

**Challenging Equivalency: A Two-Site Analysis of English Language Proficiency Pathways
and Academic Achievement in Canadian Higher Education**

Jason R. Holmes

A dissertation submitted to the Faculty of Graduate Studies in partial fulfillment of the
requirements for the degree of Doctor of Philosophy

Graduate Program in Education

York University

Toronto, Ontario

September 2025

© Jason R. Holmes, 2025

Abstract

This dissertation investigates the relationship between English language proficiency (ELP) pathways and academic achievement among international undergraduate students at two Canadian universities (University A and B) and challenges the presumed equivalence between scores on standardized English ELP tests (IELTS and TOEFL) and the completion of English for Academic Purposes (EAP) programs. Given the scarcity of comparative research on each pathway's effectiveness in predicting academic achievement, this study examined whether students admitted through different ELP pathways and at different proficiency levels (low to high) significantly differed in terms of their first-year cumulative grade point average (CGPA) and credits earned (CCE), while controlling for discipline, country of origin, gender, and age. By examining two institutions with similar ELP policies but distinct student populations, this study also explores how institutional context may moderate the relationship between ELP pathways and academic achievement.

Using a comprehensive quantitative approach, with data from 9,673 students at University A and 2,480 at University B, analyses consisted of multiple regression, analysis of covariance and interaction analyses. ELP pathways were categorized based on test scores relative to admission cutoffs and EAP completion. Regression models progressively incorporated moderators to assess their impact on academic outcomes.

Results indicated that students admitted through IELTS outperformed their peers from EAP programs in CGPA, though at University B, this trend was evident only among students with the highest IELTS scores. In contrast, the TOEFL pathway showed less consistent superiority compared to EAP students: only students with the highest TOEFL scores significantly exceeded EAP students in CGPA. For CCE, however, students admitted through IELTS and TOEFL surpassed EAP students across both institutions. Moderated by discipline, gender, age, and country of origin, outcomes varied by institution, highlighting context-specific effects.

These findings indicate that students admitted through standardized tests typically outperformed those admitted through EAP, challenging presumed equivalency between pathways. Implications for policy and practice for Canadian universities are discussed to better ensure support for international students' academic success.

Acknowledgements

Professional Acknowledgments:

I would first like to express my gratitude to my supervisor, Dr. Khaled Barkaoui. Two years into my program, I found myself seeking a new supervisor for a research direction completely different from the one I had originally proposed when I was admitted to the PhD program. I specifically sought out Khaled as a potential supervisor because of his incredible expertise not only in the content area of this dissertation, but also the methodological precision and depth that characterizes his work. His support and guidance throughout this project, as well as in various other aspects of my professional development, have been instrumental to my success and have significantly shaped my analytical abilities as a researcher. Thank you, Khaled.

I would like to extend my sincere thanks to my committee members, Dr. Saskia Van Viegan and Dr. Qiang Zha, for their support for this project. From the proposal stage onward, both provided invaluable feedback and insightful direction that strengthened the quality of this dissertation. Their willingness to engage and support my research is deeply appreciated. Thank you, Saskia and Qiang for your commitment to my work.

I would also like to thank my internal examiner Dr. James Corcoran for his interest in supporting my work. His comments and words of encouragement motivate me to continue to explore the topic of this dissertation and to disseminate the findings more broadly.

I would like to express my gratitude to my external examiner, Dr. Gregory Tweedie. When I first began exploring the topic of this dissertation, I felt both encouraged by the apparent lack of research in the field, and, simultaneously, uncertain of a viable path forward. Dr. Tweedie's work, along with that of his colleagues, provided a foundational template for a feasible project which gave me with the confidence to argue for the necessity and unique contribution of this dissertation. Without the groundwork laid by Dr. Tweedie, this project would not have been possible.

I am grateful to Dr. Hugh McCague and Stephan Bonfield for their statistical expertise and words of encouragement at critical methodological obstacles during this dissertation.

Personal Acknowledgments:

The completion of a project of this magnitude would not have been possible without the unwavering support of my friends and family.

I was very fortunate to meet, and become close friends with, many incredible people throughout the PhD program and during my time in Toronto. While I dare not attempt to name them all, their work continues to inspire me, and I am truly grateful for their friendship.

