacific attendance rate), A (Asian attendance rate), and E/P (European/Pākehā attendance rate), highlighting an inverse relationship between the two variables.

At the beginning of the year, the number of new cases remained very low from February to August, followed by a significant surge in August. Then there was a decrease in the number in September, followed by a rapid increase from late September to early November, and it started to decrease.

The attendance rates across all ethnicities, including the national average, exhibited similar patterns. In February, there was a significant drop in attendance rates. However, the significant drop did not last for a very long time; it only lasted for several days, and then it stabilized at a certain level with fluctuations. Additionally, there was a substantial decrease in attendance rates of multiple ethnicities. Similarly, there was a spike in new cases in October, corresponding to a gradual decrease in attendance rates of all groups since late November.

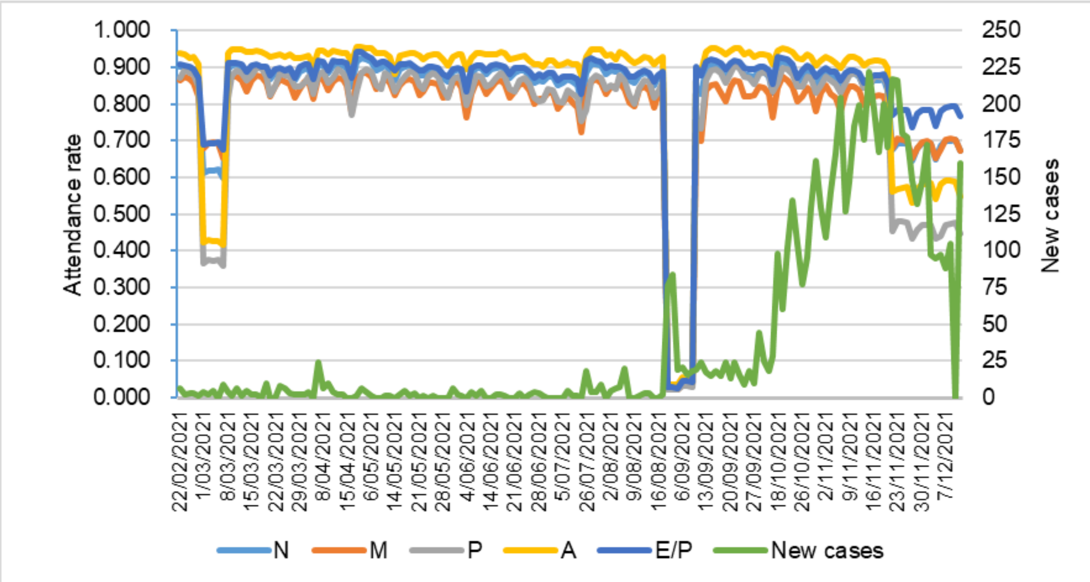


Fig. 7.1 New Cases and School Attendance Rates by Ethnicity in 2021

Fig. 7.2 illustrates the changes in new COVID-19 cases and attendance rates in New Zealand from February to December 2022. In general, there is an inverse relationship between the trends of the number of new cases and the attendance rate.

Throughout this period, the number of new cases experienced several significant fluctuations, particularly before August. There was a notable increase in new cases in February. Following a subsequent decrease, there was another significant increase in March, followed by a decline. Despite a substantial spike in July and August, the overall tendency of the data is decreasing, reaching zero by October.

The graph also depicts changes in attendance rates of each group. There are great similarities among the changes in the attendance rates of each group. At the beginning of the year, in February, when new cases surged, the attendance rates of each group decreased. From late March to April, despite fluctuations, attendance rates have increased. Then it fluctuated, while in June and July, it once reached a low point. Since September, the attendance rates of each group have tended to stabilize at a certain level.

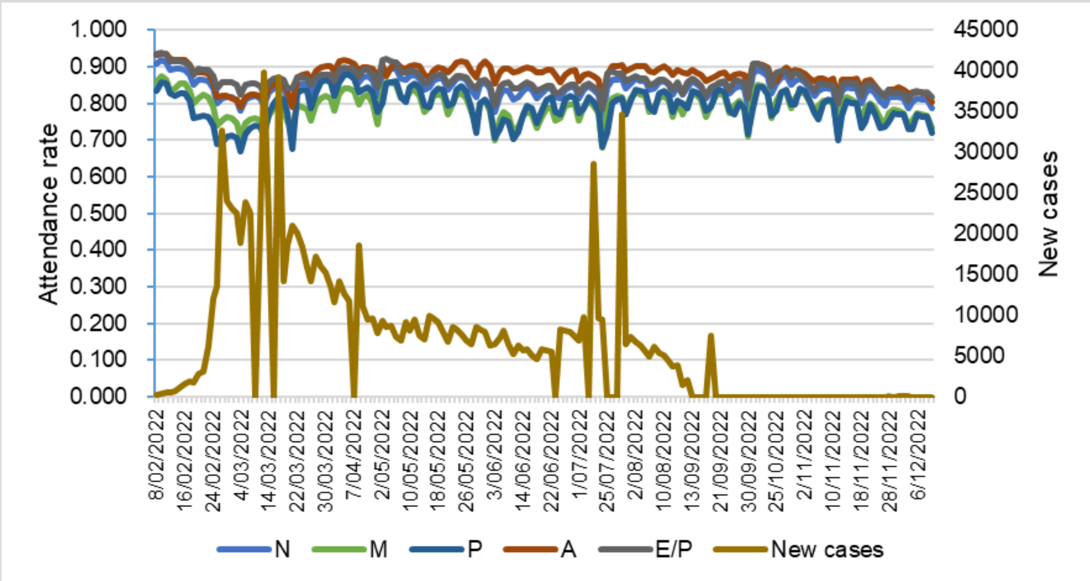


Fig. 7.2 New Cases and School Attendance Rates by Ethnicity in 2022

Based on the changes observed in the graphs for 2021 and 2022, the hypothesis for the relationship is as follows: The number of new COVID-19 cases negatively influenced the attendance rates of students across different ethnicities in 2021 and 2022.

7.3.2.2 New Vaccinations and Attendance Rates by Ethnicity

Fig. 7.3 illustrates the number of new COVID-19 vaccinations and school attendance rates among different ethnicities in New Zealand from February to November 2021. Throughout the year, the trends in vaccination numbers and attendance rates exhibit opposing patterns. The number of new vaccinations began to rise since February, reaching a marked spike in July, followed by a subsequent decline.

Concurrently, although attendance rates varied among different ethnicities, consistent patterns were observed. Notably, as new vaccinations increased at the beginning of 2021, attendance rates significantly dropped. Similarly, during the surge in vaccinations in early September, there was a notable decline in attendance rates. After the decrease in vaccinations in September, the attendance rate was back to the previous level.

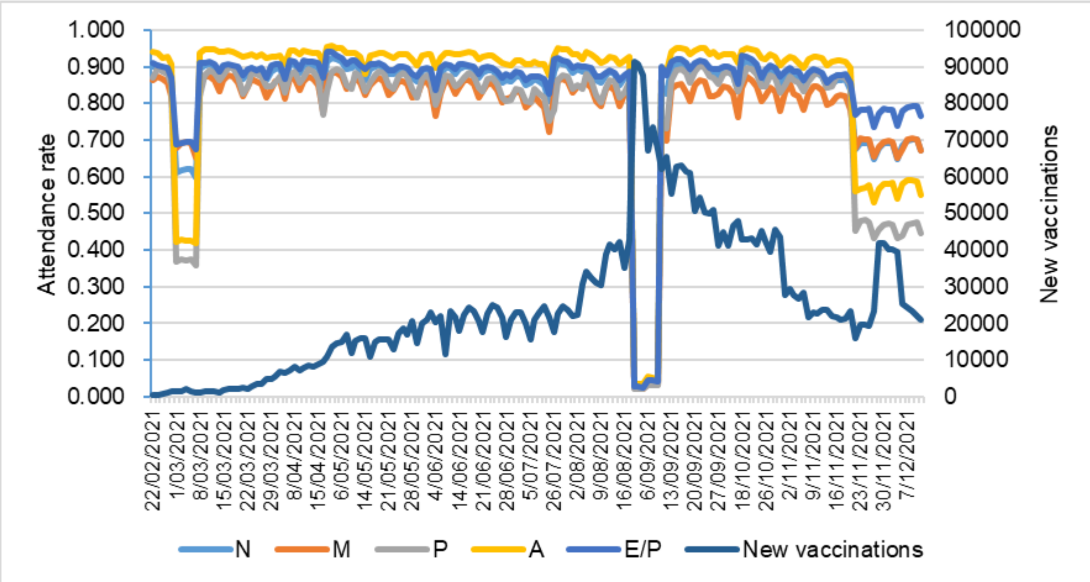


Fig. 7.3 New Vaccinations and School Attendance Rates by Ethnicity in 2021

Fig. 7.4 illustrates the changes in the number of new COVID-19 vaccinations and attendance rates among different ethnicities in New Zealand from February to December 2022. The pattern of school attendance contrasts with the changes in new vaccinations. Overall, the number of new vaccinations decreased from February to April, with notable spikes in July and August. At the beginning of 2022, there was a significant decrease in new vaccinations, while the attendance rates of each ethnicity and the national ones, despite some fluctuations, showed a slight increase. In July, a notable spike in new vaccinations coincided with a decrease in attendance rates.

The hypothesis based on the pattern of the graph is that the number of new vaccinations negatively influenced the attendance rates of students of different types of ethnicities in the years 2021 and 2022.

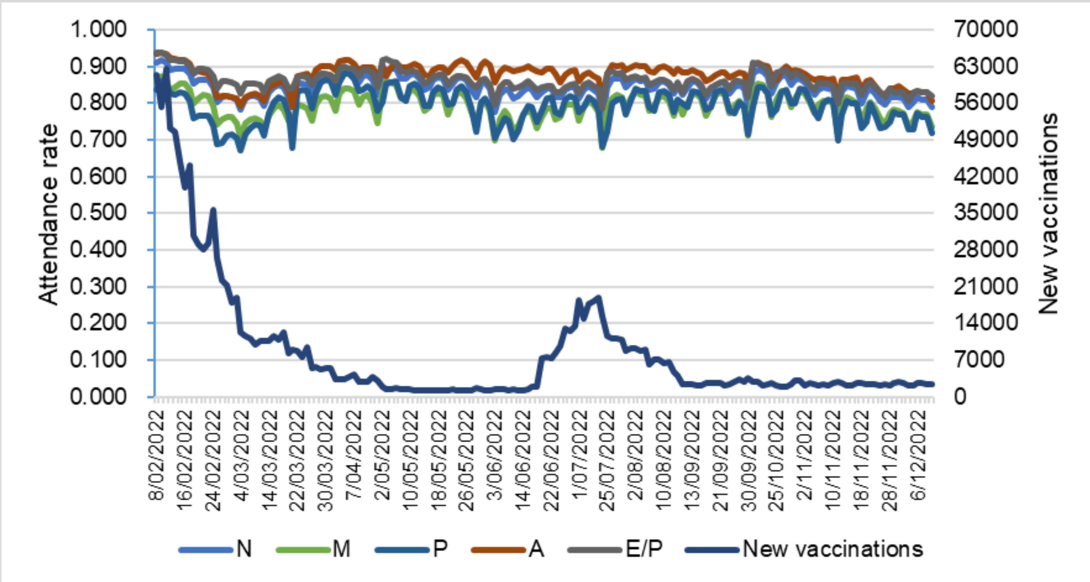


Fig. 7.4 New Vaccinations and School Attendance Rates by Ethnicity in 2022

7.3.2.3 New Cases and Attendance Rates in Māori Schools

Fig.7.5 illustrates the changes in new COVID-19 cases and attendance rates among three types of Māori schools: those with instruction exclusively in Māori, those with mixed-language instruction, and those with instruction exclusively in English, from February to December 2021. The graph shows that attendance trends in these Māori schools generally opposed the trends in new COVID-19 cases, while the trends of the attendance rates of the three different types of schools are nearly identical. In the first half of 2021, fluctuations in attendance rates across the three types of Māori schools closely mirrored the inverse of the changes in new cases. Notably, a significant increase in new cases in September was accompanied by a substantial drop in attendance rates of the three types of schools. After September, as case numbers continued to rise, attendance rates for all three types of schools gradually declined.

Fig. 7.6 illustrates the changes in the number of new COVID-19 cases and the attendance rates of three types of Māori schools in New Zealand throughout 2022. The trends in attendance rates of the three types of Māori schools generally oppose those of new cases, while the trends of the changes in the attendance rates

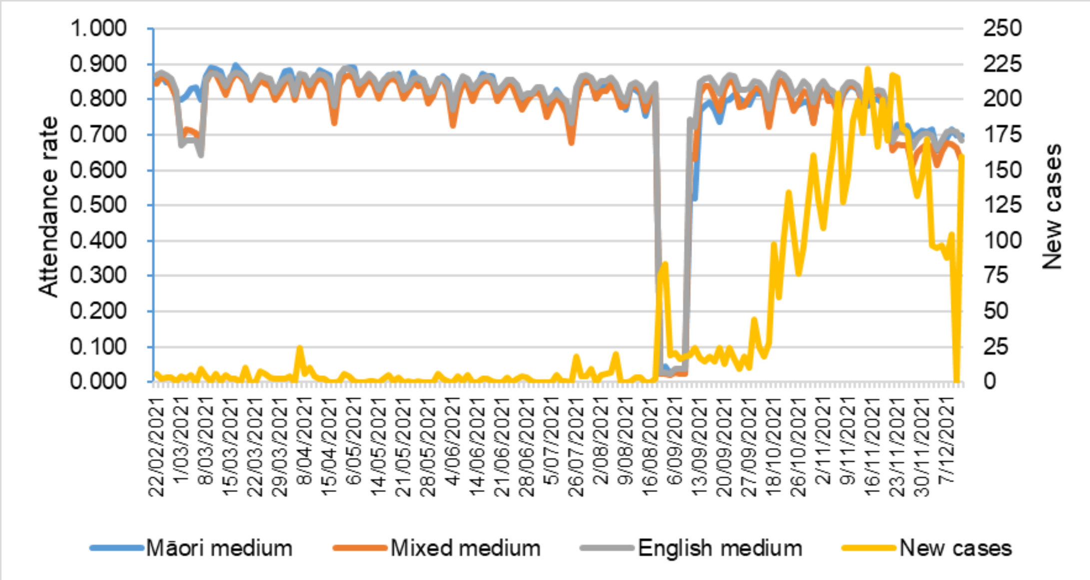


Fig. 7.5 New Cases and School Attendance Rates across Different Māori Schools in 2021

of the three types of Māori schools are nearly identical.

At the beginning of the year, in February, a spike in new cases coincided with a significant decrease in attendance rates. After substantial fluctuations in new cases, the decline in case numbers led to increased attendance rates, reaching a new level despite some ongoing volatility. However, when new cases spiked again in August, attendance rates across the three types of schools significantly decreased. Additionally, despite fluctuations, as new cases continued to decrease from August to December, attendance rates further increased.

Therefore, the hypothesis is: The COVID-19 cases had a negative influence on the attendance rate of different types of Māori schools in 2021 and 2022.

7.3.2.4 New Vaccinations and Attendance Rates in Māori Schools

Fig.7.7 illustrates the changes in the number of new vaccinations and the attendance rates of three different types of Māori schools in New Zealand throughout 2021. As the number of new vaccinations gradually increased before September, attendance rates at the schools tended to decrease from 0.800 to 0.700, despite

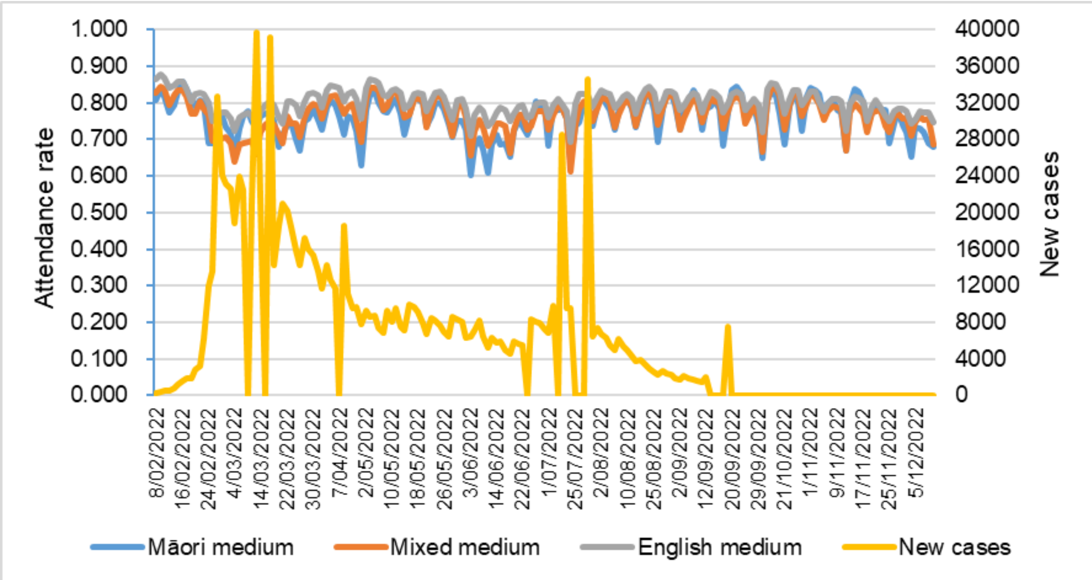


Fig. 7.6 New Cases and School Attendance Rates across Different Māori Schools in 2022

some fluctuations. In September, a significant spike in new vaccinations was accompanied by a notable decrease in attendance rates. Following this peak, the number of new vaccinations gradually decreased but increased again in November. In November, attendance rates continued to decline, reaching a lower level.

Fig.7.8 illustrates the changes in the number of new vaccinations and attendance rates among different types of Māori schools in New Zealand during 2022. The trends in these two variables appear to be opposite. At the beginning of 2022, the number of vaccinations gradually decreased, while attendance rates for different types of Māori students increased from late February to March, despite some fluctuations. However, in July, when there was a spike in the number of new vaccinations, the attendance rates of these schools tended to decrease during the spike in vaccination dosage.

The hypothesis based on the inverse relationship between new vaccinations and attendance rate is that the COVID-19 new vaccinations have a negative influence on the attendance rate of different types of Māori schools in 2021 and 2022.

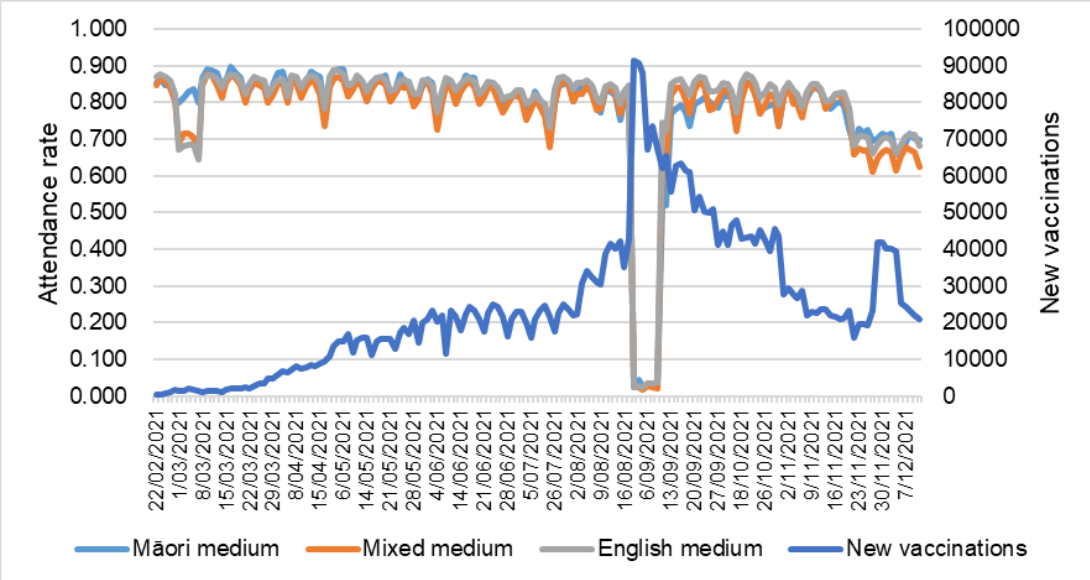


Fig. 7.7 New Vaccinations and Attendance Rates across Different Māori Schools in 2021

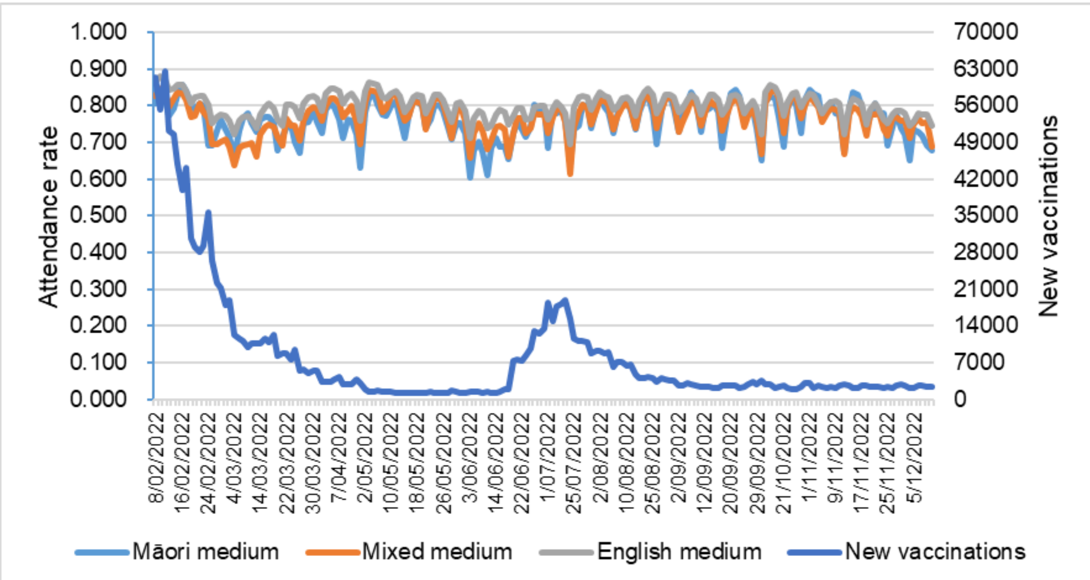


Fig. 7.8 New Vaccinations and Attendance Rates across Different Māori Schools in 2022

7.4 Methodology

This chapter employs a quantitative methodology to ensure analytical objectivity, utilizing both correlation and regression analyses.

7.4.1 Correlation Analysis

The Spearman correlation coefficient is utilized to analyze the relationship between COVID-19 pandemic factors and attendance rates. This coefficient, a measure used to estimate the monotonic relationship between variables, is chosen for its appropriateness given the nature of the data involved [94]. Specifically, in real scenarios, case numbers and vaccination numbers are integers, which are categorized as ordinal data [94]. Moreover, the Spearman correlation coefficient is based on the ranking of variables. All the independent variables are converted into rankings, and their corresponding dependent variables are converted into rankings as well [102]. It is less affected by extreme values, which are present in variables such as new cases and new vaccinations.

The Spearman correlation coefficient formula is illustrated as (7.2) [102]:

$$r = 1 - \frac{6 \sum d_t^2}{n(n^2 - 1)}, \tag{7.2}$$

where d_t is the difference in ranks between paired variables, and n represents the number of data points.

7.4.2 Regression Analysis

Various regression techniques are employed to analyze the relationships between variables. Typically, a single regression model is applied to examine the relationships among multiple variables. However, each regression model has its limitations and underlying assumptions, which may introduce bias into the results. Therefore, the selection of each method is guided by its suitability for the characteristics of the dataset and its ability to meet the analytical requirements, to explore the consistency patterns across the results. The application of multiple regression models aims to explore whether the independent variables influence the dependent variables, instead of forecasting and making predictions.

The methods applied include:

- MLR: MLR is a classical statistical technique adopted to model the linear relationship between multiple independent variables and a single dependent variable [224]. It is based on the assumption of linear distribution in the input data [107]. This method is particularly suited to the chapter as it allows for the integration of multiple predictors, aligning well with the need to analyze complex relationships within our dataset.