I am also deeply thankful for the many lifelong friends who continue to offer me encouragement and support in all aspects of my professional and personal life. Their now decades long friendships mean more to me than words can convey.

Both my paternal and maternal grandparents supported me throughout my educational journey with their kindness, curiosity, and much more. The encouragement of these four individuals and their belief in my abilities have been a powerful force behind this achievement. My grandfather's love of education and lifelong curiosity served as a constant source of inspiration, reminding me of the value of learning and perseverance. I hope to instill this same spirit of curiosity, resilience, and commitment to lifelong learning in my own children, so that they too may come to appreciate the joy and fulfillment of education.

I am thankful to my brother, Jared Holmes, for his constant support and for his genuine happiness for my achievements.

I began my post-secondary academic journey at 17 at the encouragement of my parents, Bob and Lori Holmes. Although I was initially reluctant, I soon discovered a deep love for learning in this new environment – so much so that I have spent the past decade studying various aspects of higher education, culminating in this dissertation. Throughout every stage of my life, my parents have supported me in every conceivable way. Their unwavering encouragement and guidance have been integral to my success and stand as the exemplar of love and understanding that I hope to model for my own children. Thank you, Mom and Dad, for everything.

Lastly, completing a doctorate would not have been possible without my beloved partner, Summer Cowley. No individual plays a more integral role than Summer in my continued success. After meeting during our master's program, we chose to pursue our respective PhDs in higher education, and this period has been the greatest adventure of my life. Throughout this journey, Summer was wise, capable, understanding, and supportive – she is the foundational rock that grounds me in moments of doubt. The life we built together during graduate school is truly remarkable, as are the two children, Naomi and Tristan, that she brought into our lives while she simultaneously completed her own doctorate. Her guidance and support during the most challenging parts of this dissertation were invaluable, and her extraordinary achievements continue to inspire me to reach for even greater heights.

Summer, you motivate me each day to strive to be the very best version of myself. Thank you, Darling. You are, without question, the best thing that has ever happened to me.

Funding Acknowledgements

This project was partially funded by the Social Science and Humanities Research Council Doctoral Fellowship, project # 752-2022-1353.

This project was also partially funded by Ontario Graduate Scholarships awarded from 2019-2022.

Table of Contents

<i>Abstract</i>	ii
Acknowledgements	iii
Table of Contents	vi
List of Tables	x
List of Figures	xi
Chapter 1: Introduction	1
Key Concepts	3
<i>Language Proficiency</i>	3
<i>Measures of English Language Proficiency</i>	3
<i>English Proficiency Pathway</i>	5
<i>Academic Achievement</i>	6
Rationale for the Current Study	6
Research Questions	9
Overview of Thesis Organization	9
Chapter 2: Literature Review	11
Defining and Measuring Academic Success	12
<i>Student Success versus Academic Success</i>	13
The Operationalization of Academic Achievement	16
<i>Grade point average</i>	16
<i>Credits earned</i>	21
Defining and Operationalizing English Proficiency and Pathways	24
<i>What is Language Proficiency?</i>	24
<i>The Measurement of English Language Proficiency</i>	28
<i>University Admission: English language proficiency challenges</i>	30
The Relationship Between English Language Proficiency and Academic Achievement	32
<i>The relationship between TOEFL performance and academic achievement</i>	34
<i>The relationship between IELTS performance and academic achievement</i>	41
English Proficiency Pathways and Academic Achievement	49
Moderators of Academic Achievement	58
<i>Gender</i>	59
<i>Age</i>	61
<i>Country of Origin</i>	61
<i>Academic Major</i>	63
Conclusion	65