- SVR: Originally developed for classification tasks, SVR has been adapted for regression by leveraging the principles of SVM [109]. It constructs regression models in high-dimensional feature spaces [225], enabling it to be well-suited for handling the complexity of our dataset, which involves multiple variables.

- KNN: KNN is a non-parametric method applicable to both classification and regression tasks [226]. It predicts outcomes based on the "K" nearest training samples [227] and does not rely on assumptions regarding the underlying data distribution [117]. This enhances its robustness against outliers and its effectiveness in handling varying data distributions [115]. KNN is often applied to large and complex datasets [115]. Its simplicity and adaptability to diverse data characteristics enhance its suitability for the analysis, especially given its success in pattern recognition tasks, such as human action type recognition [228]. In this chpater, KNN is appropriate for modeling the complexity and fluctuations present in the data.

- GPR: GPR is a probabilistic, non-parametric method that excels in modeling complex relationships in multi-dimensional spaces [229]. It is particularly effective in addressing challenges such as high dimensionality, nonlinearity, and small sample sizes, and is frequently used in time series prediction and dynamic system model identification [119]. Its flexibility in capturing diverse data relationships makes it a valuable tool for our analysis, especially as the datasets in the chapter involve multiple features, intricate dependencies, and a relatively small sample size.

- DTR: DTR employs a tree-like structure to represent decisions and their possible outcomes. It operates by recursively splitting the dataset into homogeneous subsets until all data within a subset are considered to belong to the same class [128]. This versatile method supports both numerical and categorical data types [115] and has been widely applied in domains such as medical diagnostics and finance, where high accuracy is essential [128]. DTR is well-suited to the diverse datasets, which include numerical and ordinal variables. However, techniques such as pruning may be necessary to mitigate overfitting [230].

- PLS: PLS is effective for both classification and regression tasks, particularly in the presence of multicollinearity among predictor variables [120]. It has been widely adopted for pattern classification and function approximation [231]. Given the potential correlations in our data, PLS provides a robust framework for managing multicollinearity, thereby supporting reliable and interpretable results.

- MLP: MLP, a type of artificial neural network, is capable of modeling complex non-linear relationships and applies to both regression and classification tasks [123]. Its ability to be trained using flexible algorithms enables it to be a powerful tool for analyzing datasets with multiple features, which aligns with the multi-variable nature of the chapter [124].

Selecting an appropriate evaluation metric is crucial for accurately assessing the performance of different regression models, while the aim in this context is to evaluate the relevance between the corresponding variables and attendance rates based on existing data, rather than to predict attendance rates in future years. This chapter chooses the MAPE as the primary metric for evaluating prediction accuracy [131]. MAPE, which calculates the average of absolute percentage errors, offers the advantage of being scale-independent [232]. This characteristic ensures that the metric is applicable across various types of datasets, as the results are expressed in percentages and remain unaffected by the size of the dataset. Given that all of the attendance rates are lower than 100 percent, relying solely on absolute differences between actual and predicted values would not provide an effective measure of prediction accuracy. Thus, MAPE is particularly suitable for this analysis, as it offers a clear and intuitive representation of prediction errors relative to the scale of the data.

The formula for the MAPE is shown in Equation (7.3) [131].

$$\text{MAPE} = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{A_t - F_t}{A_t} \right|, \tag{7.3}$$

where N denotes the number of samples in the dataset, A_t means the actual values, and F_t signifies the predicted values [131].

The application of MAPE in all regression analyses is intended to identify consistent patterns across various regression models for the years 2021 and 2022. Generally, lower MAPE values in a given year suggest that changes in the dependent variables are more closely aligned with changes in the independent variables for that year.

7.5 Experimental Results

7.5.1 Variable Selection

As detailed in the analysis sections, this chapter is structured into two main analytical components. The first component investigates the impact of the pandemic on school attendance among various ethnic groups in New Zealand. The second component examines the effects of the pandemic on attendance rates across different types of Māori schools in New Zealand.

For the analysis of the impacts of COVID-19 cases and vaccinations on attendance rates, several variables have been selected:

- **Variables for regression and correlation analysis**
  - **Total cases**: The cumulative number of confirmed and probable COVID-19 cases reported daily. The variable is utilized as an independent variable in the regression analysis.

- **New cases**: The count of newly confirmed and probable COVID-19 cases reported daily. Regression and correlation analyses utilize this variable as the independent variable.
- **Total vaccinations**: The total number of vaccine doses administered daily. This variable is utilized as an independent variable in the regression analysis.
- **People vaccinated**: The cumulative number of individuals who have received at least one dose of the vaccine daily. This variable is utilized as an independent variable in the regression analysis.
- **People fully vaccinated**: The cumulative number of individuals who have received two doses of the vaccine daily. This variable is utilized as an independent variable in the regression analysis.
- **New vaccinations**: The number of new vaccine doses administered daily. This variable is utilized as an independent variable in both regression and correlation analysis.

- **Attendance rates**
  - **N**: National attendance rate.
  - **A**: Attendance rate for Asian students.
  - **E/P**: Attendance rate for European/Pākehā students.
  - **M**: Attendance rate for Māori students.
  - **P**: Attendance rate for Pacific students.
  - **Māori medium**: Attendance rate for Māori schools where the medium of instruction is solely in the Māori language.
  - **Mixed medium**: Attendance rate for Māori schools where the medium of instruction is a mix of languages.
  - **English medium**: Attendance rate for Māori schools where the medium of instruction is English.

7.5.2 Effects of the COVID-19 Pandemic on Ethnicity Attendance Rates

7.5.2.1 Impacts of COVID-19 Cases on Ethnicity Attendance Rates

To analyze the impact of COVID-19 cases on ethnicity attendance rates, this section explores the correlations between new COVID-19 cases and attendance rates across various ethnicities, including the national average, for 2021 and 2022, as shown in Table 7.1.

The significance of the results is compared against a threshold of 0.05. If the *p*-value is lower than 0.05, we conclude that there is a statistically significant correlation between the two variables in the correlation analysis. In the table, "**" denotes *p*-values less than 0.01, indicating great statistical significance, while "*" indicates *p*-values less than 0.05, denoting statistical significance. This notation clarifies the strength of the correlations observed in the analysis.

Table 7.1 illustrates significant variations in correlation coefficient values between the two years. In 2021, all *p*-values for correlation coefficients, except those of the Pacific ethnicity, are below 0.05, indicating statistically significant correlations between new case numbers and attendance rates. The observed correlation coefficients are uniformly negative, ranging from 0.1719 to 0.2439, suggesting that higher new case numbers are associated with lower attendance rates in most ethnicities' schools. Māori ethnicity shows the strongest correlation (r=-0.2439). In contrast, in 2022, no *p*-values are below 0.05, signifying no statistically significant correlations between new case numbers and attendance rates for any ethnicity.

Table 7.1 Correlation Between New COVID-19 Cases and School Attendance Rates by Ethnicity for 2021 and 2022

Ethnicity	2021	2022
N	-0.2186**	-0.0002
A	-0.1719*	0.0213
E/P	-0.1742*	0.0901
M	-0.2439**	-0.1418
P	-0.0145	0.0354

This section further examines MAPE values associated with predicting ethnicity-specific attendance rates using a range of regression models, based on COVID-19 case-related variables, as presented in Table 7.2 and Fig. 7.9. Total COVID-19 cases and new daily cases in New Zealand were employed as independent variables to provide a more objective basis for estimating attendance rates. As shown in both the table and the figure, across all regression models utilizing machine learning methods, the MAPE values for 2021 are consistently higher than those observed for 2022 across all ethnic groups, including the national average. In 2022, all MAPE values fell below 0.1, whereas in 2021, the majority of values exceeded 0.2, with several models yielding MAPE values substantially greater than 0.5. These findings indicate that the predictive accuracy of attendance rates based on COVID-19 case variables, such as total and new cases, was considerably lower in 2021 compared to 2022.

Table 7.2 MAPE of Ethnicity-Specific Attendance Rates Predictions Based on Case Variables for Different Regression Models in 2021 and 2022

	N		A		E/P		M		P	
	2021	2022	2021	2022	2021	2022	2021	2022	2021	2022
MLR	0.5975	0.0268	0.5428	0.0294	0.5803	0.0206	0.7230	0.0357	0.8322	0.0567
SVM	0.5936	0.0250	0.5367	0.0308	0.5758	0.0193	0.6986	0.0367	0.8147	0.0520
KNN	0.2487	0.0190	0.2435	0.0163	0.2461	0.0152	0.2708	0.0282	0.3506	0.0432
GPR	0.5975	0.0268	0.5428	0.0294	0.5803	0.0206	0.7230	0.0357	0.8320	0.0567
DTR	0.0209	0.0210	0.0246	0.0144	0.0287	0.0167	0.0271	0.0381	0.0428	0.0518
PLS	0.5975	0.0268	0.5428	0.0294	0.5803	0.0206	0.7230	0.0357	0.8322	0.0567
MLP	0.5081	0.0821	0.4737	0.0782	0.5603	0.0750	0.7003	0.0799	0.7356	0.0679

7.5.2.2 Impacts of COVID-19 Vaccinations on Ethnicity Attendance Rates

To analyze the impact of COVID-19 vaccinations on attendance rates by ethnicity, this section begins by examining the Spearman correlations between new

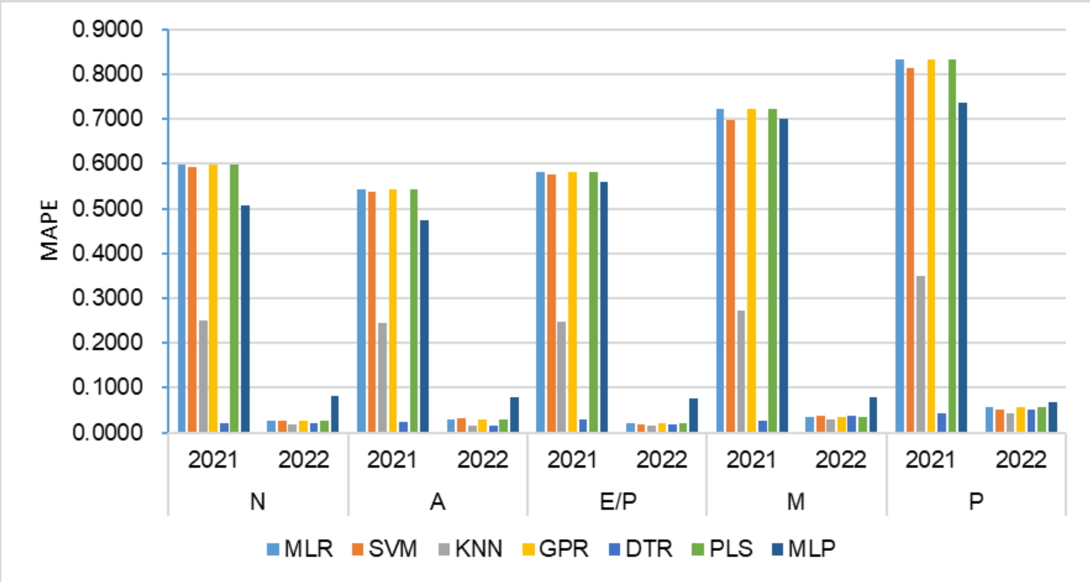


Fig. 7.9 MAPEs of Ethnicity-Specific Attendance Rate Predictions Based on Case Variables for Different Regression Models in 2021 and 2022

vaccinations and attendance rates across various ethnicities, including the national average, for 2021 and 2022, as illustrated in Table 7.3. The results for 2021 differ markedly from those for 2022. In 2021, all p-values, except those related to the Pacific ethnicity, are below 0.05, with absolute values ranging from 0.1580 to 0.3511, indicating statistically significant correlations between the number of new vaccinations and school attendance rates for these ethnicities. The negative correlation coefficients suggest that higher numbers of new vaccinations are associated with lower attendance rates in 2021, with the Māori ethnic group exhibiting the most pronounced effect. In contrast, for 2022, all p-values exceed 0.05, indicating the absence of statistically significant correlations between new vaccinations and attendance rates across all ethnicities in New Zealand.

Additionally, the analysis presents the MAPE values for attendance rate predictions across various ethnicities, based on multiple vaccination variables, including total vaccinations, people vaccinated, people fully vaccinated, and new vaccinations in New Zealand, as demonstrated in Table 7.4 and Fig. 7.10. The values for 2022 are substantially lower than those for 2021, indicating that predictions based on vaccination variables were more accurate in 2022. In 2021, all

Table 7.3 Correlation Between New COVID-19 Vaccinations and School Attendance Rates by Ethnicity for 2021 and 2022

	2021	2022
N	-0.2540**	-0.0092
A	-0.1580*	-0.1218
E/P	-0.1683*	-0.0901
M	-0.3511**	-0.0304
P	-0.1339	-0.1025

of the corresponding values are lower than 0.1, while in 2022, nearly all are larger than 0.3. This improvement in predictive accuracy suggests that the relationship between vaccinations and attendance rates was stronger in 2022 compared to 2021.

Table 7.4 MAPEs of Ethnicity-Specific Attendance Rates Predictions Based on Vaccination Variables for Different Regression Models in 2021 and 2022

	N		A		E/P		M		P	
	2021	2022	2021	2022	2021	2022	2021	2022	2021	2022
MLR	0.4953	0.0236	0.4687	0.0192	0.4824	0.0192	0.5759	0.0326	0.7059	0.0503
SVM	0.4640	0.0250	0.4532	0.0308	0.4649	0.0193	0.5297	0.0367	0.6357	0.0520
KNN	0.3809	0.0170	0.3658	0.0158	0.3722	0.0127	0.4236	0.0259	0.5382	0.0360
GPR	0.4996	0.0236	0.4839	0.0192	0.4839	0.0192	0.5792	0.0326	0.7353	0.0503
DTR	0.0230	0.0231	0.0191	0.0188	0.0172	0.0165	0.0309	0.0344	0.0290	0.0428
PLS	0.4870	0.0237	0.4866	0.0197	0.4632	0.0190	0.5595	0.0351	0.7381	0.0512
MLP	0.4205	0.0560	0.4240	0.0676	0.3771	0.0636	0.4889	0.0994	0.6767	0.1158

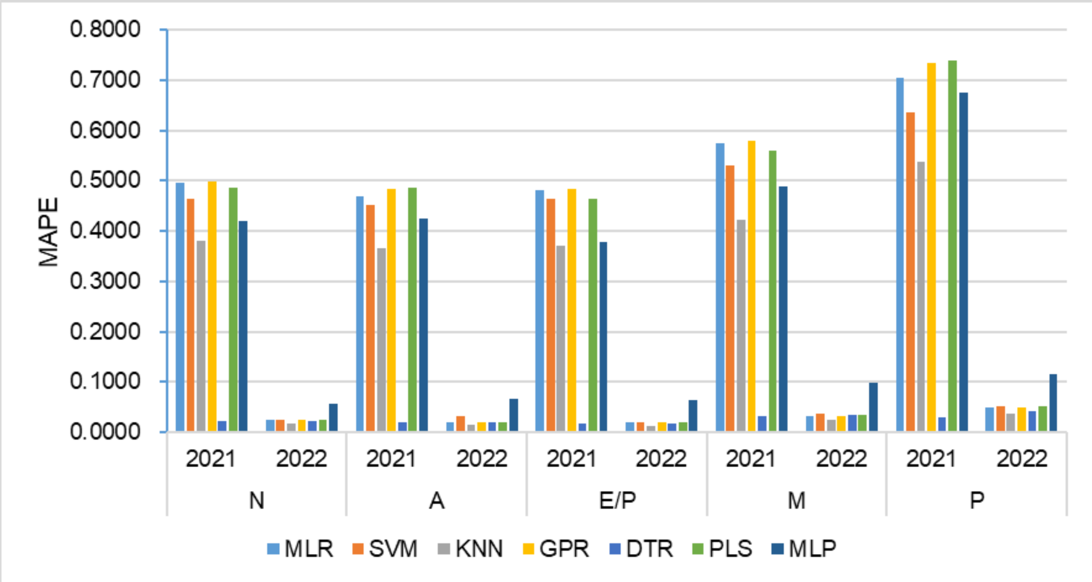


Fig. 7.10 MAPEs of Ethnicity-Specific Attendance Rates Predictions Based on Vaccination Variables for Different Regression Models in 2021 and 2022

7.5.3 Effects of the COVID-19 Pandemic on Attendance Rates in Māori Schools

7.5.3.1 Impacts of COVID-19 Cases on Māori School Attendance Rates

To further investigate the impact of COVID-19 vaccinations on Māori school attendance rates, we calculated the correlations between new vaccination numbers and attendance rates for different types of Māori schools in 2021 and 2022, as presented in Table 7.7. The correlations observed in 2021 are notably stronger than those in 2022. In 2021, all corresponding p-values are below 0.05, and all correlation coefficients are negative. These results indicate a statistically significant negative correlation between the number of new vaccinations and attendance rates across all school types in 2021. Among the three types, Māori-medium schools exhibit the strongest negative correlation (r = -0.5839), followed by Mixed-medium (r = -0.3793) and English-medium schools (r = -0.3215). In contrast, none of the p-values for 2022 fall below the 0.05 threshold, suggesting that new vaccinations were not statistically significantly correlated with attendance rates for any of the three school types in that year.

Table 7.5 Correlation Between New Cases and Māori School Attendance Rates for 2021 and 2022

	2021	2022
Māori Medium	-0.4170**	-0.2481**
Mixed Medium	-0.2422**	-0.2136**
English Medium	-0.2344**	-0.1228

Further, we examine the MAPE values for attendance rate predictions across different types of Māori schools, based on multiple case variables, including total COVID-19 cases and new cases in 2021 and 2022, as presented in Table 7.6 and Fig. 7.11. The MAPE values for 2022 are consistently below 0.1, while most of the corresponding values in 2021 exceed 0.6. This indicates that the predictions for 2022 are generally more accurate than those for 2021, suggesting a stronger predictive performance in 2022.

Table 7.6 MAPEs of Māori School Attendance Rates Predictions Based on Case Variables for Different Regression Models in 2021 and 2022

	Māori Medium		Mixed Medium		English Medium	
	2021	2022	2021	2022	2021	2022
MLR	0.6830	0.0622	0.8985	0.0598	0.7000	0.0405
SVM	0.6197	0.0709	0.8560	0.0666	0.6860	0.0420
KNN	0.1911	0.0594	0.3144	0.0523	0.2688	0.0344
GPR	0.6830	0.0622	0.8984	0.0598	0.6999	0.0405
DTR	0.0443	0.0625	0.0435	0.0569	0.0365	0.0335
PLS	0.6830	0.0622	0.8985	0.0598	0.7000	0.0405
MLP	0.6307	0.0911	0.6893	0.1013	0.6283	0.0810

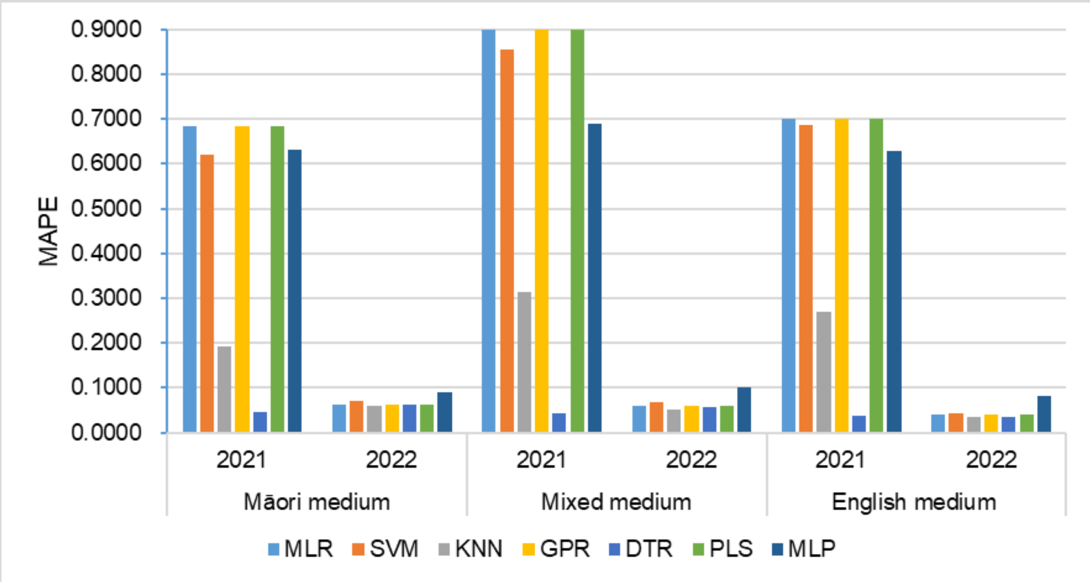


Fig. 7.11 MAPEs of Māori School Attendance Rate Predictions Based on Case Variables for Different Regression Models in 2021 and 2022

7.5.3.2 Impacts of COVID-19 Vaccinations on Māori School Attendance Rates

To further investigate the impact of COVID-19 vaccinations on Māori school attendance rates, we calculated the correlations between new vaccinations and attendance rates across different types of Māori schools in 2021 and 2022, as presented in Table 7.7. The correlation coefficients observed in 2021 are notably stronger than those in 2022. In 2021, all corresponding p-values are below 0.05, and all correlation coefficients are negative, indicating statistically significant negative correlations between the number of new vaccinations and attendance rates for each school type. Among the three types, Māori-medium schools exhibit the strongest correlation (0.5839), followed by Mixed-medium (0.3793) and English-medium schools (0.3215). In contrast, in 2022, none of the p-values are below 0.05, suggesting that new vaccinations are not statistically significantly correlated with attendance rates for any of the three school types during that year.

Further regression analysis examines the MAPE values of attendance rate predictions for various types of Māori schools during 2021 and 2022, using multiple vaccination variables such as total vaccinations, people vaccinated, people fully

Table 7.7 Correlation Between New Vaccinations and Māori School Attendance Rates for 2021 and 2022

	2021	2022
Māori Medium	-0.5839**	0.0420
Mixed Medium	-0.3793**	-0.1075
English Medium	-0.3215**	0.0015

vaccinated, and new vaccinations. The results are presented in Table 7.8 and Fig. 7.12. The table and figure illustrate that, across all types of Māori schools, the MAPE values in 2022 are substantially lower compared to those in 2021 for each regression model utilizing machine learning methods. This significant reduction in MAPE values suggests that the predictions based on vaccination data for 2022 are markedly more accurate than those made for 2021.

Table 7.8 MAPEs of Māori School Attendance Rates Predictions Based on Vaccination Variables for Different Regression Models in 2021 and 2022

	Māori Medium		Mixed Medium		English Medium	
	2021	2022	2021	2022	2021	2022
MLR	0.5141	0.0573	0.6951	0.0549	0.5623	0.0377
SVM	0.4430	0.0677	0.6366	0.0666	0.5169	0.0420
KNN	0.3177	0.0586	0.5011	0.0499	0.4186	0.0338
GPR	0.5172	0.0573	0.6997	0.0549	0.5656	0.0377
DTR	0.0402	0.0603	0.0443	0.0552	0.0287	0.0334
PLS	0.4964	0.0598	0.6809	0.0593	0.5463	0.0405
MLP	0.4718	0.0840	0.5749	0.0705	0.4707	0.0602

7.5.4 Discussion

The results reveal that the negative association between COVID-19 variables, such as case numbers and vaccination rates, and school attendance was more pronounced in 2021 compared to 2022. This pattern was consistent across various ethnic groups and types of Māori schools. A plausible explanation for the

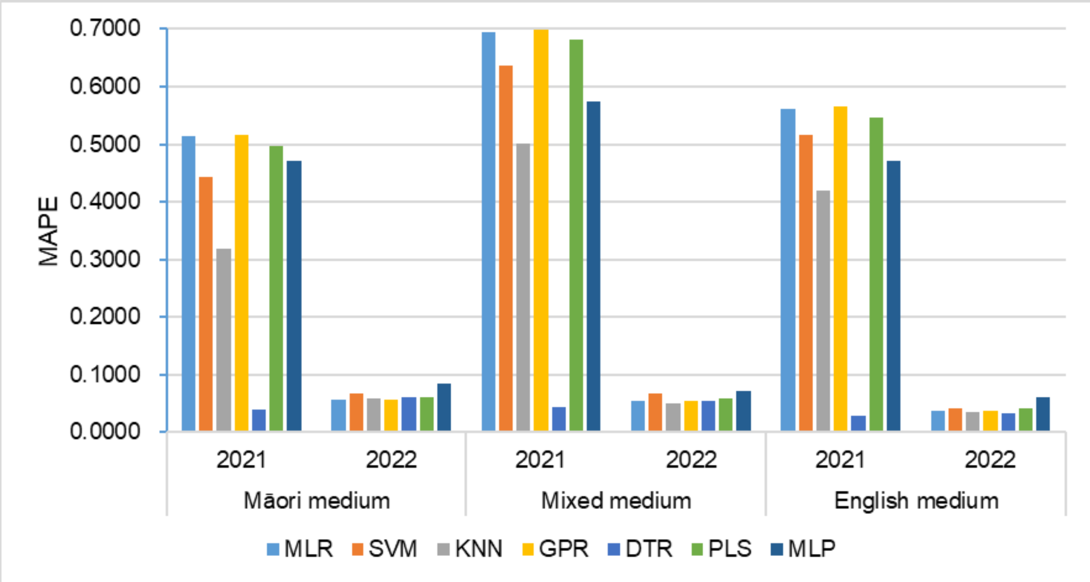


Fig. 7.12 MAPEs of Māori School Attendance Rates Predictions Based on Vaccination variables for Different Regression Models in 2021 and 2022

stronger association observed in 2021 is the heightened severity of the COVID-19 pandemic during that year. Specifically, 2021 was marked by stricter public health policies and the emergence of more virulent strains of the virus, which likely contributed to the greater impact on school attendance relative to 2022 [233]. By 2022, the overall severity of the pandemic had diminished due to widespread vaccination efforts and other mitigating factors. Consequently, the negative impact of COVID-19 on school attendance was more pronounced under the stringent conditions of 2021. Numerous studies have shown that school disengagement among children significantly increased during the onset of the pandemic [234]. Additionally, the initial side effects of early vaccine doses, predominantly administered in 2021, may have further contributed to the observed negative correlation [235].

Notably, significant disparities were observed in the associations between COVID-19 variables and attendance rates across different ethnic groups in 2021. The negative correlations between new COVID-19 cases, vaccination rates, and school attendance were particularly strong among Māori ethnic groups, especially in schools where Māori is the primary language of instruction. This trend is likely attributable to the inadequacy of pandemic policies in addressing long-standing disparities in resources and services. As a result, the adverse effects of COVID-19, including vaccine-related side effects, were more pronounced within the Māori community [179]. These findings are consistent with existing literature. For example, research on Māori and Pacific peoples in New Zealand demonstrates that the odds of hospitalization for Māori are significantly higher than for non-Māori [236], indicating that Māori communities were more vulnerable to the impacts of the pandemic. Another study on health inequalities found that Māori individuals faced greater challenges in accessing essential health services during the pandemic [237]. These findings underscore the need for targeted interventions to address such disparities and improve outcomes for Māori students.

It is also important to reconcile these findings with the MAPE values, which highlight discrepancies when compared to the correlation coefficient results. The higher MAPE observed in 2021 suggests increased prediction errors and greater variability in attendance rates across different ethnic groups and Māori school types, based on COVID-19 case and vaccination data. Nevertheless, the overall association between case numbers and attendance rates remained stronger in 2021. This complexity likely stems from a more pronounced relationship between attendance and COVID-19 cases in 2021, coupled with greater data variability relative to 2022. The Spearman correlation, which measures monotonic relationships based on rank order, captures the general trend but does not fully reflect the extent of variability. As such, attendance rates in 2021 were less stable than in 2022, as evidenced by earlier analyses revealing more extreme values in 2021. This instability adversely affected prediction accuracy, resulting in higher MAPE values.

Based on these findings, governments should implement timely and responsive measures in the face of similar future crises. At the onset of a pandemic, school attendance and overall educational performance may be severely disrupted. Therefore, it is crucial to introduce adaptive policies and alternative strategies to support the continuity of education across all schools. Furthermore, it is imperative to enhance the understanding of the needs of diverse ethnic communities, particularly in multi-ethnic countries. Certain groups may face limited access to critical services such as healthcare, vaccinations, and emergency support. Ensuring equitable access to high-quality services, especially for minority populations, is essential for reducing the disproportionate impact of public health crises.

Despite these insights, this chapter has several limitations. One key limitation pertains to the dataset employed. Although the data were collected by a consistent organization, the voluntary nature of responses and potential regional gaps during certain periods may have influenced the results, as some students' attendance was recorded in both online and offline categories, which may have affected the accuracy of the data. Furthermore, due to data availability constraints, certain key years—such as 2020 and 2023—were excluded from the analysis, representing a potential area for future research.

Additionally, the chapter focuses solely on school attendance rates, representing only one dimension of educational outcomes. Finally, as the analysis is confined to the context of New Zealand, the findings may not be directly generalizable to other countries or regions similarly affected by the COVID-19 pandemic.

7.6 Conclusion

This chapter investigates the impacts of the COVID-19 pandemic on school attendance rates among various ethnic groups and different types of Māori schools in New Zealand. The findings reveal a more pronounced association between COVID-19-related factors, such as new case numbers and vaccination rates, and attendance rates in 2021 compared to 2022. This stronger association in 2021 is attributed to the greater severity of the pandemic that year, which resulted in more significant disruptions and fluctuations in school attendance.

Furthermore, it highlights the particularly adverse effects experienced by Māori ethnic groups, especially in Māori-medium schools where instruction is delivered exclusively in the Māori language. These negative outcomes are largely due to the insufficiency of government policies in addressing the specific needs and long-standing disparities affecting the Māori community. Accordingly, governments need to implement targeted measures to maintain the continuity of school operations during public health crises and to ensure that all ethnic groups, particularly minorities, have equitable access to quality services and support.

Future research will focus on collecting more detailed and precise data for more accurate analyses. Subsequent studies will also examine additional dimensions of education, such as academic performance and study time. They will extend the scope to include data from a wider range of countries to develop a more comprehensive understanding of the pandemic's broader impact on education.

Chapter 8

Synthesizing the Educational Impact of COVID-19

8.1 The Impact of COVID-19 Pandemic on Academic Performance

The findings indicate that the pandemic's influence was significant across various global regions, especially at the beginning of the outbreak. However, notable differences in the impacts of academic performance were observed between the two tests. The TOEFL scores illustrate a significant increase in average performance across multiple sections from 2020 to 2021 worldwide. In contrast, the GMAT scores illustrated that at the beginning of the pandemic, both the overall average scores and the number of examinees decreased.

There are several reasons for the differences in academic performance changes across these standardized English tests at the beginning of the pandemic. One important reason lies in the differences between the contents of TOEFL and GMAT. TOEFL primarily assesses English level for non-English speakers, including reading, writing, and other skills. It is often taken when participants intend to enter certain universities or apply for a visa [238]. In contrast, GMAT is a much more comprehensive test, assessing English skills, mathematical skills, and logical reasoning skills. Designed for the Master of Business Administration programs, GMAT tests are much more rigorous [239]. Therefore, the more demanding examinations are more likely to result in corresponding stress, because of the high-level critical thinking skills requirements. The findings are aligned with the ones from some existing research studies during the pandemic; for example, a recent study demonstrates that academic performance in assessments such as TIMSS and PISA tests, which require mathematical and logical skills, has been negatively impacted by the pandemic [56].

Another contributing factor is the limitations of the home edition or online versions of the tests. The home-based TOEFL tests did not impose restrictions on the number of attempts and often included extended time and breaks [240]. On the contrary, there were multiple limitations of the GMAT online versions, while a typical example is that it is limited to 5 times a year [241], especially at the beginning of the pandemic. The added inconvenience of the GMAT format may have further affected performance by increasing test anxiety due to stricter regulations and limited testing frequency.

An additional factor is the difference in test participants. Those taking the TOEFL are generally applying for bachelor's or master's degree programs [242], so these students may have more flexible schedules to prepare for the test, and they tend to be at a young age [242]. In contrast, GMAT candidates are typically preparing for business school admission [239], which tends to be more urgent and time-sensitive, and the participants tend to be older than those of the TOEFL test. These differences in participant characteristics may also explain the variation in academic performance, as TOEFL candidates may have faced less time pressure, while GMAT candidates likely experienced higher levels of stress.

Therefore, it is essential to set up appropriate and corresponding strategies. A key measure is to develop the targeted programs for the enhancement of mathematical and logical skills under high-pressure conditions. Additionally, policymakers should provide alternative plans, including multiple computer-based testing options, flexible frequency limits, and adjusted intervals between test attempts. The cultural and age factors of the examinations also need to be carefully considered, which contributes to promoting greater fairness among the participants. Furthermore, there should be a more comprehensive analysis of profiles of the examinees, so appropriate schemes and alternative test formats can be developed for emergencies, such as pandemics.

8.2 The Influence of COVID-19 on Attendacne Rates

The findings illustrate significant differences in the influence of the pandemic on attendance rates between childcare and school institutions in New Zealand.

One important distinction is that the negative effect of the pandemic was more evident in 2022 for childcare attendance rates across multiple ethnicities, whereas for school attendance rates, the negative impact was more pronounced in 2021 across various ethnic groups. Several factors may account for these findings, with a primary explanation relating to both age and policy differences. Generally, childcare attendance is more flexible [243], as parents can choose whether or not to send their children to these institutions. As a result, in 2021, the pandemic's negative effect on childcare attendance was not apparent. Parents who had already opted for at-home care likely continued this practice, which would not have caused significant changes in attendance rates. However, in 2021, the government introduced a plan to ease border restrictions at the beginning of 2022 [244], while vaccination campaigns were greatly expanded [244]. With multiple restrictions being lifted, children were allowed to interact more freely, leading to a substantial rise in infections in 2022. According to a report on hospitalisations in New Zealand, 2,101,473 COVID-19 cases were recorded in 2022—far exceeding the 12,032 cases reported in 2021 [245]. Consequently, the increased number of infections among children likely explains why the correlation results became more evident in 2022.

In contrast, school attendance is more strictly regulated, as primary and secondary education is compulsory for children aged 6 to 16 years [243]. In 2021, the pandemic may have had a stronger influence on school attendance because many students were exposed to illness and became infected. By 2022, strengthened immune responses following earlier infections may have contributed to a decline in infection rates, thereby reducing the pandemic's negative impact on school attendance.

Additionally, the pandemic's influence varied across different ethnicities. For childcare attendance, the most significant impact was observed among Asian ethnic groups, while for school attendance, the most notable impact was among Māori students.

There may be several reasons for these differences. A key reason is that during the pandemic, many parents who had previously sent their children to childcare centers chose to keep them at home in New Zealand [246], and Asian parents placed particular emphasis on safety concerns [247], resulting in a substantial decline in attendance rates and the corresponding highly visible negative impacts. Additionally, Chinese children are likely to face unique challenges related to racism [248]. Moreover, according to another study, non-English-speaking residents in New Zealand experienced difficulties in accessing timely information during the pandemic, which may have hindered their ability to respond appropriately [249].

For Māori students, limited access to healthcare and educational resources during the pandemic contributed to higher infection rates. General access control and management methods have been studied for many years [250, 251, 252, 253], and privacy and security challenges have been discussed [254, 255, 256, 257, 258]. The applications of the existing methods are still on the way in real life [259, 260, 261]. Some research has noted that Māori patients reported barriers to accessing health services [262, 263], and another study shows that nearly 50% of Māori in New Zealand require primary care [264], indicating that they may not receive timely support. Higher infection rates may have subsequently resulted in increased school absenteeism. In addition, numerous studies have highlighted persistent inequalities in educational resources in New Zealand [265, 266].

Consistent with the previous analysis presented in this chapter, the attendance rates of Māori-medium schools, where the primary language of instruction is Māori, show the strongest association with the number of new COVID-19 cases and vaccinations, which aligns with these findings.

Therefore, governments need to develop targeted policies for different ethnic groups. Since healthcare needs and available resources vary across ethnicities, and cultural backgrounds influence how services are accessed and utilized, policy responses must be data-driven. A clear statistical analysis of each ethnic group, particularly minority groups, is necessary. Appropriate emergency response plans can be established based on service availability and cultural diversity.

8.3 Implications and Future Directions

Based on the findings regarding academic performance and attendance rates, relevant institutions and stakeholders must consider multiple factors when formulating policies.

One important consideration is the magnitude of the pandemic. As COVID-19 has had a global impact, various tests and attendance rates across different ethnic groups have been affected. Therefore, it is essential to analyze both the scope and duration of pandemics—particularly global ones—when designing policies.

Furthermore, a systematic analysis of age and cultural factors in relation to examinations and educational policies is essential. Participants of different ages and cultural backgrounds exhibited varying academic performances in the TOEFL and GMAT during the early stages of the pandemic. Similarly, the impact of the pandemic also varied among different age groups within a country, as well as among ethnic minority groups.

Moreover, the effects of the pandemic differed across its various phases. As a result, it is important to develop adaptable plans. For test-takers, the impact was most pronounced at the onset of the pandemic, whereas for school-aged children and those in childcare, the effects were more evident in the subsequent years.

Chapter 9

A Systematic Review of LLMs in Education: Dual Roles, User Perspectives, and Cross-Disciplinary Impacts

9.1 Introduction

In contemporary society, AI plays an increasingly pivotal role across various sectors, including education. Among these technologies, LLMs have attracted significant attention for their potential to enhance teaching and learning processes [13]. LLMs can work as learning systems designed for language processing, capable of generating responses to user input based on extensive training data and internal knowledge representations [12, 13]. Representative LLMs include GPT models (e.g., GPT-3.5, GPT-4), while typical applications leveraging LLMs include ChatGPT, Bard, Bing Chat, and Perplexity [15].

The widespread adoption of LLMs has substantially influenced both students and instructors. Existing studies have demonstrated their application across a wide range of disciplines, such as medicine, computer science, and language education. LLMs also support instructional practices through automated grading, writing assistance, and content generation [45]. However, alongside these opportunities, concerns have emerged—particularly regarding academic integrity, overreliance, and the ethical implications of AI integration in educational contexts [267].

Despite a growing body of research, several critical gaps remain in the literature. First, many existing reviews concentrate predominantly on the medical domain, often limiting their scope to healthcare-related education. For example, studies have examined the role of LLMs in general medical training [42] or more narrowly within subfields such as dental education [49]. While these studies offer valuable insights, they fall short in providing comparative, cross-disciplinary analyses of LLM applications.

Second, there is a lack of systematic, stakeholder-centered comparisons regarding the educational impact of LLMs. Much of the literature offers generalized discussions or educator-focused evaluations, with limited attention to students' perspectives or the broader roles LLMs may assume in the learning process [50]. For instance, a global survey in dental education primarily assesses educators' perceptions, omitting learners' experiences and insights [49].

To address these gaps, this chapter investigates the adoption and impact of LLMs across multiple academic disciplines, with particular attention to their dual functions in education, as learners and as instructors. Moreover, it emphasizes both student and educator perspectives to capture a more holistic view of LLM-driven transformation in educational environments. Accordingly, this chapter proposes the following research questions: What roles do LLMs play in education, and how are their applications and impacts perceived from both student and educator perspectives across academic disciplines?

To answer these questions and contribute to the growing body of literature on AI in education, this chapter offers the following key contributions:

- It introduces a dual-role framework that systematically examines LLMs both as learners and as instructors—an analytical perspective that is currently underexplored.
- It evaluates the educational applications and limitations of LLMs across diverse academic disciplines, providing a cross-domain comparison that extends beyond existing domain-specific reviews.

- It integrates user perspectives by analyzing how students and educators perceive the impact of LLMs, highlighting both benefits and challenges.

Together, these contributions offer a comprehensive foundation for understanding the evolving role of LLMs in education and identify important directions for future research and practice.

The remainder of this chapter is organised as follows. Section 2 outlines the literature selection and screening process, including the search strategy and descriptive statistics of the selected studies. Sections 3 and 4 present the functional applications of LLMs in education, examining their roles both as learners and as instructional tools, respectively. Section 5 explores the perceived educational impacts of LLMs from two key perspectives: students and instructors, highlighting both the reported benefits and potential drawbacks. Section 6 discusses the key findings, implications, and limitations of the review. Finally, Section 7 concludes the chapter and outlines directions for future research.

9.2 Literature Selection and Screening

9.2.1 Search Strategy

This chapter was conducted following the PRISMA(Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [268]. Three major academic databases were utilized for the literature search: Web of Science, Scopus, and IEEE. To ensure a comprehensive and targeted search, relevant keywords were carefully selected based on the chapter's focus, specifically the influence of LLMs in education. The core keyword was "Large Language Model." However, as the term "education" is not consistently indexed across all journals, a broader range of education-related terms was incorporated to capture different educational levels and contexts. These included: "Student," "Primary School," "Secondary School," "University," "Academic Performance," "Learning," "Course," "Assignment," and "Feedback."

To increase the relevance of search results, terms associated with general artificial intelligence or unrelated technical fields were deliberately excluded. For instance, keywords such as "Machine Learning" and "Deep Learning" were filtered out, as they refer to subfields of AI that are not directly focused on LLMs in educational settings [269]. Additional exclusion terms included "System," "User," "Data Mapping," "Network," "Algorithm," "Cyber," "Retrieval," "Query," "Robot," "Patient," "Vehicle," and "Modelling."

The final search query used was as follows:

```
(‘‘Large Language Model’’ AND (‘‘Education’’ OR ‘‘Student’’ OR ‘‘Primary
School’’ OR ‘‘Secondary School’’ OR ‘‘University’’ OR ‘‘Academic
Performance’’ OR ‘‘Learning’’ OR ‘‘Course’’ OR ‘‘Assignment’’ OR ‘‘Feedback’’)
NOT (‘‘Machine Learning’’ OR ‘‘Deep Learning’’ OR ‘‘System’’ OR ‘‘User’’
OR ‘‘Data Mapping’’ OR ‘‘Network’’ OR ‘‘Algorithm’’ OR ‘‘Cyber’’ OR
‘‘Retrieval’’ OR ‘‘Query’’ OR ‘‘Robot’’ OR ‘‘Patient’’ OR ‘‘Vehicle’’
OR ‘‘Modelling’’))
```

Only peer-reviewed journal articles and conference papers published within the five years before March 2025 were considered. As illustrated in Fig. 9.1, the initial search yielded 1,587 articles from Scopus, 475 from Web of Science, and 316 from IEEE, resulting in a total of 2,378 records. After removing 266 duplicates, 2,112 articles remained for title and abstract screening.

Only articles published in Q1 journals or A-level conferences were retained to ensure quality and relevance. This screening process eliminated 1,671 records, leaving 441 articles for full-text review. After further assessment, 82 articles were selected for inclusion in the final analysis. Papers were excluded at this stage if they did not directly address the educational impact of LLMs. Many excluded articles focused instead on broader generative AI technologies, while others were inaccessible.

9.2.2 Statistics of Selected Papers

To provide a clearer understanding of the current research landscape, this section presents a descriptive overview of the selected papers regarding publication year, country of origin, article type, and topical focus.

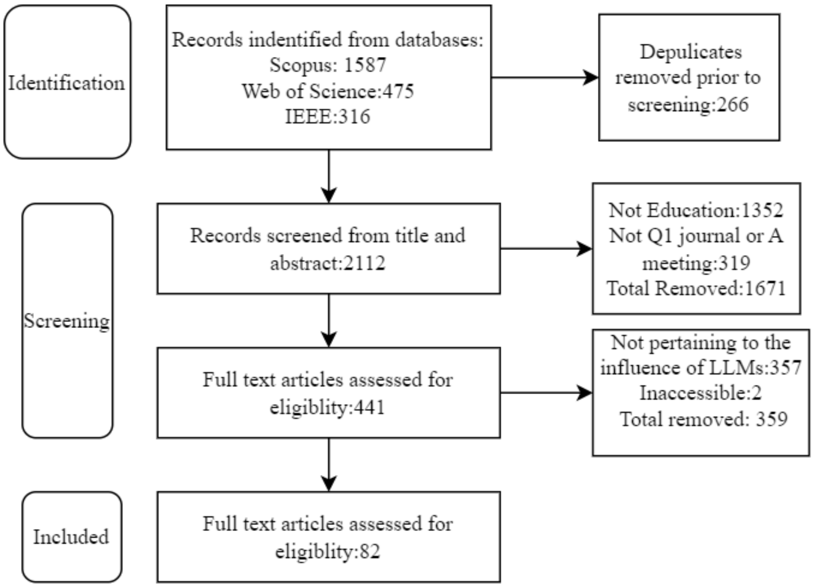


Fig. 9.1 Flow Diagram of the Article Selection Process for the Systematic Review

Fig. 9.2 illustrates the distribution of the selected articles by publication year. Among the selected articles, 59 (72%) were published in 2024, 20 (24%) in 2023, and only 3 (4%) in 2025. As the selection was limited to papers published before March 2025, only three articles from 2025 were ultimately included.

This trend aligns with the rapid development of large language models, particularly following the release of advanced systems such as GPT-4 and GPT-4o. As these models have gained significant traction in academic and educational contexts, scholarly interest in their impact on education has increased markedly since 2023.

In addition to temporal trends, Fig. 9.3 shows the geographic distribution of the first authors' affiliations. America (the U.S.) (20) and China (17) account for the largest share of publications, followed by Australia (5), Germany (4), and the United Kingdom, India, Korea, and Canada (each with 3). A broad international contribution is evident, with additional articles originating from Other countries, such as Algeria, Colombia, Croatia, Ghana, Italy, Japan, Kazakhstan, Luxembourg, the Netherlands, Norway, Peru, Poland, Qatar, Saudi Arabia, Singapore, South Africa, Spain, Turkey, and the United Arab Emirates. This wide distribution reflects the global interest in LLMs and their role in educational

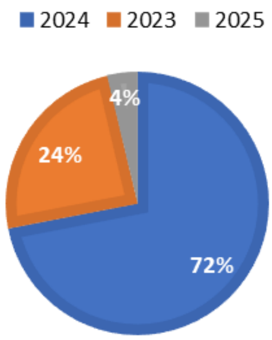


Fig. 9.2 Year-Wise Distribution of Selected Articles

transformation.

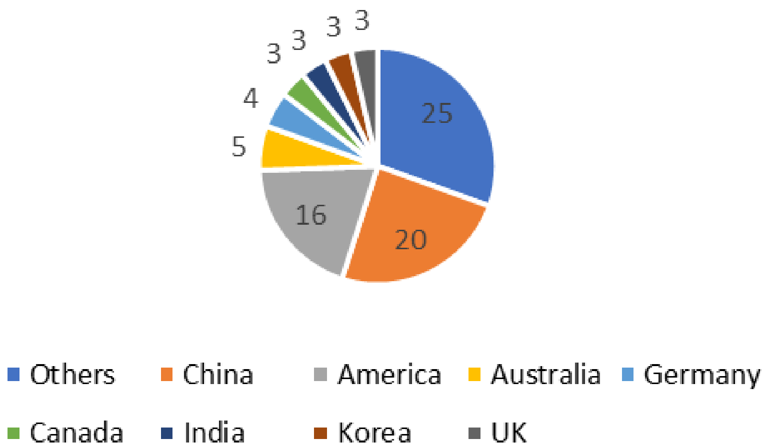


Fig. 9.3 Geographic Distribution of Reviewed Articles by Country

Fig. 9.4 presents the classification of reviewed articles by type to further characterize the literature. The majority—approximately 76 papers—are original empirical studies or exploratory research. Only 6 articles fall into the category of systematic reviews, narrative reviews, or survey papers. This imbalance highlights a clear gap in the literature: while many studies investigate specific applications of LLMs in education, few provide integrative or summative analyses that consolidate current findings.