Chapter 3: Method	69
Institutional Context	69
<i>University A</i>	72
<i>University B</i>	73
Overview of Dataset	75
Description of Variables	77
<i>English Proficiency Pathway</i>	78
<i>Credits earned</i>	80
<i>Discipline of study</i>	81
<i>Age</i>	83
<i>Gender</i>	83
<i>Country of Citizenship</i>	84
Data Cleaning	84
<i>Exclusion based on pathway</i>	85
<i>Exclusion based on other predictor variables</i>	88
<i>Exclusion based on dependent variable</i>	89
Final Sample	90
<i>English proficiency pathways and ELP levels</i>	90
<i>Discipline of study</i>	93
<i>Country of origin</i>	94
<i>Gender</i>	95
<i>Age</i>	96
Statistical Analyses	97
Multiple Regression Analysis	98
Method of Analysis and Assumption Testing	101
<i>Linearity</i>	101
<i>Homoscedasticity</i>	102
<i>Normality of Residuals</i>	105
<i>Lack of Multicollinearity</i>	109
Conclusion	111
Chapter 4: Findings	113
Relationship between ELP pathways and CGPA	113
Descriptive Statistics	114
Regression Analysis Results	120
Analysis of Covariance	129

<i>Comparative Analysis between IELTS and TOEFL Pathways</i>	130
<i>Comparative Analysis within IELTS and TOEFL Pathways</i>	135
Interaction Effects	138
<i>Interaction Effects between English Language Proficiency Pathways and Academic Discipline</i>	139
<i>Interaction Effects between English Language Proficiency Pathways and Country of Origin</i>	141
<i>Interaction Effects between English Language Proficiency Pathways and Gender</i>	143
Relationship between ELP pathways and Credits Earned	145
Descriptive Statistics	146
Regression Analysis Results	152
Analysis of Covariance	161
<i>Comparative analysis between IELTS and TOEFL pathways</i>	161
<i>Comparative analysis within IELTS and TOEFL pathways</i>	163
Interaction Effects	165
<i>Interaction Effects between English Language Proficiency Pathways and Academic Discipline</i>	166
<i>Interaction Effects between English Language Proficiency Pathways and Country of Origin</i>	168
<i>Interaction Effects between English Language Proficiency Pathways and Gender</i>	170
Chapter 5: Discussion	173
Synthesis of Main Findings	173
<i>Standardized Tests vs. EAP Pathways: Key Findings</i>	174
<i>Reevaluating IELTS–TOEFL Equivalencies</i>	181
<i>The Moderating Roles of Discipline, Origin, Gender, and Age</i>	183
Significance and Contributions for Practice and Policy	186
Limitations	190
Recommendations for Future Research	191
References	194
Appendix A: Pairwise Comparison for CGPA by ELP Pathway and Discipline for University A .	223
Appendix B: Pairwise Comparison for CGPA by ELP Pathway and Discipline for University B .	228
Appendix C: Pairwise Comparison for CGPA by ELP Pathway and Origin for University A	230
Appendix D: Pairwise Comparison for CGPA by ELP Pathway and Origin for University B	233
Appendix E: Pairwise Comparison for CGPA by ELP Pathway and Gender for University A	235
Appendix F: Pairwise Comparison for CGPA by ELP Pathway and Gender for University B	238
Appendix G: Pairwise Comparison for CCE by ELP Pathway and Gender for University A	239
Appendix H: Pairwise Comparison for CCE by ELP Pathway and Discipline for University B	241
Appendix I: Pairwise Comparison for CCE by ELP Pathway and Origin for University A	243
Appendix J: Pairwise Comparison for CCE by ELP Pathway and Origin for University B	246

Appendix K: Pairwise Comparison for CCE by ELP Pathway and Gender for University A	248
Appendix L: Pairwise Comparison for CCE by ELP Pathway and Gender for University B.....	251

List of Tables

<i>Table 3.1:</i> Raw data counts	75
<i>Table 3.2:</i> Variable overview	77
<i>Table 3.3:</i> Discipline data received	82
<i>Table 3.4:</i> Discipline data recoded	83
<i>Table 3.5:</i> Final pathway count by ELP level, pathway, and university	91
<i>Table 3.6:</i> Subject group by university	94
<i>Table 3.7:</i> Country of citizenship by university	95
<i>Table 3.8:</i> Gender by university	96
<i>Table 3.9:</i> Age by university	96
<i>Table 3.10:</i> Overview of modelled variables	100
<i>Table 4.1:</i> Mean CGPA by English language proficiency pathway by university	114
<i>Table 4.2:</i> Mean CGPA by discipline of study by university	116
<i>Table 4.3:</i> Mean CGPA by country of citizenship by university	117
<i>Table 4.4:</i> Mean CGPA by gender by university	118
<i>Table 4.5:</i> Mean CGPA by age (range) by university	120
<i>Table 4.6:</i> Regression Analysis Estimates (and SE) for CGPA for University A	124
<i>Table 4.7:</i> Regression Analysis Estimates (and SE) for CGPA for University B	128
<i>Table 4.8:</i> Pairwise comparisons between pathways at University A	132
<i>Table 4.9:</i> Pairwise comparisons between pathways at University B	134
<i>Table 4.10:</i> Pairwise Comparisons within IELTS and TOEFL Pathways at University A	136
<i>Table 4.11:</i> Pairwise Comparisons within IELTS and TOEFL Pathways at University B	137
<i>Table 4.12:</i> Mean credits earned by English language proficiency pathway by university	146
<i>Table 4.13:</i> Mean credits earned by discipline of study by university	149
<i>Table 4.14:</i> Mean credits earned by country of citizenship by university	150
<i>Table 4.15:</i> Mean credits earned by gender by university	150
<i>Table 4.16:</i> Mean CGPA by age (range) by university	152
<i>Table 4.17:</i> Regression Analysis Estimates (and SE) for CCE for University A	155
<i>Table 4.18:</i> Regression Analysis Estimates (and SE) for CCE for University B	159
<i>Table 4.19:</i> Pairwise Comparisons between IELTS and TOEFL Pathways at University A for Cumulative Credits Earned	162
<i>Table 4.20:</i> Pairwise Comparisons between IELTS and TOEFL Pathways at University B for Cumulative Credits Earned	163
<i>Table 4.21:</i> Pairwise Comparisons within IELTS and TOEFL Pathways at University A for Cumulative Credits Earned	164
<i>Table 4.22:</i> Pairwise Comparisons within IELTS and TOEFL Pathways at University B for Cumulative Credits Earned	165

List of Figures

<i>Figure 3.1:</i> University A Scatterplot (fitted values and residuals)	104
<i>Figure 3.2:</i> University B Scatterplot (fitted values and residuals)	105
<i>Figure 3.3:</i> University A Histogram (normality of residuals)	106
<i>Figure 3.4:</i> University B Histogram (normality of residuals)	106
<i>Figure 3.5:</i> University A QQ-plot (normality of residuals)	108
<i>Figure 3.6:</i> University B QQ-plot (normality of residuals)	108
<i>Figure 3.7:</i> University A variance inflation factor test	110
<i>Figure 3.8:</i> University B variance inflation factor test	111
<i>Figure 4.1:</i> Predictive margins with 95% CIs - interaction effects between pathway and discipline (University A)	140
<i>Figure 4.2:</i> Predictive margins with 95% CIs - interaction effects between pathway and discipline (University B)	141
<i>Figure 4.3:</i> Predictive margins with 95% CIs - interaction effects between pathway and origin (University A)	142
<i>Figure 4.4:</i> Predictive margins with 95% CIs - interaction effects between pathway and origin (University B)	143
<i>Figure 4.5:</i> Predictive margins with 95% CIs - interaction effects between pathway and gender (University A)	144
<i>Figure 4.6:</i> Predictive margins with 95% CIs - interaction effects between pathway and gender (University B)	145
<i>Figure 4.7:</i> Predictive margins with 95% CIs - interaction effects between pathway and discipline (University A)	167
<i>Figure 4.8:</i> Predictive margins with 95% CIs - interaction effects between pathway and discipline (University B)	168
<i>Figure 4.10:</i> Predictive margins with 95% CIs - interaction effects between pathway and origin (University A)	169
<i>Figure 4.11:</i> Predictive margins with 95% CIs - interaction effects between pathway and origin (University B)	170
<i>Figure 4.12:</i> Predictive margins with 95% CIs - interaction effects between pathway and gender (University A)	171
<i>Figure 4.13:</i> Predictive margins with 95% CIs - interaction effects between pathway and gender (University B)	172