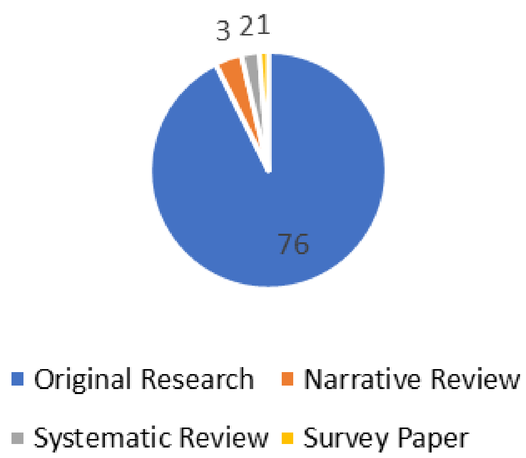


Fig. 9.4 Classification of Reviewed Articles by Type

Finally, to clarify the organizational logic of this review, Fig. 9.5 presents the conceptual structure that guides the analysis. The review is structured around two primary components: the functional roles that LLMs assume in educational settings, and the perceived impacts of these roles from the perspectives of learners and instructors.

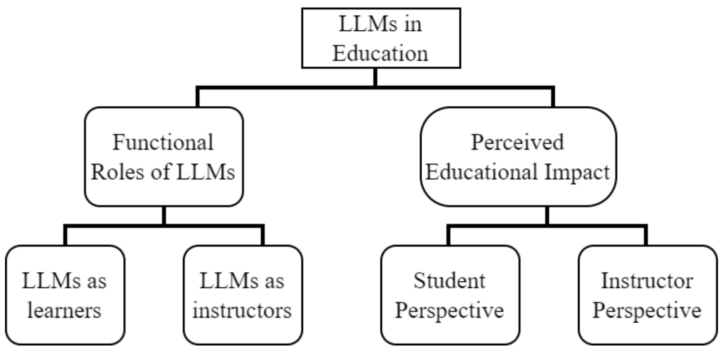


Fig. 9.5 Conceptual Structure of the Review: Functional Roles of LLMs and Their Perceived Educational Impacts

The first component, *Functional Roles of LLMs*, examines how large language models operate as both learners and instructors within educational contexts. A substantial number of studies investigate LLMs as learners, particularly in domains such as medical education, while others focus on their instructional functions, including automated grading, content generation, and feedback provision. This component is therefore divided into two subsections: LLMs as learners and LLMs as instructors.

The second component, *Perceived Educational Impact*, addresses how the integration of LLMs is experienced and evaluated by end users. This includes examining both the benefits and challenges reported by students and educators. By analyzing the perspectives of these two groups separately, the review provides a balanced and comprehensive account of LLMs' influence on educational practice.

9.3 Functional Applications of LLMs in Education

9.3.1 LLMs as Learners: Evaluating Academic Performance

9.3.1.1 Performance of LLMs in Medical Education

Recent studies have extensively examined the performance of LLMs in medical examinations and academic assessments, highlighting significant advancements. Notably, as shown in Table 9.1, GPT-4o and GPT-4 have demonstrated exceptional accuracy across multiple medical licensing examinations. For example, GPT-4o attained an accuracy rate of over 80% across most subjects in the experiment of the 48-part I sample questions of the Hong Kong Medical Licensing Examination (HKMLE) [46]. In addition, the test results show that GPT-4 consistently achieved a higher accuracy rate than 80% for all subjects in the test of 30 sample questions from part I of the Professional and Linguistic Assessment Board (PLAB) examination [46]. Furthermore, these models maintained high accuracy rates in the National Medical Licensing Examination (NMLE) [46]. Similarly, a research study on multiple choice questions in the Korean pharmacist license examination(KPLE), excluding image questions, illustrates that GPT-4 consistently achieved scores above the passing threshold [270]. Another study on the UK Medical Licensing Assessment (UKMLA) reported that the average accuracy rate of GPT in the test of 191 single-best-answer questions on the three attempts is over 75% [271].

Table 9.1 LLMs as Learners in Medicine

Exam Category	**Specific Exam**	**Strengths**	**Challenges**
Medicine Licensing	USMLE	GPT-4o and GPT-4 got 90% accuracy [46].	GPT accuracy was just over 30% for some subjects in Step 2 CK. Other LLMs performed much worse [46].
	PLAB; HKMLE	GPT-4o scored over 80% [46].	GPT-4's accuracy in medical ethics and orthopedics was only 50% in PLAB and 66% in HKMLE.
	NMLE	GPT-4o scored over 70% in nearly all subjects [46].	GPT-4, GPT-3.5, and Bard scored below 70% [46].
	NPLE; NNLE		ChatGPT did not pass NPLE and NNLE [62].
	NCLEX-RN	GPT-4 scored over 79% for both Chinese and English NCLEX-RN practical questions [272].	
	Japanese Medical Licensing Exam	Proficient in text-centric questions [63].	Not proficient in image questions [273].

Continued on next page

Table 9.1 – *Continued from previous page*

Exam Category	**Specific Exam**	**Strengths**	**Challenges**
	KPLE	GPT-4 consistently achieved passing scores [270].	GPT-3.5 did not reach the 60% passing threshold in 1 of 3 years [270].
	UKMLA	GPT-4 answered nearly 3/4 SBAS correctly [271].	Did not perform well in clinical hematology and other subjects [271].
Biomedical	Graduate-level exams in biomedical sciences	Not lower than students' level in 7 out of 9 subjects [274].	Below average in figure questions [274].
	BMAT		Correct response rate was significantly lower than incorrect answers in Section 2 [275].
Specialized Assessments	ESNR Exams	GPT-4 scored 70% [276].	Other models did not pass all four exams [276].
	RDE	Accuracy of GPT-4 was up to 72% [277].	Inconsistent performance in dental field questions [277].
	Orthopedic Education	Scored higher on some subjects compared to the control group [278].	

Continued on next page

Table 9.1 – *Continued from previous page*

Exam Category	**Specific Exam**	**Strengths**	**Challenges**
	UK Postgraduate Medical Exams	Performed above passing level on most exams [279].	GPT-3.5 failed eight out of nine exams [279].
Other	Psychosomatic Medicine Exam Questions	Scored over 90% [280].	Errors mainly in the "remember" and "understand" cognitive levels [280].
	National Premedical Exam	GPT-4 outperformed other models in some subjects [281].	Limited accessibility due to financial resources [281].
	Polish Medical Final Exam	GPT-3.5 passed 2 out of 3 versions; GPT-4 passed all versions [282].	Low accuracy in ecosystem and environmental topics [281].
	Neurology Specialist Exam		GPT-3.5 did not pass [283].
	Taiwan Audiologist Exam	Both proficient in grasping fundamental concepts [284].	Information source errors [284].
	Higher Education Courses	ChatGPT Plus and CoPilot-Bing scored highest on MCQs [67].	
	ACR TXIT Radiation Oncology Exam	GPT-3.5's accuracy rate is over 63%; GPT-4's over 74% [67].	Limited knowledge in gynecology and other sectors [67].

Continued on next page

Table 9.1 – *Continued from previous page*

Exam Category	Specific Exam	Strengths	Challenges
	K-CBMSE		Performance dependent on multiple factors [67].
	Physiology and Biochemistry Questions	GPT-4 performs better in biochemistry [285].	

Beyond medical licensing assessments, LLMs, mainly GPT-4, have also demonstrated strong performance in biomedical-related examinations and other specialized courses and assessments. For example, in a research study on the biomedical science graduate examinations for corresponding scientific trainees, GPT models outperformed or matched human participants in seven out of nine subjects, including virology, microbiology, and some other courses [274]. Likewise, GPT-4o exhibited a 72% accuracy rate in 151 Restorative Dentistry and Endodontics (RDE) Student Assessment Questions [277], while GPT achieved an accuracy rate of up to 70% in European Society of Neuroradiology (ESNR) examinations [276]. Another research study on the assistance of ChatGPT showed that students who adopted ChatGPT performed much better than those who were prohibited in both the long-term and short-term Orthopedic Education assessments [278].

The accuracy of LLMs remains notably high across other examinations, including multiple postgraduate medical examinations and specialization certifications. In the United Kingdom's Postgraduate Medical Examinations, GPT achieved an accuracy rate exceeding the passing threshold for the majority of exams without additional fine-tuning [279]. Similarly, GPT-4 achieved a success rate surpassing 90 % based on Bloom's taxonomy in 307 multiple-choice questions of the Psychosomatic Medicine Exam[280]. Researchers of the Polish Medical Final Examination (MFE) found that GPT-4 successfully passed all three editions in both English and Polish versions, with the mean accurate rate of 79.7% [282]. Meanwhile, findings from the American College of Radiology (ACR) radiation oncology in-training (TXIT) examination indicate that GPT-4 achieved an accuracy rate exceeding 70% [286]. Both GPT-3 and GPT-4 completed the tests in undergraduate medical examinations, with GPT-4 exhibiting a higher accuracy rate [285]. Moreover, a comparative study between ChatGPT and neurologists in Spain revealed that GPT-4 outperformed certain neurologists in an examination [283].

In addition to high accuracy rates, LLMs excel in text-based and standardized questions. A study on the Japanese medical licensing examination indicated that GPT-4 was particularly proficient in text-based questions [63]. Similarly, research on ESNR examinations suggested that LLMs perform more effectively on lower-order cognitive questions, which are well-suited to text-based responses due to their fixed-answer nature [276]. Additionally, a study on the Taiwan Audiologist Qualification Examination demonstrated that both GPT-3.5 and GPT-4 effectively understood key concepts in radiation oncology [284]. Further, in graduate-level biomedical science examinations, GPT-4 exhibited strong performance in fill-in-the-blank and short-answer questions [274].

Another key advantage is their efficiency in processing examination questions. ChatGPT Plus, CoPilot-Bing Precise, Gemini Pro, Llama 3.1405B, Mistral Large, and Claude 3 Opus completed the tasks significantly efficiently than humans in multiple-choice questions and final examination questions from an introductory biomedical and health informatics course [67].

However, despite these advancements, several limitations persist in the application of LLMs to medical examinations and assessments. One is the inconsistency in the performance. For example, GPT-3.5 and Google Bard demonstrated accuracy rates below 60% in both the USMLE, PLAB [46] as well as NMLE [62]. Additionally, another study found that GPT-3.5 and Google Bard performed poorly in the ESNR examinations, with results approaching failure [276]. Similarly, GPT-3.5 did not pass UK postgraduate medical examinations [279].

Second, LLMs' performance varies across specific topics and datasets. Some limitations were observed in PLAB, where GPT-4 demonstrated reduced accuracy in medical ethics and orthopedics-related questions [46]. In UKMLA, LLMs struggled with specialized topics such as gynecology, brachytherapy, and dosimetry [284], while even GPT displayed limitations in clinical hematology and some other topics [271]. Additionally, GPT exhibited suboptimal performance on simulated data in nine graduate-level biomedical science examinations for the corresponding trainees in around 150 questions [274]. Moreover, performance inconsistencies were identified in dental-related assessments [277] and radiation oncology in-training examinations [286].

Third, LLMs' performance is highly dependent on the nature of the questions posed. Studies indicate that LLMs struggle with image-based questions. In the Japanese Medical Licensing Examination, response rates for image recognition tasks were significantly lower than for text-based questions [63]. Likewise, the introduction of image-based questions in the Japanese National Medical Licensing Examination led to a measurable decline in overall accuracy rates [273]. Furthermore, when responding to higher-order thinking questions, LLM-generated answers tended to be less accurate [66].

9.3.1.2 Performance of LLMs in Other Academic Domains

In addition to the research on medicine, as shown in the Table 9.2, LLMs also demonstrate great breakthroughs in multiple disciplines, including programming, linguistics, engineering, physics, and others [43, 44]. In programming, for example, LLMs such as GPT enable students to quickly generate solutions and engage with complex analogies [287]. Moreover, LLMs exhibit a robust response rate to specific standardized questions in undergraduate computer science education. [288].

There are also some breakthroughs of LLMs in language and engineering studies. In English studies, the effectiveness of LLMs is particularly notable. A study in the Netherlands revealed that GPT-4 significantly outperformed high school students in English reading comprehension exams, whereas GPT-3 performed comparably to students [290]. This trend extends to engineering and physics education as well. For instance, in India, research in computer science education (CSE) demonstrated that GPT achieved a 90% consistency rate [20]. Similarly, studies in physics education suggest that GPT's problem-solving capabilities in physics are comparable with human performance [68].

Despite these successes, LLMs face several limitations. One notable concern is the inconsistent reliability of the content generated, particularly in programming [287, 291]. Additionally, in engineering, research on undergraduate com-

Table 9.2 LLMs as Learners in Other Subjects

Subject	Strengths	Challenges
Programming	Promising ability to rapidly create and engage with analogies [287].	Specific prompts influence the range of topics [287].
	High accuracy for subjective and theoretical questions [288].	Varying accuracy rate in different domains [288].
	ChatGPT can subdivide questions [289].	Course difficulty and chat lengths influence results [289].
Language	GPT-3.5 performs similarly to average Dutch students; GPT-4 outperforms humans with reprompting [290].	
Engineering	91.87% concordance for accurate explanations in CSE [20].	Poor performance in GATE and JEE [20]; temperature setting significantly influences results [21].
Physics	Ability to solve some physics calculation problems up to human level [68].	
Law		Low correct response rate in LNAT [275].

puter science assessments indicates that different LLMs exhibit biases, potentially impacting performance in exams such as GATE(Graduate Aptitude Test in Engineering) and JEE(Joint Entrance Examination) [288], while accuracy rates remain relatively low in standardized tests LNAT(Law National Aptitude Test), where correct response rates in these assessments are suboptimal [275].

Another challenge lies in the variability of results, as LLMs' performance often depends on question types and subjects. For example, a study on programming analogies suggests that when the prompts are highly specific, the generated analogies tend to cover a wider range of topics [287], highlighting the context-dependent nature of LLM performance. In undergraduate science education and certain programming examinations, accuracy rates vary considerably across different domains [288, 291]. Another study on data science education revealed that prolonged chat sessions with GPT are more likely to produce blank responses [289], while environmental engineering indicates that temperature settings can influence the precision of GPT responses [21]. Lastly, a linguistic study has shown that LLMs' correction capabilities in Chinese are limited to the word level, with little improvement in contextually complex sentences [292].

9.4 LLMs as Instructors: Supporting Teaching and Assessment

LLMs play a pivotal role in course instruction across the globe. As shown in the Table 9.3, their versatile functionalities encompass a wide range of educational tasks, including automated essay scoring, writing assistance, provision of feedback, teaching support, and the generation of questions.

Table 9.3 LLMs as Instructors

Functions	Strengths	Challenges
Automatic Scoring	GPT-4 is the most accurate in annotation and prediction of learning levels. The agreement coefficient between GPT-4 and human raters outperforms that among human raters [293].	The length of the requirements influences scoring [294].
	GPT-3.5 is more accurate than BERT in multiclass tasks [295].	The architecture and amount of training data influence the final scoring [295].
Continued on next page		

Table 9.3 – continued from previous page

Functions	**Strengths**	**Challenges**
	PALM2, GPT-3.5, and GPT-4 illustrate great reliability in automatic scoring [296]. GPT-4 outperforms others in interrater reliability due to a fuller range of rubric scores [296].	Some LLMs rely on a continuous scale [296].
	High efficiency in scoring binomial and trinomial tasks [297].	Not appropriate for course-related lectures [298].
	More inclined to avoid extreme scores [298].	Inconsistency in grading grammar, organization, content, and mechanics [299].
	Proficient in distinguishing varying levels of sophistication [299]. No significant differences between GPT-4 evaluations and teachers' evaluations [300].	
Writing Assistance	Increases engagement [301]. Provides complete ChatGPT feedback [303]. Provides positive comments and highlights improvement areas [303].	Accessibility issues [302].
	Not limited by topics; provides feedback on key evaluation points [302].	Limited to the linguistic level, lacking a humanistic approach. Relies on previous texts, lacking human experience [302].
Feedback	High speed of generation [45].	Feedback is inconsistent across different versions [45].
	Improves scalability and consistency of feedback [304].	Over 30% of hints are incorrect [305]. Difficulty in distinguishing the correct code [291].

Continued on next page

Table 9.3 – continued from previous page

Functions	Strengths	Challenges
	Feedback by GPT is significantly more readable for open-ended tasks [47].	Most feedback is positive, lacking negative ones [47].
	Ability to generate process-focused feedback [47].	
	AI-enhanced tool "Typography Evaluator" has similar output to human ones [306].	Overcorrects at the word level, ignoring corresponding context (CGEC) [292].
Instructor Development	Creative with automated assessment features [307].	Accuracy and relevance issues. Difficulty in converting manual assessments into electronic ones [307].
	Contributes to clinical practical skills and human-computer interaction [308].	Limited medical knowledge affects accuracy [308].
Question Generation	No statistically significant difference in difficulty between questions generated by GPT and humans [309].	Only 50% of questions are deemed usable without modification [310].
	Ability to generate examination questions of equivalent quality to expert-generated ones [311, 312].	Misguidance in medical examinations [311].
	Improved quality, appropriate for large-scale applications [312, 313].	ChatGPT-3 cannot improve question-answer pairs and automated summaries [314].
	Potential to generate multiple-choice question examinations much faster [315].	
	Potential to bolster programming education [313].	

GPT-based techniques generally achieve the best performance among various LLMs for automatic scoring. Research on non-native Japanese learners demonstrates that the GPT-4-based method outperforms BERT and OCLL in automatic scoring [293]. Two other studies indicate that the average accuracy rate of GPT-3.5 surpasses that of BERT [295] and GPT-4 outperforms GPT-3 [297]. Moreover, the capability of automated scoring is comparable to manual grading, as supported by several studies. EFL(English as a foreign language) essay grading demonstrates that GPT produces explainable and interpretable outcomes and distinguishes different levels of sophistication [299], while another indicates no significant differences between evaluations generated by GPT-4 models and manual scoring based on the mini papers from 25 third-year Chinese medical students[300].

Despite these advantages, LLMs also present certain limitations in automatic scoring, primarily related to technical constraints. For instance, a research study on undergraduate medical education and another study based on multiple assessment tasks in middle and high schools suggest that both the length of input data and the amount of training data significantly influence the results [294, 295]. Additionally, one important reason is that multiple large language models only rely on the continuous scale [296], while the effects of marking general questions are more appropriate for course-related lectures and not consistent in different tasks [298, 299].

In the context of writing assistance and feedback provision, both have their respective strengths and weaknesses. Feedback generated by LLMs tends to be more comprehensive, with comments that are more positive while also identifying areas for improvement [302]. This enables a large number of students to report having a mixed emotional experience when receiving feedback from LLMs, with the results being more acceptable to students [47, 301]. The response rate is significantly faster [45]. In a study on EFL writing, ChatGPT produced substantially more feedback under identical conditions compared to human graders [303], thereby improving the stability and consistency of the feedback [304]. Moreover, based on a study in higher education, GPT-generated feedback is generally more readable for open-ended tasks, making it more easily accepted by students [47]. Additionally, researchers in a Typography course found that the outputs from the AI-enhanced tool, Typography Evaluator, were very similar to manual evaluations [306], and some experts observed that most of the feedback adheres to effective feedback principles [304].

However, LLMs also have limitations. First, GPT-generated feedback is largely confined to standardized text structures, while affective feedback is limited to linguistic aspects [302]. Second, ChatGPT's feedback is not consistently accurate. For instance, different versions of the model show variability in written corrective feedback [45], and most of the feedbacks from ChatGPT tend to be very positive, lacking in comprehensiveness [47]. A research study on learning gains in mathematical skills illustrates that over 30 % of the hints provided are inaccurate [305], while GPT-3's code correctness rate was over 70% based on 1211 student code submissions in a programming exam [291].

LLMs also influence the instructor's professional development, presenting both opportunities and challenges. One research study on implementing ChatGPT to improve teaching skills in English illustrates that chatbots are more creative [307], which may enhance professional development, while another study in nuclear medicine research in Chongqing Medicine University shows that LLMs benefit the interaction between humans and computers [308]. However, many instructors are still concerned about the accuracy and relevance issues of LLMs [307], because the pre-training sets of the medical knowledge may influence accuracy rates and result in misuse behaviors [308]. Moreover, it is challenging to convert some manual versions into electronic ones in actual practice [307].

In addition to these functions, another important adaptation of large language models is the generation of learning materials. Regarding question generation, the quality of questions generated by LLMs is comparable to human-generated ones. A study in medical education indicates that the difficulty level of the 25 AI-generated multiple-choice questions aligns closely with those created with a medical educator [309]. Similarly, 44 dental board-style examination questions produced by LLMs based on the corresponding textbook contents are of the same quality as those developed by human experts [311]. Moreover, LLMs generate questions at a much faster rate [315], and a study on introductory programming education shows that LLMs have the potential for large-scale question analysis, as the code-tracing generated in the experiments is of high quality [313], while automated question generation can improve the quality of assessments [312].

Nonetheless, the accuracy of AI-generated questions remains a crucial factor that requires further consideration. A study on exam generation reflects that despite the ability to generate multiple-choice questions, around half of the questions generated by ChatGPT need to be modified [310], while another illustrates that the ability of ChatGPT to improve question and answer pairs still needs to be enhanced, as the quality of the automated summaries cannot be guaranteed [314]. In addition, LLMs still need to be supervised in real practice because they cannot replace human interactions [311].

9.5 Perceived Educational Impacts of LLMs

9.5.1 Impacts from the Student Perspective

9.5.1.1 Perceived Benefits for Students

As shown in Table 9.4, LLMs primarily enhance students' critical thinking, argumentation, and language skills while also fostering motivation and improving their overall learning experience.

The integration of LLMs significantly strengthens students' comprehension and argumentation abilities. For instance, a study involving 95 undergraduate students found that LLMs help students grasp complex concepts more effectively while simultaneously improving their argumentation skills in international relations debates [316]. Similarly, an experiment comparing two groups of Ghana university students demonstrated that the incorporation of LLMs into education led to improved critical thinking scores among undergraduates [317]. Another study involving 67 first-year university students showed that ChatGPT significantly enhanced argumentation performance across various English proficiency levels [318].

LLMs also play a crucial role in improving language proficiency and stimulating student motivation. Research on 25 Chinese medical students highlights that ChatGPT facilitates translation and grammar correction while supporting literature review processes, with manual scoring increasing by more than four points following assistance from GPT-3.5 and GPT-4 [300]. Additionally, both GPT and Bard have been shown to assist 47 medical students in generating essays that reflect their perspectives on ethical issues, fostering more structured and coherent writing [319]. A study on the capabilities of ChatGPT in English

Table 9.4 Positive Influence of LLMs on Academic Skills of Students

Skills	**Positive Influence**
Critical Thinking	Improvement of students' understanding of complex concepts, critical thinking, and argumentation skills [316]. A significant increase in critical thinking scores. Improvements in creative thinking skills and reflective thinking skills [317]. ChatGPT-CA approach enhances argumentative speaking performance, critical thinking awareness, and collaboration tendency for different English levels [318].
Language Skills	Improved papers in terms of the structure, logic, and total score [300]. AI-generated essays' language is more affect-related and tends to be more emotionally positive and authentic [303, 319].
Study Experience	Improved research engagement and clinical skills development [42, 48, 50]. Increased knowledge and understanding with much wider options [42, 48, 50]. Augments the learning process in concept validation [42]. Better understandings of ethical theories[320].

language studies further illustrated that feedback provided by LLMs enhances student engagement by increasing receptiveness to constructive criticism, ultimately strengthening learning motivation [303].

Moreover, LLMs contribute to a more personalized and enriched learning experience. Studies in healthcare and medical education suggest that LLMs enable personalized learning, enhancing students' comprehension of complex concepts [42, 48, 50]. Research on the integration of GPT in education and healthcare indicates that this tailored approach sustains student engagement and deepens their understanding of specialized medical terminology [42]. Furthermore, a study examining students' perspectives on the integration of LLMs in professional ethics assignments suggests that these models facilitate a deeper understanding of ethical theories [320].

9.5.1.2 Perceived Drawbacks for Students

As depicted in Table 9.5, the negative effects of LLMs in education primarily fall into two categories: skill development and ethical concerns.

Table 9.5 Negative Influence of LLMs on Academic Skills of Students

Aspects	**Negative Influence**
Skills	Overreliance issues in writing papers or preparing for examinations [48].
	Overreliance may misguide students, so they will not undergo the enculturation and cognitive development [321, 322].
Ethical Concerns	Academic integrity [48].
	Bias and transparency [48].
	Data privacy and security [48, 322].
	Limited knowledge and training [322].
	Cannot identify ethical issues [323].

A significant challenge associated with LLMs in skill development is students' overreliance on these models. Research on the complex role of LLMs in medical education indicates that students may complete assignments and prepare for

examinations solely based on LLM-generated responses [48], potentially undermining their literacy skill development as they become overly dependent on these tools [322]. Moreover, the risk of misinformation poses an additional concern, as students may be misled by incorrect or misleading information produced by LLMs, further hindering their learning process [321].
Ethical concerns also constitute a critical issue in LLMs. One major challenge is academic integrity, as LLMs increase the risk of plagiarism, with some students generating entire assignments using multiple LLM platforms [48]. Data privacy is another pressing concern, as personal information shared with LLMs for academic guidance may be exposed to third parties [48]. Furthermore, LLMs lack human emotions, well-being considerations, and moral reasoning, leading students to make misguided decisions, particularly in ethical and legal contexts [322]. Additionally, technical limitations in input processing can result in inaccuracies, further increasing the risk of students making incorrect decisions based on flawed outputs [323]. Consequently, LLMs are not equipped to assist students in developing ethical judgment, making it more challenging for them to recognize and navigate ethical dilemmas [323].
Beyond the inherent characteristics of LLMs, students' acceptance levels also influence their usage and the associated effects. A study on clinical medical students in China found that attitude is a key predictor of students' behavioral intentions to adopt LLMs [324]. Similarly, research conducted in Peruvian universities highlights that effort expectancy and behavioral intention significantly impact students' actual adoption of LLMs [325]. Additionally, a study on Chinese EFL learners demonstrates that students' intention to adopt LLMs directly influences their engagement with these models [326].

9.5.2 Impacts from the Instructor Perspective

9.5.2.1 Perceived Benefits for Instructors

As illustrated in Table 9.6, LLMs contribute significantly to instructors' knowledge acquisition, research practices, teaching methodologies, and the preparation of instructional materials.

Table 9.6 Positive Influence of LLMs from the Perspective of Instructors

Knowledge Acquisition	**Research Practice**	**Materials Preparation**	**Teaching Practice**
Facilitation of information gathering [49].	Support research activities, including manuscript writing and experimental design [49].	Reduction of teachers' cognitive load [307] and supplement educators' feedback provision [303].	Build up more suitable pedagogies and effective instructional methods [42, 48].
Culturally relevant and personalized knowledge [287] improves understanding of real-life scenarios [327] and healthcare concepts [42].	Support medical data retrieval, literature summarization, and dataset analysis [48].	Recently introduced image recognition functions facilitate automated assessment [307].	Students had higher submission rates with Wisdom Bot [328].
		Rapid preparation for course materials [42, 323].	

LLMs enhance instructors' knowledge acquisition by facilitating information retrieval and promoting critical thinking and analytical skills. For instance, a global survey of dental educators highlights that LLMs aid in information gathering and foster deeper analytical engagement [49]. Additionally, the recursion analogy study involving over 350 first-year computing students indicates that LLM-generated content is more memorable for instructors due to its cultural relevance [287]. Furthermore, a comprehensive review of healthcare education demonstrates that LLMs contribute to a clearer understanding of multiple concepts within the field [42]. Similarly, research on nursing education shows that personalized knowledge delivery through ChatGPT enhances students' ability to comprehend real-life applications of theoretical concepts [327].

Beyond knowledge acquisition, LLMs play a crucial role in supporting instructors' research activities. The global survey of dental educators underscores how LLMs streamline literature searches, facilitate experimental design, and enhance various research processes, thereby improving efficiency by reducing time and effort [49]. Likewise, a study in medical education demonstrates that LLMs are effective in analyzing datasets and assisting with manuscript editing, further optimizing research workflows [48].

LLMs also enhance daily teaching practices by enabling personalized feedback. Research in medical education illustrates that LLMs support personalized learning experiences by adapting to individual students' needs [48]. Another study emphasizes that ChatGPT-generated feedback can be customized to align with the specific requirements of both instructors and students, improving the overall learning process [42].

Instructors benefit from LLMs in the preparation of teaching and learning materials. A study on ChatGPT's application in English language teacher training demonstrates that automated assessments reduce instructors' workload, while image recognition capabilities and file upload functions enhance content creation efficiency [307]. Similarly, research on ChatGPT's role in EFL writing instruction highlights how LLMs assist in generating feedback and refining instructional content based on students' initial responses [303]. In medical education, the efficiency of material preparation is significantly improved due to LLMs' rapid processing capabilities [42]. Additionally, research on ethics education indicates that LLMs effectively summarize complex ideas and concepts, thereby supporting educators in content development [323].

The integration of LLMs into instructional processes extends beyond material preparation, positively influencing actual teaching practices. For instance, LLMs facilitate personalized learning by generating feedback and suggestions that align with both instructors' pedagogical objectives and students' learning needs [48]. This adaptability allows educators to design more tailored curricula [42]. Furthermore, experimental studies suggest that the incorporation of LLMs in the learning environment may lead to higher student submission rates, indicating increased engagement and participation [328].

9.5.2.2 Perceived Drawbacks for Instructors

For instructors, there are three main limitations of LLMs: technological constraints, potential impacts on student motivation, and ethical concerns, as shown in the Table 9.7. While the techniques underlying LLMs still require further development, concerns have also been raised regarding their influence on student motivation and associated ethical issues. Some experts argue that educators and universities should establish policies to address these potential challenges [329], and guidelines have already been developed to assist educators in leveraging LLMs effectively [330].

Table 9.7 Negative Influence of LLMs from the Perspective of Instructors

Technique Issues	**Study Motivation**	**Ethical Concern**
Hard to find specific matching models and high costs to access [49].	Lengthy feedbacks strengthen anxiety [303].	Potential biases and errors [49].
No clear prompt results in ambiguous outputs [287].	Overreliance on LLMs influences cognitive development and study skills [323, 327, 331].	Breach of academic integrity [327, 332].
Longer arguments lead to lower precision and recall rates [302, 303, 327].	Lacking in enculturation and cognitive development [322].	Concerns towards ChatGPT being listed as an author [332].
The diagnostic errors, such as neglecting disease histories [42].		Difficult to identify the AI scoring and manual scoring [298].
Challenges of integration into the education system [42].		Enforcing fairness and privacy issues [42, 284, 323].

One major limitation is technological shortcomings. Since LLMs generate responses based on pre-existing databases, they often fail to provide adequate answers to novel or complex questions posed by lecturers and students [49]. A study in dental education highlights that ChatGPT may lack sufficient medical knowledge, leading to diagnostic errors with potentially serious consequences [42]. Additionally, the quality of LLM-generated outputs is often inconsistent. For instance, in EFL writing, responses tend to be overly lengthy, making it difficult for students and instructors to extract key points [303], while in the study on recursion analogies by 350 first-year computing students, responses were sometimes ambiguous due to unclear prompts [287]. Furthermore, while another research study on the feedback of ChatGPT on the argumentations of the undergraduate students showed that accuracy remained low when there were lengthy arguments [302]. These limitations largely stem from issues related to training datasets. Several studies in healthcare education illustrate that LLM accuracy is restricted to the scope of existing databases, making response quality unreliable [42, 327]. For example, LLMs struggle to distinguish ethical issues effectively [323].

Moreover, experts worry that technological limitations can negatively influence study motivation. Some students may find lengthy feedback overwhelming and find the key points difficult, leading to decreased motivation in writing tasks [303]. Additionally, instructors are increasingly concerned about the cognitive development and academic literacy skills of students who heavily rely on LLMs, as these students may lack the necessary experience of enculturation in academic disciplines [322, 323, 331], while some students may even generate answers directly rather than engage in critical thinking [327, 328, 331].

Ethical concerns are another critical issue raised by instructors. Each large language model has its own perspectives based on the existing databases, which may result in potential biases when providing solutions to certain questions [49]. Academic integrity issues, including plagiarism and cheating, have become more prevalent with the use of LLMs [327]. Detecting AI-generated work remains challenging in practice, as some students may use anti-plagiarism software to lower repetition rates [298], raising concerns about the potential misuse of LLM technology [332]. Additionally, debates persist over whether AI-generated content should be credited as authorship [332]. Issues related to fairness and privacy also remain unresolved [42, 284, 323]. For instance, in many countries, both instructors and students lack access to free versions of certain models, exacerbating educational disparities [323]. Furthermore, there is a high likelihood that personal information may be exposed to third parties through the adoption of LLMs [42, 284, 323].

Furthermore, educators' familiarity with LLMs significantly influences their effectiveness in education. A survey of K-12 teachers reveals that a substantial proportion of educators lack sufficient knowledge about LLMs, with some even unaware of whether their students utilize them [333].

9.6 Discussion

Existing studies demonstrate that LLMs, particularly advanced versions of GPT, have achieved significant breakthroughs across various disciplines. Notably, GPT-4 excels in answering discipline-specific questions, showing high accuracy in medical studies, programming, linguistics, and physics education. LLMs perform exceptionally well on numerous medical exams, including USMLE, HKLME, NMLE, ESNR, RDE student assessments [46, 62, 276, 277], certain programming assignments, and language exams [20, 290]. Their accuracy is especially evident in standardized assessments, including multiple-choice and text-based questions, and they demonstrate a strong understanding of fundamental concepts [63, 67, 288].

In addition to their high accuracy as learners, LLMs also demonstrate impressive performance as instructors. They enhance efficiency by generating high-quality course materials and feedback at scale, significantly reducing preparation and grading time [42, 304]. The volume of feedback generated by LLMs often exceeds that of human instructors, maintaining comparable quality and personalization when provided with detailed inputs [304, 334].

Several factors contribute to the performance of LLMs in education. One key reason is the advanced technology underlying them, including enhanced knowledge bases and newly introduced features such as image recognition [307, 335]. Additionally, the evolving attitudes of educators, many of whom are increasingly applying LLMs in grading and other tasks [293, 296, 309], have fostered greater acceptance and adaptability of these tools.

Despite these advancements, LLMs face limitations that hinder broader and more reliable applications. A major challenge is inconsistent accuracy, as evidenced by failures to pass exams such as NPLE and NNLE [62]. Moreover, when acting as learners, LLMs' effectiveness is heavily influenced by input contexts and subjects. Research shows that ambiguous or incomplete queries often lead to inaccurate responses [276], and their performance declines in complex reasoning, image-based questions, and domain-specific inquiries [63, 273].

The issues primarily arise from the limitations of the underlying databases. While LLMs possess vast knowledge, they often lack the most recent information, particularly in fields such as medical practice and advanced areas. To enhance the quality of answers, developers should broaden the knowledge base to include recent developments, verify the accuracy of existing information, and, if possible, introduce new features, such as enhanced analytical and graphical capabilities.

Regarding the impact of LLMs on both students and instructors, LLMs have positively influenced students' study skills and overall learning experience. They help students gain a deeper understanding of concepts, enhancing skills such as argumentation, language proficiency, and critical thinking [300, 316], as well as improving their grasp of key ideas [320]. For educators, LLMs support research and the preparation of course materials, enabling the creation of more personalized, culturally relevant curricula that are tailored to students' needs [42, 48, 287].

However, challenges remain from both students' and instructors' perspectives. One concern is overreliance on LLMs. Excessive dependence on AI reduces students' engagement in analytical reasoning and problem-solving [48, 321, 322, 336]. Additionally, LLMs, lacking an understanding of real-world human interactions, struggle with addressing complex moral and social issues effectively [42, 323], and this concern is heightened by the constant accessibility of AI tools [322].

Another significant challenge is ethical concerns, particularly regarding academic integrity. The adoption of AI-generated content raises issues of plagiarism and dishonesty [284]. Some students submit LLM-generated assignments without modification, making it difficult to differentiate between original and AI-assisted work. While plagiarism detection tools exist, they often struggle to reliably identify AI-generated content [298]. However, these ethical concerns are not confined to specific educational institutions, as they are closely associated with international guidelines and standards. For instance, they are directly related to the key principle of "Fairness" [337], since inappropriate adoption of LLMs may compromise the integrity of daily academic activities across multiple institutions, which designers and researchers must carefully consider. Additionally, the principle of "Sustainable Development" [337] represents another critical OECD guideline, emphasizing that AI technologies should inform the long-term development of the younger generation, but a limited understanding of AI systems among students may hinder their future educational growth.

Both positive and negative influences are present across various AI tools. Educators note that AI software can contribute to cheating, with some schools even considering banning its use in daily studies [338]. Conversely, others view it as a tool to reduce educators' workload.

In response, a systematic evaluation of LLMs' adoption is necessary to assess both their potential and risks. Institutions should implement detailed policies on LLM applications rather than simply banning them. Furthermore, experts should guide students and instructors on how to effectively use LLM functionalities.

Despite these findings, this chapter has limitations. For instance, some newer LLMs, such as DeepSeek, were not analyzed. Additionally, most existing studies did not account for the cultural background of these models, which could influence the accuracy of responses. The effectiveness of LLMs may also vary across different educational levels, including primary, secondary, high school, and university settings.

9.7 Conclusion

This chapter provides a comprehensive analysis of how LLMs, particularly GPT-4 and GPT-4o, are currently applied and perceived in educational settings. These models demonstrate high accuracy across a wide range of disciplines—including medicine, programming, linguistics, engineering, and physics—particularly excelling at answering standardized, text-based questions. In addition to their learning performance, LLMs also show strong potential as instructional tools, especially in automated scoring, feedback generation, and content creation.

Despite these promising capabilities, several limitations remain. LLMs continue to face challenges with handling multimodal input (e.g., image-based questions), and their performance is sensitive to prompt clarity and contextual variation. Furthermore, ethical concerns such as academic dishonesty, overreliance, and reduced motivation for independent learning present ongoing issues that require critical attention.

A key contribution of this chapter lies in its dual analytical lens: it examines LLMs not only in terms of their functional roles—as learners and as instructors—but also in terms of their perceived educational impact from both student and educator perspectives. This dual-structured framework enables a more nuanced understanding of how LLMs are integrated into, and affect, different dimensions of teaching and learning.

Notably, current literature has yet to fully investigate the educational applications of newer models such as DeepSeek, nor has it sufficiently explored how cultural and educational diversity may shape the use and impact of LLMs. Future research should therefore prioritize emerging models and consider cross-cultural, cross-level comparisons to guide the responsible and effective integration of LLMs in global education systems.

Chapter 10

Conclusion and Prospect

In conclusion, this thesis explores the impacts of the COVID-19 pandemic and large language models on various aspects of education, including multiple standardized tests, attendance rates, and a wide range of other factors.

The research demonstrates that the pandemic significantly altered overall academic performance at the onset of the crisis; however, the trends vary across different types of international tests. The skills assessed in each examination, the requirements of home-based online tests, and the urgency experienced by test participants may all contribute to the observed differences in academic performance. Therefore, the thesis provides valuable insights for policymakers, suggesting that testing policies should be flexible and tailored to the nature of each test and the needs of the students.

Furthermore, the pandemic had a substantial impact on attendance rates in New Zealand across different ethnic groups and types of Māori schools. However, the years during which childcare and school attendance rates were most affected differ, as does the ethnicity most impacted by the pandemic. These variations may be attributed to differences in educational systems, cultural backgrounds, and diverse knowledge systems. Consequently, it is essential to understand the cultural context of each ethnicity, as well as their access to health services and financial resources. Governments and policymakers should develop targeted policies to ensure equitable access to services across all ethnicities and establish contingency plans within education systems to respond effectively to future pandemics.

Therefore, it is important to maintain the perspective of the pandemic as a global one because the pandemic influences all countries worldwide, while cultural factors, age factors, and multiple other factors need to be taken into consideration when setting up policies. The phases of the pandemic are also very important.

Moreover, LLMs play an increasingly significant role in education, offering strong performance in various examinations and functioning effectively as virtual tutors. The findings from the perspectives of students and educators reveal that large language models achieve high accuracy rates in answering specific questions and provide useful feedback. However, their accuracy is not consistently reliable, and issues such as cheating and misuse are prevalent. Developers should enhance these systems, while educators must implement appropriate guidelines and detection mechanisms to support responsible student use.

Therefore, LLMs are considered a breakthrough in computer science, particularly in the field of education; however, they still require substantial improvements to mitigate associated issues, such as ethnic biases. In the future, across various academic disciplines—including medicine, mathematics, and others—specific guidelines should be developed to ensure the appropriate adoption of LLMs within each subject area. Moreover, mature IT techniques established in certain disciplines can serve as valuable references for their application in other fields. Furthermore, policy authorities need to establish corresponding regulations to guide the adoption of computer science technologies in each discipline.

Despite some breakthroughs in the research, the thesis also acknowledges certain limitations that should be addressed in future studies. First, there are some inherent limitations of the public databases. Public databases are lacking in records of the academic performance of certain countries and attendance rates on some specific dates, and some children can claim that they belong to more than one ethnicity, which may slightly influence the accuracy of the final results. In future studies, more types of databases will be adopted to analyze the influences of the pandemic on education. In addition to the inherent limitations of the public databases, the standardized tests are limited to TOEFL and GMAT, so future research should cover more assessments, such as IELTS and other international exams, to enable a more comprehensive evaluation of the pandemic's impact on skill-based testing. Future studies should also examine attendance rates of different ethnicities in other multicultural countries for a comprehensive analysis of the influences of pandemics on different ethnicities. For the LLMs, more advanced technologies, including the latest versions of GPT and DeepSeek, will also be incorporated to reflect the more advanced roles and functions of LLMs in education.

List of References

[1] S. Ahmed, A. Shrestha, and J. Yong, "Towards a user-level self-management of covid-19 using mobile devices supported by artificial intelligence, 5g and the cloud," in *Health Information Science: 10th International Conference, HIS 2021, Melbourne, VIC, Australia, October 25–28, 2021, Proceedings 10*. Springer, 2021, pp. 33–43.

[2] J. Zhang, "People's responses to the covid-19 pandemic during its early stages and factors affecting those responses," *Humanities and Social Sciences Communications*, vol. 8, no. 1, 2021.

[3] A. Atalan, "Is the lockdown important to prevent the covid-19 pandemic? effects on psychology, environment and economy-perspective," *Annals of medicine and surgery*, vol. 56, pp. 38–42, 2020.

[4] R. Sarker, A. Roknuzzaman, Nazmunnahar, M. Shahriar, M. J. Hossain, and M. R. Islam, "The who has declared the end of pandemic phase of covid-19: Way to come back in the normal life," *Health science reports*, vol. 6, no. 9, p. e1544, 2023.

[5] R. Sarki, K. Ahmed, H. Wang, Y. Zhang, and K. Wang, "Automated detection of covid-19 through convolutional neural network using chest x-ray images," *Plos one*, vol. 17, no. 1, p. e0262052, 2022.

[6] M. N. Asadullah and E. Tham, "Learning and happiness during covid-19 school closure in urban malaysia," *International Journal of Educational Development*, vol. 101, p. 102822, 2023.

[7] J.-P. Peltola, E. Lindfors, and E. Luukka, "Exploring crisis management measures taken by school leaders at the unpredictable crisis–case covid-19," *Journal of Educational Change*, vol. 25, no. 4, pp. 727–743, 2024.

[8] H. Wang, Y. Zhang, and J. Cao, "Effective collaboration with information sharing in virtual universities," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 6, pp. 840–853, 2008.

[9] B. E. Schueler and M. R. West, "How online learning can engage students and extend the reach of talented teachers: evidence from a pandemic-era national virtual summer program," *Journal of Educational Change*, vol. 24, no. 4, pp. 759–803, 2023.

[10] V. Maity and D. P. Sikdar, "Online platform and covid-19 pandemic situation," *International Journal of Advanced Multidisciplinary Research and Studies*, vol. 3, no. 3, pp. 1010–1014, 2023.

[11] S. Maity, T. N. Sahu, and N. Sen, "Covid-19 and digital primary education: Impact and strategies for sustainable development," *Journal of Development Policy and Practice*, vol. 7, no. 1, pp. 10–30, 2022.

[12] S. Jana, R. Biswas, K. Pal, S. Biswas, and K. Roy, "The evolution and impact of large language model systems: A comprehensive analysis," *Alochana Journal*, 2024.

[13] J. Deng, Y. Zhang, T.-M. Lau, and S. Huang, "Exploring the cultivation of digital intelligence design talents: A case study of human-ai co-creation in forward-looking robotic application scenarios," *Frontiers of Digital Education*, vol. 2, no. 1, p. 9, 2025.

[14] J. Li, "The evolution, applications, and future prospects of large language models: An in-depth overview," *Applied and Computational Engineering*, vol. 35, pp. 234–244, 2024.

[15] A. C. Carius and A. J. Teixeira, "Artificial intelligence and content analysis: the large language models (llms) and the automatized categorization," *AI & SOCIETY*, pp. 1–12, 2024.

[16] Y. Zhang, M.-P. Kwan, and L. Fang, "An llm driven dataset on the spatiotemporal distributions of street and neighborhood crime in china," *Scientific Data*, vol. 12, no. 1, p. 467, 2025.

[17] R. Li, D. Zhang, Y. Wang, Y. Jiang, Z. Zheng, S.-W. Jeon, and H. Wang, "Open-vocabulary multi-object tracking with domain generalized and temporally

adaptive features," *IEEE Transactions on Multimedia*, vol. 27, pp. 3009–3022, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusID:277555494

[18] G. Chen, M. Wang, S. Han, J. Yin, H. Wang, and J. Cao, "Deep reinforcement learning-based cloud-edge offloading for wbans," *IEEE Transactions on Consumer Electronics*, vol. 71, pp. 4053–4064, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusID:274316523

[19] Y.-F. Ge, H. Wang, E. Bertino, J. Cao, and Y. Zhang, "Multiobjective privacy-preserving task assignment in spatial crowdsourcing," *IEEE Transactions on Cybernetics*, vol. 55, pp. 3584–3597, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusID:279308723

[20] R. G. Bhardwaj and H. S. Bedi, "Chatgpt as an education and learning tool for engineering, technology and general studies: performance analysis of chatgpt 3.0 on cse, gate and jee examinations of india," *Interactive Learning Environments*, pp. 1–14, 2025.

[21] V. Pursnani, Y. Sermet, M. Kurt, and I. Demir, "Performance of chatgpt on the us fundamentals of engineering exam: Comprehensive assessment of proficiency and potential implications for professional environmental engineering practice," *Computers and Education: Artificial Intelligence*, vol. 5, p. 100183, 2023.

[22] C. Huang, Q. Huang, X. Huang, H. Wang, M. Li, K.-J. Lin, and Y. Chang, "Xkt: Toward explainable knowledge tracing model with cognitive learning theories for questions of multiple knowledge concepts," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, pp. 7308–7325, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:270734295

[23] L. Chen, W. Liu, H. Wang, S.-W. Jeon, Y. Jiang, and Z. Zheng, "Consistency-guided adaptive alternating training for semi-supervised salient object detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, pp. 7033–7046, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusID:276198496

[24] C. Ganjavi, M. Eppler, D. O'Brien, L. S. Ramacciotti, M. S. Ghauri, I. Anderson, J. Choi, D. Dwyer, C. Stephens, V. Shi *et al.*, "Chatgpt and large language models (llms) awareness and use. a prospective cross-sectional survey of us medical students," *PLOS Digital Health*, vol. 3, no. 9, p. e0000596, 2024.

[25] F. Zhang, Y. A. Wang, S. Liu, and H. Wang, "Decision-based evasion attacks on tree ensemble classifiers," *World Wide Web*, vol. 23, pp. 2957 – 2977, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:215819534

[26] M. E. Kabir, H. Wang, and E. Bertino, "A conditional role-involved purpose-based access control model," *Journal of Organizational Computing and Electronic Commerce*, vol. 21, pp. 71 – 91, 2011. [Online]. Available: https://api.semanticscholar.org/CorpusID:2151736

[27] Y.-F. Ge, M. E. Orlowska, J. Cao, H. Wang, and Y. Zhang, "Mdde: multitasking distributed differential evolution for privacy-preserving database fragmentation," *The VLDB Journal*, vol. 31, pp. 957 – 975, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:245654000

[28] W. Channell, "Making a difference: The role of the llm in policy formulation and reform," in *The Export of Legal Education*. Routledge, 2016, pp. 13–21.