Chapter 1: Introduction

For more than two decades, federal and institutional policies have actively promoted the expansion of international enrollment in Canada's post-secondary sector (Government of Canada, 2020). Over that span, the number of international students grew dramatically—from roughly one hundred thousand at the turn of the millennium to well beyond half a million by the late 2010s—which continued to rise in the years following the COVID-19 pandemic (Crossman et al., 2022). This sustained growth eventually prompted the federal government to impose a temporary cap on new study-permit approvals in 2024 reducing the number of international students coming to Canada by 40%, which was, additionally, reduced further (10%) in 2025 (Government of Canada, 2024; 2025). Post-secondary institutions generally have actively recruited international students because they pay, on average, five times more in tuition than domestic learners, making them an essential revenue source (Statistics Canada, 2024). Provincial operating funding have not kept pace, further encouraging schools to rely on this key income stream (Metcalf, 2010). This reliance is now a risk as federal immigration policy changes are expected to flatten—or even reduce—future international enrolment, creating significant budgetary concerns.

The demographics of Canada's international student body has also shifted dramatically. In 2010/11 the top source countries were China, France, India, the United States, and South Korea, but by 2020/21 they had become India, China, France, Iran, and Vietnam; Indian enrolment alone grew by 1,078 percent (Government of Canada, 2023). At the time of writing, most international learners come from India and China, bringing greater cultural and linguistic diversity to campuses. Because many students come from non-English-speaking backgrounds,

they must meet institutional English-language proficiency standards—usually by submitting scores from standardized tests, though many universities offer in-house pathway programs that substitute test scores with preparatory language coursework. Regardless of pathway – either standardized or course-based, the assumption by post-secondary admissions is that these options are equivalent in their abilities to ensure that students have met the minimum English language proficiency necessary for them to be successful in an English medium environment (Davies et al., 1999; Schoepp, 2018; Tweedie & Chu, 2019).

While there is significant research on the academic outcomes of international students who meet post-secondary admission requirements through the submission of standardized English language proficiency test scores (Al-Malki, 2014; Ingram & Bayliss, 2007; Cho & Bridgeman, 2012; Dooley & Oliver, 2002; Feast, 2002; Harsh et al., 2017; O'Dwyer et al., 2018; Schoepp & Garinger, 2016; Schoepp, 2018; Yen & Kuzma, 2009), there is significantly less research examining the academic outcomes of students who meet admission standards through the course-based pathway (Baik & Greig, 2009; Dooley, 2010; Floyd, 2015; Fox et al., 2006; Fox et al., 2013; Roessingh, & Douglas, 2012). In Canada, these course-based pathways, typically referred to as English for Academic Purposes (EAP) programs, are expensive, intellectually intensive, and often lack research on the outcomes of students who undertake such options (Roy & Landry, 2021).

Therefore, to address this significant gap and to examine the assumption of equivalency of these pathways, this study examines and compares the relationship between English language proficiency pathway and academic achievement using quantitative data from the course based and the test based pathway options. Crucially, the analysis draws on cohorts from two large

English-medium Canadian universities (hereafter University A and University B), enabling a direct comparison of whether effects hold constant—or diverge—across institutional contexts.

Key Concepts

This section establishes key concepts referred to throughout this thesis. A more detailed discussion of the more complex terms (e.g. English language proficiency) is presented in the literature review chapter; however, this section establishes the necessary nomenclature for a reader to progress through the remainder of the thesis.

Language Proficiency

What precisely defines language proficiency ranges with seminal articles proposing models that have been debated and built upon for decades (e.g. Bachman & Palmer, 1996; Canale & Swain, 1980). A more detailed discussion of this complex construct is presented in the following chapter; however, for definitional purposes language proficiency is defined as “the extent to which an individual possesses the linguistic cognition necessary to function in a given communicative situation, in a given modality (listening, speaking, reading, or writing)” (Hulstijn, 2011, p. 242). Language proficiency is typically measured or operationalized by standardized test scores or some other form of academic assessment (e.g., course grades). Specifically, in this dissertation, English language proficiency (ELP) will be exclusively referred to, and the above definition provides the necessary context for a reader to navigate through subsequent material.