[29] E. Huang, Z. Zhao, J. Yin, J. Cao, and H. Wang, "Transformer-enhanced adaptive graph convolutional network for traffic flow prediction," *ACM Transactions on Intelligent Systems and Technology*, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusID:277766169

[30] X. Sun, M. Li, and H. Wang, "A family of enhanced (l, alpha)-diversity models for privacy preserving data publishing," *Future Gener. Comput. Syst.*, vol. 27, pp. 348–356, 2011. [Online]. Available: https://api.semanticscholar.org/CorpusID:20012270

[31] J.-Q. Yang, Q.-T. Yang, K.-J. Du, C.-H. Chen, H. Wang, S.-W. Jeon, J. Zhang, and Z. hui Zhan, "Bi-directional feature fixation-based particle swarm optimization for large-scale feature selection," *IEEE Transactions on Big Data*, vol. 9, pp. 1004–1017, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:255292831

[32] C. Wang, B. Sun, K.-J. Du, J.-Y. Li, Z. hui Zhan, S.-W. Jeon, H. Wang, and J. Zhang, "A novel evolutionary algorithm with column and sub-block local search for sudoku puzzles," *IEEE Transactions on Games*, vol. 16, pp. 162–172, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:255879353

[33] B. Kereopa-Yorke, “Building resilient smes: Harnessing large language models for cyber security in australia,” *Journal of AI, Robotics & Workplace Automation*, vol. 3, no. 1, pp. 15–27, 2024.

[34] Y.-F. Ge, E. Bertino, H. Wang, J. Cao, and Y. Zhang, “Distributed cooperative coevolution of data publishing privacy and transparency,” *ACM Transactions on Knowledge Discovery from Data*, vol. 18, pp. 1 – 23, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:260614898

[35] Y.-F. Ge, H. Wang, E. Bertino, Z. hui Zhan, J. Cao, Y. Zhang, and J. Zhang, “Evolutionary dynamic database partitioning optimization for privacy and utility,” *IEEE Transactions on Dependable and Secure Computing*, vol. 21, pp. 2296–2311, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:260729441

[36] R. Nowrozy, K. Ahmed, A. S. M. Kayes, H. Wang, and T. R. Mcintosh, “Privacy preservation of electronic health records in the modern era: A systematic survey,” *ACM Computing Surveys*, vol. 56, pp. 1 – 37, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:268577931

[37] S. Fowler, M. Korolkiewicz, and R. Marrone, “First 100 days of chatgpt at australian universities: An analysis of policy landscape and media discussions about the role of ai in higher education,” *Learning Letters*, vol. 1, pp. 1–1, 2023.

[38] P. Lambert, “Measuring remote work using a large language model (llm),” in *EconPol Forum*, vol. 24, no. 3. Munich: CESifo GmbH, 2023, pp. 44–49.

[39] M. N. A. Tawhid, S. Siuly, K. N. Wang, and H. Wang, “Automatic and efficient framework for identifying multiple neurological disorders from eeg signals,” *IEEE Transactions on Technology and Society*, vol. 4, pp. 76–86, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:256290344

[40] M. N. A. Tawhid, S. Siuly, K. N. Wang, and H. Wang, “Genet: A generic neural network for detecting various neurological disorders from eeg,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 16, pp. 1829–1842, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:269028855

[41] A. M. Alvi, M. J. Khan, N. T. Manami, Z. A. Miazi, K. N. Wang, S. Siuly, and H. Wang, “Xcr-net: A computer aided framework to detect covid-19,” *IEEE Transactions on Consumer Electronics*, vol. 70, pp. 7551–7561, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:272140068

[42] Z. T. Hamad, N. Jamil, and A. N. Belkacem, "Chatgpt's impact on education and healthcare: Insights, challenges, and ethical consideration," *IEEE Access*, 2024.

[43] M. Alharbi, J. Yin, Y. Miao, and J. Cao, "From data to insights: Constructing and evaluating a hospitality dataset for quadruple aspect-based sentiment analysis," in *International Conference on Web Information Systems Engineering.* Springer, 2024, pp. 102–113.

[44] M. Alharbi, J. Yin, and H. Wang, "Surveying the landscape: compound methods for aspect-based sentiment analysis," in *Australasian Database Conference.* Springer, 2023, pp. 101–115.

[45] S. Lin and P. Crosthwaite, "The grass is not always greener: Teacher vs. gpt-assisted written corrective feedback," *System*, vol. 127, p. 103529, 2024.

[46] Y. Chen, X. Huang, F. Yang, H. Lin, H. Lin, Z. Zheng, Q. Liang, J. Zhang, and X. Li, "Performance of chatgpt and bard on the medical licensing examinations varies across different cultures: a comparison study," *BMC Medical Education*, vol. 24, no. 1, p. 1372, 2024.

[47] W. Dai, J. Lin, H. Jin, T. Li, Y.-S. Tsai, D. Gašević, and G. Chen, "Can large language models provide feedback to students? a case study on chatgpt," in *2023 IEEE international conference on advanced learning technologies (ICALT).* IEEE, 2023, pp. 323–325.

[48] L. Zhui, N. Yhap, L. Liping, W. Zhengjie, X. Zhonghao, Y. Xiaoshu, C. Hong, L. Xuexiu, R. Wei *et al.*, "Impact of large language models on medical education and teaching adaptations," *JMIR Medical Informatics*, vol. 12, no. 1, p. e55933, 2024.

[49] S. E. Uribe, I. Maldupa, A. Kavadella, M. El Tantawi, A. Chaurasia, M. Fontana, R. Marino, N. Innes, and F. Schwendicke, "Artificial intelligence chatbots and large language models in dental education: Worldwide survey of educators," *European Journal of Dental Education*, 2024.

[50] T. M. Benítez, Y. Xu, J. D. Boudreau, A. W. C. Kow, F. Bello, L. Van Phuoc, X. Wang, X. Sun, G. K.-K. Leung, Y. Lan *et al.*, "Harnessing the potential of large language models in medical education: promise and pitfalls," *Journal of the American Medical Informatics Association*, vol. 31, no. 3, pp. 776–783, 2024.

[51] Y. J. Su, A. Xia, D. Gong, and A. Yan, "The impact of zero-covid policies on academic performance: Evidence from primary and secondary students in china," *Journal of Asian Economics*, vol. 93, p. 101762, 2024.

[52] P. A. Hendriksen, A. H. Koyun, J. Garssen, A.-K. Stock, and J. C. Verster, "Student perceptions of academic functioning during the covid-19 pandemic in germany," *COVID*, vol. 4, no. 11, pp. 1764–1775, 2024.

[53] G. Anakpo, S. Nkungwana, and S. Mishi, "Impact of covid-19 on school attendance in south africa. analysis of sociodemographic characteristics of learners," *Heliyon*, vol. 10, no. 7, 2024.

[54] M. Kuhfeld, J. Soland, K. Lewis, E. Ruzek, and A. Johnson, "The covid-19 school year: Learning and recovery across 2020-2021," *Aera Open*, vol. 8, p. 23328584221099306, 2022.

[55] C. M. Vicario, M. Mucciardi, P. Perconti, C. Lucifora, M. A. Nitsche, and A. Avenanti, "The impact of the covid-19 pandemic on academic performance: a comparative analysis of face-to face and online assessment," *Frontiers in Psychology*, vol. 14, p. 1299136, 2024.

[56] M. Jakubowski, T. Gajderowicz, and H. A. Patrinos, "Covid-19, school closures, and student learning outcomes. new global evidence from pisa," *npj Science of Learning*, vol. 10, no. 1, p. 5, 2025.

[57] P. Xu, J. Yin, J. Cao, and Y. Zhang, "The impact of covid-19 on global gmat test-taking patterns: A statistical analysis," in *International Conference on Health Information Science*. Springer, 2025, pp. 303–312.

[58] P. Engzell, A. Frey, and M. D. Verhagen, "Learning loss due to school closures during the covid-19 pandemic," *Proceedings of the National Academy of Sciences*, vol. 118, no. 17, p. e2022376118, 2021.

[59] S. P. Zhdanov, K. M. Baranova, N. Udina, A. E. Terpugov, E. V. Lobanova, and O. V. Zakharova, "Analysis of learning losses of students during the covid-19 pandemic." *Contemporary Educational Technology*, vol. 14, no. 3, 2022.

[60] J. Gatley, "Intrinsic value and educational value," *Journal of Philosophy of Education*, vol. 55, no. 4-5, pp. 675–687, 2021.

[61] T. K. Kannisto, “Basic education as a collective good: In defence of the school as a public social institution,” *Journal of Philosophy of Education*, vol. 56, no. 2, pp. 305–317, 2022.

[62] H. Zong, J. Li, E. Wu, R. Wu, J. Lu, and B. Shen, “Performance of chatgpt on chinese national medical licensing examinations: a five-year examination evaluation study for physicians, pharmacists and nurses,” *BMC medical education*, vol. 24, no. 1, p. 143, 2024.

[63] S. Takagi, M. Koda, T. Watari *et al.*, “The performance of chatgpt-4v in interpreting images and tables in the japanese medical licensing exam,” *JMIR Medical Education*, vol. 10, no. 1, p. e54283, 2024.

[64] G. Jošt, V. Taneski, and S. Karakatič, “The impact of large language models on programming education and student learning outcomes,” *Applied Sciences*, vol. 14, no. 10, p. 4115, 2024.

[65] S. Filippi and B. Motyl, “Large language models (llms) in engineering education: A systematic review and suggestions for practical adoption,” *Information*, vol. 15, no. 6, p. 345, 2024.

[66] R. Bhayana, S. Krishna, and R. R. Bleakney, “Performance of chatgpt on a radiology board-style examination: insights into current strengths and limitations,” *Radiology*, vol. 307, no. 5, p. e230582, 2023.

[67] W. Hersh and K. Fultz Hollis, “Results and implications for generative ai in a large introductory biomedical and health informatics course,” *NPJ Digital Medicine*, vol. 7, no. 1, p. 247, 2024.

[68] Y. Liang, D. Zou, H. Xie, and F. L. Wang, “Exploring the potential of using chatgpt in physics education,” *Smart Learning Environments*, vol. 10, no. 1, p. 52, 2023.

[69] N. Faizi and Y. Alvi, *Biostatistics manual for Health Research: A practical guide to data analysis.* Elsevier, 2023.

[70] R. K. Rana, R. Singhal, and P. Dua, “Deciphering the dilemma of parametric and nonparametric tests,” *Journal of the Practice of Cardiovascular Sciences*, vol. 2, no. 2, pp. 95–98, 2016.

[71] M. Muhammed Al-Kassab, "The use of one sample t-test in the real data," *Journal of advances in mathematics*, vol. 21, 2022.

[72] S. C. Loftus, *Basic statistics with R: reaching decisions with data.* Academic Press, 2021.

[73] G. Francis and V. Jakicic, "Equivalent statistics for a one-sample t-test," *Behavior Research Methods*, vol. 55, no. 1, pp. 77–84, 2023.

[74] G. Di Leo and F. Sardanelli, "Statistical significance: p value, 0.05 threshold, and applications to radiomics—reasons for a conservative approach," *European radiology experimental*, vol. 4, pp. 1–8, 2020.

[75] S. Greenland, S. J. Senn, K. J. Rothman, J. B. Carlin, C. Poole, S. N. Goodman, and D. G. Altman, "Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations," *European journal of epidemiology*, vol. 31, no. 4, pp. 337–350, 2016.

[76] A. Rovetta, L. Piretta, and M. A. Mansournia, "p-values and confidence intervals as compatibility measures: guidelines for interpreting statistical studies in clinical research," *The Lancet Regional Health-Southeast Asia*, vol. 33, 2025.

[77] J. Gao, "P-values–a chronic conundrum," *BMC Medical Research Methodology*, vol. 20, pp. 1–8, 2020.

[78] A. Olsson-Collentine, M. A. Van Assen, and C. H. Hartgerink, "The prevalence of marginally significant results in psychology over time," *Psychological science*, vol. 30, no. 4, pp. 576–586, 2019.

[79] A. I. Talikan, R. Salapuddin, J. A. Aksan, R. J. Rahimulla, A. Ismael, R. Jimlah, N. Idris, R. B. Dammang, D. A. Jamar, E. Sarahadil *et al.*, "On paired samples t-test: Applications, examples and limitations," *Ignatian*, vol. 2, no. 4, pp. 943–951, 2024.

[80] P. Martínez-Camblor, N. Corral, and J. María de la Hera, "Hypothesis test for paired samples in the presence of missing data," *Journal of Applied Statistics*, vol. 40, no. 1, pp. 76–87, 2013.

[81] X. Manfei, D. Fralick, J. Z. Zheng, B. Wang, M. T. Xin, and F. Changyong, "The differences and similarities between two-sample t-test and paired t-test," *Shanghai archives of psychiatry*, vol. 29, no. 3, p. 184, 2017.

[82] A. Ross and V. L. Willson, *Basic and advanced statistical tests: Writing results sections and creating tables and figures.* Springer, 2018.

[83] H.-Y. Kim, "Statistical notes for clinical researchers: the independent samples t-test," *Restorative Dentistry & Endodontics*, vol. 44, no. 3, p. e26, 2019.

[84] B. Gerald, "A brief review of independent, dependent and one sample t-test," *International journal of applied mathematics and theoretical physics*, vol. 4, no. 2, p. 50, 2018.

[85] S. Acharya and B. Rani, "Limitations involved in a two-sample independent t-test," *Authorea Preprints*, 2024.

[86] T. K. Kim, "Understanding one-way anova using conceptual figures," *Korean journal of anesthesiology*, vol. 70, no. 1, p. 22, 2017.

[87] J. Verma and J. Verma, "One-way anova: Comparing means of more than two samples," *Data analysis in management with SPSS software*, pp. 221–254, 2013.

[88] A. Chatzi and O. Doody, "The one-way anova test explained," *Nurse researcher*, vol. 32, no. 4, 2024.

[89] M. Alassaf and A. M. Qamar, "Improving sentiment analysis of arabic tweets by one-way anova," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 6, pp. 2849–2859, 2022.

[90] P. Schober and T. R. Vetter, "Nonparametric statistical methods in medical research," *Anesthesia & Analgesia*, vol. 131, no. 6, pp. 1862–1863, 2020.

[91] K. M. Ramachandran and C. P. Tsokos, *Mathematical statistics with applications in R.* Academic Press, 2020.

[92] A. P. King and R. J. Eckersley, "Inferential statistics iii: nonparametric hypothesis testing," *Statistics for Biomedical Engineers and Scientists*, pp. 119–145, 2019.

[93] S. J. Wisniewski and G. D. Brannan, "Correlation (coefficient, partial, and spearman rank) and regression analysis," in *StatPearls [Internet].* StatPearls Publishing, 2024.

[94] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: appropriate use and interpretation," *Anesthesia & analgesia*, vol. 126, no. 5, pp. 1763–1768, 2018.

[95] Q. L. Her and J. Wong, "Significant correlation versus strength of correlation," *American Journal of Health-System Pharmacy*, vol. 77, no. 2, pp. 73–75, 2020.

[96] M. F. Ikhwan, W. Mansor, Z. I. Khan, M. K. A. Mahmood, A. Bujang, and K. Haddadi, "Pearson correlation and multiple correlation analyses of the animal fat s-parameter," *TEM Journal*, vol. 13, no. 1, p. 155, 2024.

[97] L. Gust and X. B. D'journo, "The use of correlation functions in thoracic surgery research," *Journal of thoracic disease*, vol. 7, no. 3, p. E11, 2015.

[98] A. Rovetta, "Raiders of the lost correlation: a guide on using pearson and spearman coefficients to detect hidden correlations in medical sciences," *Cureus*, vol. 12, no. 11, 2020.

[99] I. Erb, "Partial correlations in compositional data analysis," *Applied Computing and Geosciences*, vol. 6, p. 100026, 2020.

[100] R. C. Van Aert and C. Goos, "A critical reflection on computing the sampling variance of the partial correlation coefficient," *Research Synthesis Methods*, vol. 14, no. 3, pp. 520–525, 2023.

[101] D. Y. Kenett, X. Huang, I. Vodenska, S. Havlin, and H. E. Stanley, "Partial correlation analysis: Applications for financial markets," *Quantitative Finance*, vol. 15, no. 4, pp. 569–578, 2015.

[102] J.-S. Kang, D.-H. Shin, J.-W. Baek, and K. Chung, "Activity recommendation model using rank correlation for chronic stress management," *Applied Sciences*, vol. 9, no. 20, p. 4284, 2019.

[103] I. Sazonets, S. Yekimov, J. Hinke, D. Šálková, and R. Křečková, "Spearman's correlation coefficient: The case of duplicate ranks," in *E3S Web of Conferences*, vol. 538. EDP Sciences, 2024, p. 02031.

[104] J. Rymarz, A. Niewczas, P. Hołyszko, and E. Dębicka, "Application of spearman's method for the analysis of the causes of long-term post-failure downtime of city buses," *Applied Sciences*, vol. 12, no. 6, p. 2921, 2022.

[105] S. M. Miraftabzadeh, M. Longo, F. Foiadelli, M. Pasetti, and R. Igual, "Advances in the application of machine learning techniques for power system analytics: A survey," *Energies*, vol. 14, no. 16, p. 4776, 2021.

[106] D. Maulud and A. M. Abdulazeez, "A review on linear regression comprehensive in machine learning," *Journal of applied science and technology trends*, vol. 1, no. 2, pp. 140–147, 2020.

[107] F. Zhang and L. J. O'Donnell, "Support vector regression," in *Machine learning*. Elsevier, 2020, pp. 123–140.

[108] P. Anandhi and D. Nathiya, "Application of linear regression with their advantages, disadvantages, assumption and limitations," *International Journal of Statistics and Applied Mathematics*, vol. 8, no. 6, pp. 133–137, 2023.

[109] R. Rodríguez-Pérez and J. Bajorath, "Evolution of support vector machine and regression modeling in chemoinformatics and drug discovery," *Journal of Computer-Aided Molecular Design*, vol. 36, no. 5, pp. 355–362, 2022.

[110] P. Rivas-Perea, J. Cota-Ruiz, D. G. Chaparro, J. A. P. Venzor, A. Q. Carreón, and J. G. Rosiles, "Support vector machines for regression: a succinct review of large-scale and linear programming formulations," 2012.

[111] C. Shoko and C. Sigauke, "Short-term forecasting of covid-19 using support vector regression: An application using zimbabwean data," *American Journal of Infection Control*, vol. 51, no. 10, pp. 1095–1107, 2023.

[112] M. Abdel-Sattar, A. M. Aboukarima, and B. M. Alnahdi, "Application of artificial neural network and support vector regression in predicting mass of ber fruits (ziziphus mauritiana lamk.) based on fruit axial dimensions," *Plos one*, vol. 16, no. 1, p. e0245228, 2021.

[113] Z. Li, Y. Xie, X. Li, and W. Zhao, "Prediction and application of porosity based on support vector regression model optimized by adaptive dragonfly algorithm," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 43, no. 9, pp. 1073–1086, 2021.

[114] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing k-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications," *Journal of Big Data*, vol. 11, no. 1, p. 113, 2024.

[115] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of k-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning," *Decision Analytics Journal*, vol. 3, p. 100071, 2022.

[116] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of k-nearest neighbour (knn) algorithm and its different variants for disease prediction," *Scientific Reports*, vol. 12, no. 1, p. 6256, 2022.

[117] P. Srisuradetchai and K. Suksrikran, "Random kernel k-nearest neighbors regression," *Frontiers in big Data*, vol. 7, p. 1402384, 2024.

[118] S. Winarni and S. Indratno, "Gaussian process regression with average hyperparameter," in *Journal of Physics: Conference Series*, vol. 1496, no. 1. IOP Publishing, 2020, p. 012002.

[119] Z. Sun, L. Zhong, X. Chen, and J. Guo, "Application of gaussian process regression model in industry," in *2022 2nd International Conference on Robotics, Automation and Artificial Intelligence (RAAI).* IEEE, 2022, pp. 221–225.

[120] G. Mateos-Aparicio, "Partial least squares (pls) methods: Origins, evolution, and application to social sciences," *Communications in Statistics-Theory and Methods*, vol. 40, no. 13, pp. 2305–2317, 2011.

[121] L. M. Pereira, V. Sanchez Rodrigues, and F. G. M. Freires, "Use of partial least squares structural equation modeling (pls-sem) to improve plastic waste management," *Applied Sciences*, vol. 14, no. 2, p. 628, 2024.

[122] S. G. Purnama and D. Susanna, "Attitude to covid-19 prevention with large-scale social restrictions (psbb) in indonesia: Partial least squares structural equation modeling," *Frontiers in public health*, vol. 8, p. 570394, 2020.

[123] J. Yin, W. Hong, H. Wang, J. Cao, Y. Miao, and Y. Zhang, "A compact vulnerability knowledge graph for risk assessment," *ACM Transactions on Knowledge Discovery from Data*, vol. 18, no. 8, pp. 1–17, 2024.

[124] Y.-S. Park and S. Lek, "Artificial neural networks: Multilayer perceptron for ecological modeling," in *Developments in environmental modelling.* Elsevier, 2016, vol. 28, pp. 123–140.

[125] A. Rana, A. S. Rawat, A. Bijalwan, and H. Bahuguna, "Application of multi layer (perceptron) artificial neural network in the diagnosis system: a systematic review," in *2018 International conference on research in intelligent and computing in engineering (RICE).* IEEE, 2018, pp. 1–6.

[126] J. Isabona, A. L. Imoize, S. Ojo, O. Karunwi, Y. Kim, C.-C. Lee, and C.-T. Li, "Development of a multilayer perceptron neural network for optimal predictive modeling in urban microcellular radio environments," *Applied Sciences*, vol. 12, no. 11, p. 5713, 2022.

[127] L. Ying *et al.*, "Decision tree methods: applications for classification and prediction," *Shanghai archives of psychiatry*, vol. 27, no. 2, p. 130, 2015.

[128] I. D. Mienye and N. Jere, "A survey of decision trees: Concepts, algorithms, and applications," *IEEE access*, 2024.

[129] T. O. Hodson, "Root-mean-square error (rmse) or mean absolute error (mae): When to use them or not," *Geoscientific Model Development*, vol. 15, no. 14, pp. 5481–5487, 2022.

[130] D. S. K. Karunasingha, "Root mean square error or mean absolute error? use their ratio as well," *Information Sciences*, vol. 585, pp. 609–629, 2022.

[131] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," *International Journal of Forecasting*, vol. 32, no. 3, pp. 669–679, 2016.

[132] A. Tiwari, "Supervised learning: From theory to applications," in *Artificial intelligence and machine learning for EDGE computing.* Elsevier, 2022, pp. 23–32.

[133] A. Schleicher, "Pisa 2018: Insights and interpretations." *oecd Publishing*, 2019.

[134] M. O. Martin, M. Von Davier, and I. V. Mullis, "Methods and procedures: Timss 2019 technical report." *International Association for the Evaluation of Educational Achievement*, 2020.

[135] M. O. Martin, I. V. Mullis, and M. Hooper, "Methods and procedures in pirls 2016." *International Association for the Evaluation of Educational Achievement*, 2017.

[136] E. A. Hanushek, G. Schwerdt, S. Wiederhold, and L. Woessmann, "Returns to skills around the world: Evidence from piaac," *European Economic Review*, vol. 73, pp. 103–130, 2015.

[137] E. Haugom, "The effect of changing from campus-based to digital teaching on student attendance: A case study of norwegian business students," *Heliyon*, vol. 8, no. 11, p. e11307, 2022.

[138] Y. Chen, S. Han, G. Chen, J. Yin, K. N. Wang, and J. Cao, "A deep reinforcement learning-based wireless body area network offloading optimization strategy for healthcare services," *Health Information Science and Systems*, vol. 11, no. 1, p. 8, 2023.

[139] R. Sarki, K. Ahmed, H. Wang, Y. Zhang, and K. Wang, "Convolutional neural network for multi-class classification of diabetic eye disease," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 9, no. 4, pp. e5–e5, 2022.

[140] J. Yin, M. You, J. Cao, H. Wang, M. Tang, and Y.-F. Ge, "Data-driven hierarchical neural network modeling for high-pressure feedwater heater group," in *Australasian Database Conference*. Springer, 2020, pp. 225–233.

[141] U. Ludewig, R. Kleinkorres, R. Schaufelberger, T. Schlitter, R. Lorenz, C. König, A. Frey, and N. McElvany, "Covid-19 pandemic and student reading achievement: Findings from a school panel study," *Frontiers in Psychology*, vol. 13, 2022.

[142] J. Manuel Prieto, J. Salas Sánchez, J. Tierno Cordón, L. Álvarez-Kurogi, H. González-García, and R. Castro López, "Social anxiety and academic performance during covid-19 in schoolchildren," *Plos one*, vol. 18, no. 1, p. e0280194, 2023.

[143] A. C. NWOKOCHA and U. Elijah ETUKUDO, "The effect of covid – 19 on academic achievement of senior secondary school students in english language," *International Journal of English Language Education*, vol. 10, no. 1, p. 44, 2022.

[144] T. J. Patston, J. Kennedy, W. Jaeschke, H. Kapoor, S. N. Leonard, D. H. Cropley, and J. C. Kaufman, "Secondary education in covid lockdown: More anxious and less creative—maybe not?" *Frontiers in psychology*, p. 391, 2021.

[145] J. Cavanaugh, S. Jacquemin, and C. Junker, "A look at student performance during the covid-19 pandemic," *Quality Assurance in Education*, vol. 31, no. 1, pp. 33–43, 2023.

[146] A. T. Capinding, "Analysis of learning and academic performance of education students before and during the coronavirus disease pandemic." *European Journal of Educational Research*, vol. 10, no. 4, pp. 1953–1962, 2021.

[147] S. Jahan and F. Ali, "Advancing health information system with system thinking: Learning challenges of e-health in bangladesh during covid-19," in *Health Information Science: 10th International Conference, HIS 2021, Melbourne, VIC, Australia, October 25–28, 2021, Proceedings 10*. Springer, 2021, pp. 15–23.

[148] P. Vimalachandran, H. Liu, Y. Lin, K. Ji, H. Wang, and Y. Zhang, "Improving accessibility of the australian my health records while preserving privacy and security of the system," *Health Information Science and Systems*, vol. 8, pp. 1–9, 2020.

[149] H. Wang, X. Yi, E. Bertino, and L. Sun, "Protecting outsourced data in cloud computing through access management," *Concurrency and Computation: Practice and Experience*, vol. 28, 04 2014.

[150] M. Diykh, S. Abdulla, A. Y. Oudah, H. A. Marhoon, and S. Siuly, "A novel alcoholic eeg signals classification approach based on adaboost k-means coupled with statistical model," in *Health Information Science: 10th International Conference, HIS 2021, Melbourne, VIC, Australia, October 25–28, 2021, Proceedings 10*. Springer, 2021, pp. 82–92.

[151] M. You, J. Yin, H. Wang, J. Cao, K. Wang, Y. Miao, and E. Bertino, "A knowledge graph empowered online learning framework for access control decision-making," *World Wide Web*, vol. 26, no. 2, pp. 827–848, 2023.

[152] W. Hong, J. Yin, M. You, H. Wang, J. Cao, J. Li, M. Liu, and C. Man, "A graph empowered insider threat detection framework based on daily activities," *ISA transactions*, 2023.

[153] X. Zhao and W. Xue, "From online to offline education in a post-pandemic era: Challenges encountered by international students in uk universities," *Frontiers in Psychology*, vol. 13, p. 8642, 2023.

[154] L. Sun and H. Wang, "Access control and authorization for protecting disseminative information in e-learning workflow," *Concurrency and Computation: Practice and Experience*, vol. 23, no. 16, pp. 2034–2042, 2011.

[155] G. Wang, H. Wang, X. Tao, and J. Zhang, "Positive influence dominating set in e-learning social networks," in *Advances in Web-Based Learning-ICWL 2011: 10th International Conference, Hong Kong, China, December 8-10, 2011. Proceedings 10*. Springer, 2011, pp. 82–91.

[156] M. Hartnett, P. Butler, and P. Rawlins, "Online proctored exams and digital inequalities during the pandemic," *Journal of Computer Assisted Learning*, vol. 39, no. 4, pp. 1103–1115, 2023.

[157] N. Ruiz-Robledillo, J. Vela-Bermejo, V. Clement-Carbonell, R. Ferrer-Cascales, C. Alcocer-Bruno, and N. Albaladejo-Blázquez, "Impact of covid-19 pandemic on academic stress and perceived classroom climate in spanish university students," *International Journal of Environmental Research and Public Health*, vol. 19, no. 7, p. 4398, 2022.

[158] P. Xu, W. Hong, J. Yin, K. Wang, and Y. Zhang, "Analyzing the impact of covid-19 on education: A comparative study based on toefl test results," in *International Conference on Health Information Science*. Springer, 2023, pp. 175–184.

[159] K. Bhopal and M. Myers, "The impact of covid-19 on a level exams in england: Students as consumers," *British Educational Research Journal*, vol. 49, no. 1, pp. 142–157, 2023.

[160] L. V. Sandvik, B. Svendsen, A. Strømme, K. Smith, O. Aasmundstad Sommervold, and S. Aarønes Angvik, "Assessment during covid-19: Students and teachers in limbo when the classroom disappeared," *Educational Assessment*, vol. 28, no. 1, pp. 11–26, 2023.

[161] J. Lee, "Mental health effects of school closures during covid-19," *The Lancet Child & Adolescent Health*, vol. 4, no. 6, p. 421, 2020.

[162] A. Q. Noori, "The impact of covid-19 pandemic on students' learning in higher education in afghanistan," *Heliyon*, vol. 7, no. 10, 2021.

[163] S. Adzaku, H. Agbleze, and P. Afutor, "Perceived impact of covid-19 on students' academic performance in colleges of education in ghana," *African Journal of Education and Practice*, vol. 8, no. 6, pp. 17–29, 2022.

[164] Z. Meiyi and Y. Liu, "Impact of fear of covid-19 on students' performance, moderating role of mindfulness: Hsk students' perception-based view," *Frontiers in public health*, vol. 10, p. 967125, 2022.

[165] Y. Liu, D. Luo, K. Wang, and J. Yin, "Unveiling the pandemic's impact: A dataset for probing covid-19's effects on e-learning activities and academic performance," in *International Conference on Health Information Science*. Springer, 2023, pp. 149–160.

[166] C. Miao, L. Kahle-Piasecki, and S. Qian, “Gmat, ggpa, and conscientiousness as predictors of career success,” *Journal of Management Policy and Practice*, vol. 18, no. 1, 2017.

[167] E. Mathieu and L. Rodés-Guirao, “Our world in data will rely on data from the who to track confirmed covid-19 cases and deaths,” *Our World in Data*, 2023, https://ourworldindata.org/covid-jhu-who.

[168] J. A. Proudfoot, T. Lin, B. Wang, and X. M. Tu, “Tests for paired count outcomes,” *General psychiatry*, vol. 31, no. 1, 2018.

[169] M. D. Riina, C. Stambaugh, N. Stambaugh, and K. E. Huber, “Continuous variable analyses: t-test, mann–whitney, wilcoxin rank,” in *Translational Radiation Oncology*. Elsevier, 2023, pp. 153–163.

[170] H.-Y. Kim, “Statistical notes for clinical researchers: Nonparametric statistical methods: 1. nonparametric methods for comparing two groups,” *Restorative Dentistry and Endodontics*, vol. 39, no. 3, pp. 235–239, 2014.

[171] E. van den Heuvel and Z. Zhan, “Myths about linear and monotonic associations: Pearson'sr, spearman's ρ, and kendall's τ,” *The American Statistician*, vol. 76, no. 1, pp. 44–52, 2022.

[172] Y.-J. Hao, Y.-L. Wang, M.-Y. Wang, L. Zhou, J.-Y. Shi, J.-M. Cao, and D.-P. Wang, “The origins of covid-19 pandemic: A brief overview,” *Transboundary and emerging diseases*, vol. 69, no. 6, pp. 3181–3197, 2022.

[173] C. Del Rio and P. N. Malani, “Covid-19 in 2022—the beginning of the end or the end of the beginning?” *Jama*, vol. 327, no. 24, pp. 2389–2390, 2022.

[174] P. K. Mo, Y. Yu, M. M. Lau, R. H. Ling, and J. T. Lau, “Time to lift up covid-19 restrictions? public support towards living with the virus policy and associated factors among hong kong general public,” *International Journal of Environmental Research and Public Health*, vol. 20, no. 4, p. 2989, 2023.

[175] H. Onyeaka, C. K. Anumudu, Z. T. Al-Sharify, E. Egele-Godswill, and P. Mbaegbu, “Covid-19 pandemic: A review of the global lockdown and its far-reaching effects,” *Science progress*, vol. 104, no. 2, p. 00368504211019854, 2021.

[176] R. Kindred and G. W. Bates, "The influence of the covid-19 pandemic on social anxiety: a systematic review," *International journal of environmental research and public health*, vol. 20, no. 3, p. 2362, 2023.

[177] M. Bratti and E. Lippo, "Covid-19 and the gender gap in university student performance," 2022.

[178] L. Li, "Factors affecting the student performance during the covid-19 pandemic," *Open Journal of Social Sciences*, vol. 10, no. 7, pp. 410–425, 2022.

[179] J. Ritchie, "Wellbeing learnings from pandemic pedagogies in aotearoa new zealand," *New Zealand Journal of Educational Studies*, vol. 58, no. 1, pp. 109–132, 2023.

[180] K. Goodman, "Conceptualising sel in the cross-cultural spaces of primary schools in aotearoa new zealand," *New Zealand Journal of Educational Studies*, vol. 56, no. 2, pp. 285–299, 2021.

[181] D. Joseph and R. Trinick, "'staying apart yet keeping together': challenges and opportunities of teaching during covid-19 across the tasman," *New Zealand Journal of Educational Studies*, vol. 56, no. 2, pp. 209–226, 2021.

[182] J. Clarke, R. Kipping, S. Chambers, K. Willis, H. Taylor, R. Brophy, K. Hannam, S. A. Simpson, and R. Langford, "Impact of covid-19 restrictions on preschool children's eating, activity and sleep behaviours: a qualitative study," *BMJ open*, vol. 11, no. 10, p. e051497, 2021.

[183] C.-h. Ma, L. Jiang, L.-t. Chu, C.-c. Zhang, Y. Tian, J.-j. Chen, and Y. Wang, "Mental health problems of preschool children during the covid-19 home quarantine: A cross-sectional study in shanghai, china," *Frontiers in Psychology*, vol. 13, p. 1032244, 2022.

[184] L. Richter, M. Black, P. Britto, B. Daelmans, C. Desmond, A. Devercelli, T. Dua, G. Fink, J. Heymann, J. Lombardi *et al.*, "Early childhood development: an imperative for action and measurement at scale," *BMJ global health*, vol. 4, no. Suppl 4, p. e001302, 2019.

[185] A. Likhar, P. Baghel, and M. Patil, "Early childhood development and social determinants," *Cureus*, vol. 14, no. 9, 2022.

[186] T. K. Hoffman and G. H. Poll, "Challenges faced by childcare directors during covid-19: Leading during a global pandemic," *Early Childhood Education Journal*, vol. 52, no. 1, pp. 31–41, 2024.

[187] M. R. Jalongo, "The effects of covid-19 on early childhood education and care: Research and resources for children, families, teachers, and teacher educators," *Early childhood education journal*, vol. 49, no. 5, pp. 763–774, 2021.

[188] E. U. Cascio, "Covid-19, early care and education, and child development," 2021.

[189] Q. Zhang, M. Sauval, and J. M. Jenkins, "Impacts of the covid-19 pandemic on the child care sector: Evidence from north carolina," *Early childhood research quarterly*, vol. 62, pp. 17–30, 2023.

[190] C. S. Molloy, S. Guo, and S. Goldfeld, "Patterns of participation in early childhood education before and during the covid-19 pandemic in australia," *Australasian Journal of Early Childhood*, vol. 48, no. 3, pp. 182–202, 2023.

[191] I. Gelir and N. Duzen, "The effects of the covid-19 pandemic on quality of early childhood education and suggestions for a sustainable early childhood education during the health crisis," in *Proceedings of the AUBH E-Learning Conference*, 2021.

[192] R. J. Weijers, L. Ganushchak, K. Ouwehand, and B. B. de Koning, ""i'll be there": Improving online class attendance with a commitment nudge during covid-19," *Basic and Applied Social Psychology*, vol. 44, no. 1, pp. 12–24, 2022.

[193] D. Saatci, T. A. Ranger, C. Garriga, A. K. Clift, F. Zaccardi, P. San Tan, M. Patone, C. Coupland, A. Harnden, S. J. Griffin *et al.*, "Association between race and covid-19 outcomes among 2.6 million children in england," *JAMA pediatrics*, vol. 175, no. 9, pp. 928–938, 2021.

[194] D. D. Bingham, A. Daly-Smith, J. Hall, A. Seims, S. A. Dogra, S. J. Fairclough, M. Ajebon, B. Kelly, B. Hou, K. A. Shire *et al.*, "Covid-19 lockdown: Ethnic differences in children's self-reported physical activity and the importance of leaving the home environment; a longitudinal and cross-sectional study from the born in bradford birth cohort study," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 18, pp. 1–19, 2021.

[195] B. Stierman, J. Y. Abrams, S. E. Godfred-Cato, M. E. Oster, L. Meng, L. Yip, P. Patel, N. Balachandran, E. Prezzato, T. Pierce *et al.*, "Racial and ethnic disparities in multisystem inflammatory syndrome in children in the united states, march 2020 to february 2021," *The Pediatric infectious disease journal*, vol. 40, no. 11, pp. e400–e406, 2021.

[196] E. J. Anderson, C. B. Creech, V. Berthaud, A. Piramzadian, K. A. Johnson, M. Zervos, F. Garner, C. Griffin, K. Palanpurwala, M. Turner *et al.*, "Evaluation of mrna-1273 vaccine in children 6 months to 5 years of age," *New England Journal of Medicine*, vol. 387, no. 18, pp. 1673–1687, 2022.

[197] K. Attwell, L. McKenzie, S. Tomkinson, S. J. Carlson, and C. C. Blyth, "Parents' covid-19 vaccine intentions for children under 5 years: Brief reflections from a qualitative study," *Journal of Paediatrics and Child Health*, vol. 59, no. 3, pp. 453–457, 2023.

[198] S. M. de Vet, C. I. Vrijhof, S. van der Veek, J. M. Pieplenbosch, H. J. van Bakel, and H. J. Vermeer, "Child care in times of covid-19: Predictors of distress in dutch children and parents when re-entering center-based child care after a 2-month lockdown," *Frontiers in Psychology*, vol. 12, p. 718898, 2021.

[199] L. M. Carrascal, I. Galván, and O. Gordo, "Partial least squares regression as an alternative to current regression methods used in ecology," *Oikos*, vol. 118, no. 5, pp. 681–690, 2009.

[200] Z. Xie, X. Chen *et al.*, "Partial least trimmed squares regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 221, p. 104486, 2022.

[201] P. Raj and P. E. David, *The digital twin paradigm for smarter systems and environments: The industry use cases.* Academic Press, 2020.

[202] K. Murakoshi, "Avoiding overfitting in multilayer perceptrons with feeling-of-knowing using self-organizing maps," *BioSystems*, vol. 80, no. 1, pp. 37–40, 2005.

[203] S. M. Robeson and C. J. Willmott, "Decomposition of the mean absolute error (mae) into systematic and unsystematic components," *PloS one*, vol. 18, no. 2, p. e0279774, 2023.

[204] R. Hosseini and N. Askari, "A review of neurological side effects of covid-19 vaccination," *European Journal of Medical Research*, vol. 28, no. 1, pp. 1–8, 2023.

[205] P. Irizar, D. Pan, D. Kapadia, L. Bécares, S. Sze, H. Taylor, S. Amele, E. Kibuchi, P. Divall, L. J. Gray *et al.*, "Ethnic inequalities in covid-19 infection, hospitalisation, intensive care admission, and death: a global systematic review and meta-analysis of over 200 million study participants," *EClinicalMedicine*, vol. 57, 2023.

[206] A. Saw, S. Yi, L. Doan *et al.*, "Improving asian american health during the syndemic of covid-19 and racism. eclinicalmedicine. 2022; 45: 101313," 2022.

[207] Y. Zhao and J. Watterston, "The changes we need: Education post covid-19," *Journal of educational change*, vol. 22, no. 1, pp. 3–12, 2021.

[208] K. S. Schiller, K. C. Wilcox, A. Leo, M. I. Khan, and J. A. M. Ávila, "“i was on an island”: Covid-19 impacts on educator collegiality," *Journal of Educational Change*, vol. 24, no. 4, pp. 1099–1110, 2023.

[209] B. McDonald, K. J. Lester, and D. Michelson, "‘she didn't know how to go back’: School attendance problems in the context of the covid-19 pandemic—a multiple stakeholder qualitative study with parents and professionals," *British Journal of Educational Psychology*, vol. 93, no. 1, pp. 386–401, 2023.

[210] R. Gauld, "A review of public policies on covid-19: the new zealand experience," *Public administration and policy*, vol. 26, no. 1, pp. 10–20, 2023.

[211] D. Qutishat, R. Obeidallah, and Y. Qawasmeh, "An overview of attendance and participation in online class during the covid pandemic: A case study." *International Journal of Interactive Mobile Technologies*, vol. 16, no. 4, 2022.

[212] M. A. Garza, S. C. Quinn, Y. Li, L. Assini-Meytin, E. T. Casper, C. S. Fryer, J. Butler III, N. A. Brown, K. H. Kim, and S. B. Thomas, "The influence of race and ethnicity on becoming a human subject: factors associated with participation in research," *Contemporary clinical trials communications*, vol. 7, pp. 57–63, 2017.

[213] C. Highfield and M. Webber, "Mana ūkaipō: Māori student connection, belonging and engagement at school," *New Zealand Journal of Educational Studies*, vol. 56, no. 2, pp. 145–164, 2021.

[214] M. Webber and A. Macfarlane, "Mana tangata: The five optimal cultural conditions for māori student success," *Journal of American Indian Education*, vol. 59, no. 1, pp. 26–49, 2020.

[215] L. Hendrick, M.-C. Opdenakker, and W. Van der Vaart, "Students' academic engagement during covid-19 times: a mixed-methods study into relatedness and loneliness during the pandemic," *Frontiers in Psychology*, vol. 14, p. 1221003, 2023.

[216] K. Thorsteinsen, E. J. Parks-Stamm, M. Olsen, M. Kvalø, and S. E. Martiny, "The impact of covid-19-induced changes at schools on elementary students' school engagement," *Frontiers in psychology*, vol. 12, p. 687611, 2021.

[217] J. Kalenda, I. Kočvarováa, and E. Boeren, "Impact of covid-19 on participation and barriers in nonformal adult education in the czech republic," *European Journal of Education*, vol. 58, no. 4, pp. 681–698, 2023.

[218] G. Di Pietro, "The impact of covid-19 on student achievement: Evidence from a recent meta-analysis," *Educational Research Review*, vol. 39, p. 100530, 2023.

[219] T. X. Nong, H. T. Nguyen, and T. T. T. Nguyen, "The effect of perceived vaccination on students' online learning intentions: a moderated mediation model," *Human Behavior and Emerging Technologies*, vol. 2022, no. 1, p. 9387583, 2022.

[220] R. Ahmed Laar, Z. Zhang, R. Menhas, L. Zhang, S. Zhu, X. Fan, W. Wang, and S. Li, "Impact of coronavirus disease of 2019 vaccine on health and physical activities among physical education students in china," *Frontiers in Public Health*, vol. 10, p. 889311, 2022.

[221] I. S. Nugroho, E. Kusumawati, and D. D. Wahyuningsih, "Student engagement during pandemic covid-19 and its implications for guidance and counseling," *KONSELI: Jurnal Bimbingan Dan Konseling (E-Journal)*, vol. 8, no. 2, pp. 135–144, 2021.

[222] New Zealand Ministry of Education, "Attendance rates during covid-19," 2023, accessed: 2024-07-30. [Online]. Available: https://www.educationcounts.govt.nz/statistics/attendance-under-covid-19

[223] A. I. Ali and A. K. Menclova, "Schooling opportunities for 5-year-olds: some early origins of high school achievement," *New Zealand Economic Papers*, vol. 58, no. 2, pp. 179–196, 2024.

[224] J. Yin, G. Chen, W. Hong, J. Cao, H. Wang, and Y. Miao, "A heterogeneous graph-based semi-supervised learning framework for access control decision-making," *World Wide Web*, vol. 27, no. 4, p. 35, 2024.

[225] J. Hu and K. Zheng, "A novel support vector regression for data set with outliers," *Applied soft computing*, vol. 31, pp. 405–411, 2015.

[226] X. Zhang, G. Zhang, X. Qiu, J. Yin, W. Tan, X. Yin, H. Yang, L. Liao, H. Wang, and Y. Zhang, "Radiomics under 2d regions, 3d regions, and peritumoral regions reveal tumor heterogeneity in non-small cell lung cancer: a multicenter study," *La radiologia medica*, vol. 128, no. 9, pp. 1079–1092, 2023.

[227] J. Yin, G. Chen, W. Hong, H. Wang, J. Cao, and Y. Miao, "Empowering vulnerability prioritization: A heterogeneous graph-driven framework for exploitability prediction," in *International Conference on Web Information Systems Engineering*. Springer, 2023, pp. 289–299.

[228] P. Wang, Y. Zhang, and W. Jiang, "Application of k-nearest neighbor (knn) algorithm for human action recognition," in *2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, vol. 4. IEEE, 2021, pp. 492–496.

[229] V. GS and H. VS, "Prediction of bus passenger traffic using gaussian process regression," *Journal of Signal Processing Systems*, vol. 95, no. 2, pp. 281–292, 2023.

[230] P. Manoharan, J. Yin, H. Wang, Y. Zhang, and W. Ye, "Insider threat detection using supervised machine learning algorithms," *Telecommunication Systems*, pp. 1–17, 2023.

[231] P. Thomas and M.-C. Suhner, "A new multilayer perceptron pruning algorithm for classification and regression applications," *Neural Processing Letters*, vol. 42, pp. 437–458, 2015.

[232] G. Buturac, "Measurement of economic forecast accuracy: A systematic overview of the empirical literature," *Journal of risk and financial management*, vol. 15, no. 1, p. 1, 2021.

[233] M. Biancolella, V. L. Colona, R. Mehrian-Shai, J. L. Watt, L. Luzzatto, G. Novelli, and J. K. Reichardt, "Covid-19 2022 update: transition of the pandemic to the endemic phase," *Human genomics*, vol. 16, no. 1, p. 19, 2022.

[234] J. Whitley, M. H. Beauchamp, and C. Brown, "The impact of covid-19 on the learning and achievement of vulnerable canadian children and youth," *Facets*, vol. 6, no. 1, pp. 1693–1713, 2021.

[235] M. T. Segni, H. F. Demissie, M. K. Adem, A. K. Geleto, M. W. Kelkile, B. K. Sori, M. L. Heyi, D. G. Iticha, G. S. Bejiga, A. B. Guddisa *et al.*, “Post covid-19 vaccination side effects and associated factors among vaccinated health care providers in oromia region, ethiopia in 2021,” *Plos one*, vol. 17, no. 12, p. e0278334, 2022.

[236] N. Steyn, R. N. Binny, K. Hannah, S. C. Hendy, A. James, A. Lustig, K. Ridings, M. J. Plank, and A. Sporle, “Māori and pacific people in new zealand have higher risk of hospitalisation for covid-19,” *MedRxiV*, pp. 2020–12, 2020.

[237] E. Wikaire, M. Harwood, K. Wikaire-Mackey, S. Crengle, R. Brown, A. Anderson, R. Jansen, and R. Keenan, “Reducing healthcare inequities for māori using telehealth during covid-19,” *NZ Med J*, vol. 135, no. 1552, pp. 112–7, 2022.

[238] J. M. Norris and J. Lee, “The effectiveness of the toefl® essentialstm test for distinguishing english proficiency levels,” *Research Memorandum No. RM-23-07. ETS*, 2023.

[239] D. Kass, J. S. Kim, C. LaMacchia, and A. Bekisz, “Does the gmat matter? the impact of a gmat waiver policy on mba student performance,” *Journal of Education for Business*, vol. 95, no. 7, pp. 432–438, 2020.

[240] S. Papageorgiou and V. F. Manna, “Maintaining access to a large-scale test of academic language proficiency during the pandemic: The launch of toefl ibt home edition,” *Language Assessment Quarterly*, vol. 18, no. 1, pp. 36–41, 2021.

[241] GMAC, “Gmat support – official gmat exam help center,” 2025, accessed: 2025-06-08. [Online]. Available: https://support.mba.com/hc/en-us/sections/6204285008539-GMAT-

[242] C. A. Taylor and P. Angelis, “The evolution of the toefl,” in *Building a Validity Argument for the Test of English as a Foreign Language™*. Routledge, 2011, pp. 27–54.

[243] C. Mutch, “The rise and rise of early childhood education in new zealand,” *Citizenship, Social and Economics Education*, vol. 6, no. 1, pp. 1–11, 2004.

[244] G. Vattiato, D. O’Neale, O. J. Maclaren, and M. J. Plank, “Modelling aotearoa new zealand’s covid-19 protection framework and the transition away from the elimination strategy,” *Royal Society Open Science*, vol. 10, no. 2, p. 220766, 2023.

[245] Royal Commission of Inquiry into COVID-19 Lessons Learned, "B.3 cases, hospitalisations and deaths across 2020 to 2022," https://www.covid19lessons.royalcommission.nz/reports-lessons-learned/main-report/appendices/b-3-cases-hospitalisations-and-deaths-across-2020-to-2022, 2025, accessed: 2025-06-09.

[246] L. Mitchell, E. Hodgen, P. Meagher-Lundberg, and C. Wells, "Impact of covid-19 on the early childhood education sector in aotearoa new zealand: challenges and opportunities," *Wilf Malcolm Institute of Educational Research (WMIER)*, 2020.

[247] Education Review Office, "Learning in a covid-19 world: The impact of covid-19 on early childhood education," 2021, accessed: 2025-06-17. [Online]. Available: https://ero.govt.nz/sites/default/files/2021-04/Learning%20in%20a%20Covid-19%20World%20The%20Impact%20of%20Covid-19%20on%20Early%20Childhood%20Education.pdf

[248] H. Kim, Y. Han, and D. Lee, "The narratives of asians amidst the covid-19 pandemic in aotearoa new zealand: Navigating a virtual realm in the context of anti-asian racism," *Aotearoa New Zealand Social Work*, vol. 36, no. 2, pp. 39–56, 2024.

[249] L. S.-C. Park, R. Jaung, J. J. Park, and C. Song, "Examining access to and trust in sources of covid-19 information among cald asian communities in new zealand," *PloS one*, vol. 20, no. 3, p. e0319326, 2025.

[250] H. Wang, J. Cao, and Y. Zhang, "A self-scalable anonymity payment approach in cloud environment," 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:226757932

[251] H. Wang, J. Cao, and Y. Zhang, "Untraceable electronic cash system in the internet of things," 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:226721314

[252] H. Wang, J. Cao, and Y. Zhang, "Role-based access control constraints and object constraint language," 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:226460909

[253] H. Wang, J. Cao, and Y. Zhang, "Access control management in cloud environments," *Access Control Management in Cloud Environments*, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:222007500

[254] J. Yin, M. Tang, J. Cao, M. You, H. Wang, and M. Alazab, "Knowledge-driven cybersecurity intelligence: Software vulnerability coexploitation behavior discovery," *IEEE Transactions on Industrial Informatics*, vol. 19, pp. 5593–5601, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:250663438

[255] C. Huang, Q. Huang, X. Huang, H. Wang, M. Li, K.-J. Lin, and Y. Chang, "Xkt: Toward explainable knowledge tracing model with cognitive learning theories for questions of multiple knowledge concepts," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, pp. 7308–7325, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:270734295

[256] Y.-F. Ge, H. Wang, J. Cao, Y. Zhang, and X. Jiang, "Privacy-preserving data publishing: an information-driven distributed genetic algorithm," *World Wide Web*, vol. 27, pp. 1–21, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:266985891

[257] M. You, Y.-F. Ge, K. Wang, H. Wang, J. Cao, and G. Kambourakis, "Hierarchical adaptive evolution framework for privacy-preserving data publishing," *World Wide Web*, vol. 27, pp. 1–27, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:271114640

[258] J. Yin, G. Chen, W. Hong, J. Cao, H. Wang, and Y. Miao, "A heterogeneous graph-based semi-supervised learning framework for access control decision-making," *World Wide Web (WWW)*, vol. 27, p. 35, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:270029867

[259] S. Jahan, Y.-F. Ge, H. Wang, and M. E. Kabir, "Dynamic-parameter genetic algorithm for multi-objective privacy-preserving trajectory data publishing," in *WISE*, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:274657985

[260] D. Ning, Y.-F. Ge, H. Wang, and C. Zhou, "Cadif-osn: Detecting cloned accounts with missing profile attributes on online social networks," *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:273497989

[261] X. Qiao, X. Zhou, C. Zhou, H. Wang, and Y. Zhang, "Ebud: Evolving disaster burst detection over social streams," in *WISE*, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:274657859

[262] R. Graham and B. Masters-Awatere, "Experiences of māori of aotearoa new zealand's public health system: a systematic review of two decades of published qualitative research," *Australian and New Zealand journal of public health*, vol. 44, no. 3, pp. 193–200, 2020.

[263] E. Espiner, S.-J. Paine, M. Weston, and E. Curtis, "Barriers and facilitators for māori in accessing hospital services in aotearoa new zealand," *New Zealand Medical Journal*, vol. 134, no. 1546, pp. 47–58, 2021.

[264] N. Sheridan, R. M. Jansen, M. Harwood, T. Love, and T. Kenealy, "Hauora māori–māori health: a right to equal outcomes in primary care," *International journal for equity in health*, vol. 23, no. 1, p. 42, 2024.

[265] K. Tereposky, "Educational inequities for marginalized students in new zealand."

[266] D. Mayeda, A. France, T. Pukepuke, L. Cowie, and M. Chetty, "Colonial disparities in higher education: Explaining racial inequality for māori youth in aotearoa new zealand," *Social Policy and Society*, vol. 21, no. 1, pp. 80–92, 2022.

[267] S. Krause, B. H. Panchal, and N. Ubhe, "Evolution of learning: Assessing the transformative impact of generative ai on higher education," *Frontiers of Digital Education*, vol. 2, no. 2, pp. 1–15, 2025.

[268] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan *et al.*, "The prisma 2020 statement: an updated guideline for reporting systematic reviews," *bmj*, vol. 372, 2021.

[269] I. H. Sarker, "Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions," *SN computer science*, vol. 2, no. 6, pp. 1–20, 2021.

[270] H. K. Jin and E. Kim, "Performance of gpt-3.5 and gpt-4 on the korean pharmacist licensing examination: comparison study," *JMIR Medical Education*, vol. 10, p. e57451, 2024.

[271] U. H. Lai, K. S. Wu, T.-Y. Hsu, and J. K. C. Kan, "Evaluating the performance of chatgpt-4 on the united kingdom medical licensing assessment," *Frontiers in Medicine*, vol. 10, p. 1240915, 2023.

[272] Z. Wu, W. Gan, Z. Xue, Z. Ni, X. Zheng, Y. Zhang *et al.*, "Performance of chatgpt on nursing licensure examinations in the united states and china: cross-sectional study," *JMIR Medical Education*, vol. 10, no. 1, p. e52746, 2024.

[273] T. Nakao, S. Miki, Y. Nakamura, T. Kikuchi, Y. Nomura, S. Hanaoka, T. Yoshikawa, O. Abe *et al.*, "Capability of gpt-4v (ision) in the japanese national medical licensing examination: evaluation study," *JMIR Medical Education*, vol. 10, no. 1, p. e54393, 2024.

[274] D. Stribling, Y. Xia, M. K. Amer, K. S. Graim, C. J. Mulligan, and R. Renne, "The model student: Gpt-4 performance on graduate biomedical science exams," *Scientific Reports*, vol. 14, no. 1, p. 5670, 2024.

[275] P. Giannos, O. Delardas *et al.*, "Performance of chatgpt on uk standardized admission tests: insights from the bmat, tmua, lnat, and tsa examinations," *JMIR Medical Education*, vol. 9, no. 1, p. e47737, 2023.

[276] G. D'Anna, S. Van Cauter, M. Thurnher, J. Van Goethem, and S. Haller, "Can large language models pass official high-grade exams of the european society of neuroradiology courses? a direct comparison between openai chatgpt 3.5, openai gpt4 and google bard," *Neuroradiology*, pp. 1–6, 2024.

[277] P. Künzle and S. Paris, "Performance of large language artificial intelligence models on solving restorative dentistry and endodontics student assessments," *Clinical Oral Investigations*, vol. 28, no. 11, p. 575, 2024.

[278] W. Gan, J. Ouyang, H. Li, Z. Xue, Y. Zhang, Q. Dong, J. Huang, X. Zheng, and Y. Zhang, "Integrating chatgpt in orthopedic education for medical undergraduates: randomized controlled trial," *Journal of medical Internet research*, vol. 26, p. e57037, 2024.

[279] O. Vij, H. Calver, N. Myall, M. Dey, and K. Kouranloo, "Evaluating the competency of chatgpt in mrcp part 1 and a systematic literature review of its capabilities in postgraduate medical assessments," *Plos one*, vol. 19, no. 7, p. e0307372, 2024.

[280] A. Herrmann-Werner, T. Festl-Wietek, F. Holderried, L. Herschbach, J. Griewatz, K. Masters, S. Zipfel, and M. Mahling, "Assessing chatgpt's mastery of bloom's taxonomy using psychosomatic medicine exam questions: mixed-methods study," *Journal of medical Internet research*, vol. 26, p. e52113, 2024.

[281] F. Farhat, B. M. Chaudhry, M. Nadeem, S. S. Sohail, and D. Ø. Madsen, "Evaluating large language models for the national premedical exam in india: comparative analysis of gpt-3.5, gpt-4, and bard," *JMIR Medical Education*, vol. 10, p. e51523, 2024.

[282] M. Rosoł, J. S. Gąsior, J. Łaba, K. Korzeniewski, and M. Młyńczak, "Evaluation of the performance of gpt-3.5 and gpt-4 on the polish medical final examination," *Scientific Reports*, vol. 13, no. 1, p. 20512, 2023.

[283] P. Ros-Arlanzón and A. Perez-Sempere, "Evaluating ai competence in specialized medicine: Comparative analysis of chatgpt and neurologists in a neurology specialist examination in spain," *JMIR Medical Education*, vol. 10, p. e56762, 2024.

[284] S. Wang, C. Mo, Y. Chen, X. Dai, H. Wang, X. Shen *et al.*, "Exploring the performance of chatgpt-4 in the taiwan audiologist qualification examination: Preliminary observational study highlighting the potential of ai chatbots in hearing care," *JMIR Medical Education*, vol. 10, no. 1, p. e55595, 2024.

[285] W. N. V. Luke, L. Seow Chong, K. H. Ban, A. H. Wong, C. Zhi Xiong, L. Shuh Shing, R. Taneja, D. D. Samarasekera, and C. T. Yap, "Is chatgpt 'ready'to be a learning tool for medical undergraduates and will it perform equally in different subjects? comparative study of chatgpt performance in tutorial and case-based learning questions in physiology and biochemistry," *Medical Teacher*, vol. 46, no. 11, pp. 1441–1447, 2024.

[286] Y. Huang, A. Gomaa, S. Semrau, M. Haderlein, S. Lettmaier, T. Weissmann, J. Grigo, H. B. Tkhayat, B. Frey, U. Gaipl *et al.*, "Benchmarking chatgpt-4 on a radiation oncology in-training exam and red journal gray zone cases: potentials and challenges for ai-assisted medical education and decision making in radiation oncology," *Frontiers in Oncology*, vol. 13, p. 1265024, 2023.

[287] S. Bernstein, P. Denny, J. Leinonen, L. Kan, A. Hellas, M. Littlefield, S. Sarsa, and S. Macneil, "" like a nesting doll": Analyzing recursion analogies generated by cs students using large language models," in *Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1*, 2024, pp. 122–128.

[288] I. Joshi, R. Budhiraja, H. Dev, J. Kadia, M. O. Ataullah, S. Mitra, H. D. Akolekar, and D. Kumar, "Chatgpt in the classroom: An analysis of its strengths and weaknesses for solving undergraduate computer science questions," in *Proceedings of*

the 55th ACM Technical Symposium on Computer Science Education V. 1, 2024, pp. 625–631.

[289] Y. Shen, X. Ai, A. G. Soosai Raj, R. J. Leo John, and M. Syamkumar, "Implications of chatgpt for data science education," in *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, 2024, pp. 1230–1236.

[290] J. C. de Winter, "Can chatgpt pass high school exams on english language comprehension?" *International Journal of Artificial Intelligence in Education*, vol. 34, no. 3, pp. 915–930, 2024.

[291] R. Balse, B. Valaboju, S. Singhal, J. M. Warriem, and P. Prasad, "Investigating the potential of gpt-3 in providing feedback for programming assessments," in *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*, 2023, pp. 292–298.

[292] S. Lin, "Evaluating llms' grammatical error correction performance in learner chinese," *PloS one*, vol. 19, no. 10, p. e0312881, 2024.

[293] W. Li and H. Liu, "Applying large language models for automated essay scoring for non-native japanese," *Humanities and Social Sciences Communications*, vol. 11, no. 1, pp. 1–15, 2024.

[294] C. Grévisse, "Llm-based automatic short answer grading in undergraduate medical education," *BMC Medical Education*, vol. 24, no. 1, p. 1060, 2024.

[295] E. Latif and X. Zhai, "Fine-tuning chatgpt for automatic scoring," *Computers and Education: Artificial Intelligence*, vol. 6, p. 100210, 2024.

[296] A. Pack, A. Barrett, and J. Escalante, "Large language models and automated essay scoring of english language learner writing: Insights into validity and reliability," *Computers and Education: Artificial Intelligence*, vol. 6, p. 100234, 2024.

[297] G.-G. Lee, E. Latif, X. Wu, N. Liu, and X. Zhai, "Applying large language models and chain-of-thought for automatic scoring," *Computers and Education: Artificial Intelligence*, vol. 6, p. 100213, 2024.

[298] J. Flodén, "Grading exams using large language models: A comparison between human and ai grading of exams in higher education using chatgpt," *British educational research journal*, vol. 51, no. 1, pp. 201–224, 2025.

[299] F. Yavuz, Ö. Çelik, and G. Yavaş Çelik, "Utilizing large language models for efl essay grading: An examination of reliability and validity in rubric-based assessments," *British Journal of Educational Technology*, vol. 56, no. 1, pp. 150–166, 2025.

[300] J. Li, H. Zong, E. Wu, R. Wu, Z. Peng, J. Zhao, L. Yang, H. Xie, and B. Shen, "Exploring the potential of artificial intelligence to enhance the writing of english academic papers by non-native english-speaking medical students-the educational application of chatgpt," *BMC Medical Education*, vol. 24, no. 1, p. 736, 2024.

[301] S. T. S. Chan, N. P. K. Lo, and A. M. H. Wong, "Enhancing university level english proficiency with generative ai: Empirical insights into automated feedback and learning outcomes," *Contemporary Educational Technology*, vol. 16, no. 4, p. ep541, 2024.

[302] L. Wang, X. Chen, C. Wang, L. Xu, R. Shadiev, and Y. Li, "Chatgpt's capabilities in providing feedback on undergraduate students' argumentation: A case study," *Thinking Skills and Creativity*, vol. 51, p. 101440, 2024.

[303] K. Guo and D. Wang, "To resist it or to embrace it? examining chatgpt's potential to support teacher feedback in efl writing," *Education and Information Technologies*, vol. 29, no. 7, pp. 8435–8463, 2024.

[304] J. Venter, S. A. Coetzee, and A. Schmulian, "Exploring the use of artificial intelligence (ai) in the delivery of effective feedback," *Assessment & Evaluation in Higher Education*, pp. 1–21, 2024.

[305] Z. A. Pardos and S. Bhandari, "Chatgpt-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills," *Plos one*, vol. 19, no. 5, p. e0304013, 2024.

[306] M. Almasre, "Development and evaluation of a custom gpt for the assessment of students' designs in a typography course," *Education Sciences*, vol. 14, no. 2, p. 148, 2024.

[307] G. Urazbayeva, R. Kussainova, A. Aibergen, A. Kaliyeva, and G. Kantayeva, "Innovation off the bat: Bridging the chatgpt gap in digital competence among english as a foreign language teachers," *Education Sciences*, vol. 14, no. 9, p. 946, 2024.

[308] A. Deng, W. Chen, J. Dai, L. Jiang, Y. Chen, Y. Chen, J. Jiang, and M. Rao, "Current application of chatgpt in undergraduate nuclear medicine education: Taking chongqing medical university as an example," *Medical Teacher*, pp. 1–7, 2024.

[309] M. C. Laupichler, J. F. Rother, I. C. G. Kadow, S. Ahmadi, and T. Raupach, "Large language models in medical education: comparing chatgpt-to human-generated exam questions," *Academic Medicine*, vol. 99, no. 5, pp. 508–512, 2024.

[310] C. J. Edwards and B. L. Erstad, "Evaluation of a generative language model tool for writing examination questions," *American Journal of Pharmaceutical Education*, vol. 88, no. 4, p. 100684, 2024.

[311] H.-S. Kim and G.-T. Kim, "Can a large language model create acceptable dental board-style examination questions? a cross-sectional prospective study," *Journal of dental sciences*, 2024.

[312] S. Elkins, E. Kochmar, J. C. Cheung, and I. Serban, "How teachers can use large language models and bloom's taxonomy to create educational quizzes," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 21, 2024, pp. 23 084–23 091.

[313] A. X. Fan, R. H. Zhang, L. Paquette, and R. Zhang, "Exploring the potential of large language models in generating code-tracing questions for introductory programming courses," *arXiv preprint arXiv:2310.15317*, 2023.

[314] H. Gonzalez, L. Dugan, E. Miltsakaki, Z. Cui, J. Ren, B. Li, S. Upadhyay, E. Ginsberg, and C. Callison-Burch, "Enhancing human summaries for question-answer generation in education," in *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, 2023.

[315] B. H. H. Cheung, G. K. K. Lau, G. T. C. Wong, E. Y. P. Lee, D. Kulkarni, C. S. Seow, R. Wong, and M. T.-H. Co, "Chatgpt versus human in generating medical graduate exam multiple choice questions—a multinational prospective study (hong kong sar, singapore, ireland, and the united kingdom)," *PloS one*, vol. 18, no. 8, p. e0290691, 2023.

[316] M. de la Puente, J. Torres, A. L. B. Troncoso, Y. Y. H. Meza, and J. X. M. Carrascal, "Investigating the use of chatgpt as a tool for enhancing critical thinking

and argumentation skills in international relations debates among undergraduate students," *Smart Learning Environments*, vol. 11, no. 1, p. 55, 2024.

[317] H. B. Essel, D. Vlachopoulos, A. B. Essuman, and J. O. Amankwa, "Chatgpt effects on cognitive skills of undergraduate students: Receiving instant responses from ai-based conversational large language models (llms)," *Computers and Education: Artificial Intelligence*, vol. 6, p. 100198, 2024.

[318] D. Darmawansah, D. Rachman, F. Febiyani, and G.-J. Hwang, "Chatgpt-supported collaborative argumentation: Integrating collaboration script and argument mapping to enhance efl students' argumentation skills," *Education and Information Technologies*, pp. 1–25, 2024.

[319] M. Kaliterna, M. F. Žuljević, L. Ursić, J. Krka, and D. Duplančić, "Testing the capacity of bard and chatgpt for writing essays on ethical dilemmas: A cross-sectional study," *Scientific Reports*, vol. 14, no. 1, p. 26046, 2024.

[320] V. Grande, N. Kiesler, and M. A. Francisco R, "Student perspectives on using a large language model (llm) for an assignment on professional ethics," in *Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1*, 2024, pp. 478–484.

[321] A. Abd-Alrazaq, R. AlSaad, D. Alhuwail, A. Ahmed, P. M. Healy, S. Latifi, S. Aziz, R. Damseh, S. A. Alrazak, J. Sheikh *et al.*, "Large language models in medical education: opportunities, challenges, and future directions," *JMIR Medical Education*, vol. 9, no. 1, p. e48291, 2023.

[322] D. W. Anson, "The impact of large language models on university students' literacy development: a dialogue with lea and street's academic literacies framework," *Higher Education Research & Development*, pp. 1–14, 2024.

[323] V. Rahimzadeh, K. Kostick-Quenet, J. Blumenthal Barby, and A. L. McGuire, "Ethics education for healthcare professionals in the era of chatgpt and other large language models: Do we still need it?" *The American Journal of Bioethics*, vol. 23, no. 10, pp. 17–27, 2023.

[324] F. Liu, X. Chang, Q. Zhu, Y. Huang, Y. Li, and H. Wang, "Assessing clinical medicine students' acceptance of large language model: based on technology acceptance model," *BMC Medical Education*, vol. 24, no. 1, p. 1251, 2024.

[325] M. A. A. Ballesteros, B. G. A. Enríquez, E. V. R. Farroñán, H. D. G. Juárez, L. E. C. Salinas, J. E. B. Sánchez, J. C. A. Castillo, G. S. Licapa-Redolfo, and G. C. F. Chilicaus, "The sustainable integration of ai in higher education: Analyzing chatgpt acceptance factors through an extended utaut2 framework in peruvian universities," *Sustainability*, vol. 16, no. 23, pp. 1–28, 2024.

[326] G. L. Liu, R. Darvin, and C. Ma, "Exploring ai-mediated informal digital learning of english (ai-idle): A mixed-method investigation of chinese efl learners' ai adoption and experiences," *Computer Assisted Language Learning*, pp. 1–29, 2024.

[327] J. Liu, F. Liu, J. Fang, and S. Liu, "The application of chat generative pre-trained transformer in nursing education," *Nursing Outlook*, vol. 71, no. 6, p. 102064, 2023.

[328] Y. Zhu, C. Zhu, T. Wu, S. Wang, Y. Zhou, J. Chen, F. Wu, and Y. Li, "Impact of assignment completion assisted by large language model-based chatbot on middle school students' learning," *Education and Information Technologies*, vol. 30, no. 2, pp. 2429–2461, 2025.

[329] M. Landers, "Adapting to the unsanctioned use of ai-supported technologies in student assessments," *Higher Education for the Future*, vol. 12, no. 1, pp. 76–96, 2025.

[330] I. R. Indran, P. Paranthaman, N. Gupta, and N. Mustafa, "Twelve tips to leverage ai for efficient and effective medical question generation: a guide for educators using chat gpt," *Medical Teacher*, vol. 46, no. 8, pp. 1021–1026, 2024.

[331] S. Mehta and N. Mehta, "Embracing the illusion of explanatory depth: a strategic framework for using iterative prompting for integrating large language models in healthcare education," *Medical Teacher*, pp. 1–4, 2024.

[332] C. Wilkinson, M. Oppert, and M. Owen, "Investigating academics' attitudes towards chatgpt: A qualitative study," *Australasian Journal of Educational Technology*, vol. 40, no. 4, pp. 105–120, 2024.

[333] L. Hays, O. Jurkowski, and S. K. Sims, "Chatgpt in k-12 education," *TechTrends*, vol. 68, no. 2, pp. 281–294, 2024.

[334] A. V. Y. Lee, "Supporting students' generation of feedback in large-scale online course with artificial intelligence-enabled evaluation," *Studies in Educational Evaluation*, vol. 77, p. 101250, 2023.

[335] Y. Wang, "When artificial intelligence meets educational leaders' data-informed decision-making: A cautionary tale," *Studies in Educational Evaluation*, vol. 69, p. 100872, 2021.

[336] C. K. Y. Chan and L. H. Tsi, "Will generative ai replace teachers in higher education? a study of teacher and student perceptions," *Studies in Educational Evaluation*, vol. 83, p. 101395, 2024.

[337] S. Anna, "Implementing the oecd ai principles:-challenges and best practices," 2022.

[338] D. Ogugua, S. N. Yoon, and D. Lee, "Academic integrity in a digital era: Should the use of chatgpt be banned in schools?" *Global Business & Finance Review (GBFR)*, vol. 28, no. 7, pp. 1–10, 2023.