Measures of English Language Proficiency

International students who speak English as a second or additional language must typically demonstrate their ELP to meet admission requirements at universities and colleges where English is the primary language of instruction. These institutions have adopted the

massification and widespread use of standardized English language proficiency testing to meet post-secondary language admission standards (Cheng & Fox, 2008). These tests are typically either the International English Language Testing System (IELTS) or the Test of English as a Foreign Language (TOEFL) implemented in a “gate keeping role” across English higher education systems globally (Hirsh, 2006, p. 195). Sometimes referred to as high stakes tests (Solórzano, 2008), standardized English tests have come to dominate the English higher education landscape and are the most common form of evidence submitted to demonstrate English language proficiency for university admissions (Bretag, 2007; Fox, 2005).

Across Canadian institutions, there are many other standardized English language proficiency tests often accepted to demonstrate ELP for university admission purposes (e.g., Pearson Test of English Academic (PTE Academic), Canadian Academic English Language Assessment (CEAL), Duolingo test, and the Canadian English Language Proficiency Index Program (CELPIP)). Of these many tests, the TOEFL and IELTS have dominated in various national contexts for admission purposes (e.g. TOEFL in the USA and IELTS in Australia, Canada, New Zealand, Britain) (Ginther & Elder, 2014; Tomlinson, 2020). As a result of this widespread usage and adoption in almost every Canadian university admission context, and due to the extensive literature investigating these tests, this dissertation purposefully collected data for these two tests.

In addition to the standardized measurement of ELP with the IELTS or the TOEFL, post-secondary institutions have also constructed alternative pathways for international students to achieve and demonstrate the necessary ELP to meet admission requirements. Often referred to by different terms such as: *non testing*, *study pathway alternatives*, *pre-sessional*, *top-up*, or *English for Academic purposes (EAP)* programs. Typically, these alternatives are offered to students who

cannot demonstrate their ELP to a high enough standard on standardized ELP tests (e.g., IELTS). Therefore, these students are often thought of possessing insufficient English language ability to meet the English proficiency criteria required for university admission (Floyd, 2015; Thorpe et al., 2017). As a result, students are offered this alternative pathway which often takes the form of coursework and, at times, the passing of university specific proprietary English tests. This dissertation uses EAP as shorthand for this alternative course-based pathway due to the proliferation of this specific terminology. Therefore, throughout this dissertation the *EAP pathway* is used and conceptualized as the course-based pathway that students who do not achieve the required score on a standardized ELP test, take in order to meet the university admission requirements.

English Proficiency Pathway

Examination and comparison of these different pathways in terms of academic outcomes is generally lacking; therefore, agreed upon terminology regarding these pathways is scarce. For example, Schoepp (2018) refers to indirect and direct entry students, being those who submit IELTS test scores and those who complete a foundation English preparatory program, respectively. Conversely, Tweedie and Chu (2019) simply refer to different measures of ELP directly (IELTS, TOEFL, English programs), while Floyd (2015) distinguishes between “Testing and EAP pathway[s]”(p. 6). For this dissertation, three distinct pathways are referred to.

1. Meeting university admission requirements with a sufficient IELTS score
2. Meeting university admission requirements with a sufficient TOEFL score
3. Meeting university admission requirements through the successful completion of EAP programming.

Throughout this thesis, ELP testing options (e.g. IELTS, TOEFL) are referred to directly when possible; however, the *testing pathway* refers to both the IELTS and TOEFL together and no other standardized English test. Additionally, when referring to both tests and the course-based EPA pathway as a group, the term *English proficiency pathway(s)* (EPP) is used.

Academic Achievement

As with ELP, a more robust and nuanced discussion on terminology surrounding the definition and measurement of academic success or academic achievement is presented in the following literature review chapter. However, it is necessary to provide the reader with a working definition of this construct at this point. Academic achievement is complex but can be thought of as various academic outcomes within an educational setting. Typically, this means that academic achievement is operationalized through grades, semester/term or cumulative grade point average (GPA), and/or credits earned (Kuh et al., 2006). This definition provides the reader with the necessary terminological to progress through the research questions of this study – as well as other key initial material.

Rationale for the Current Study

Students can demonstrate their English language proficiency (ELP) through numerous pathways—primarily through standardized test-taking routes (e.g., TOEFL or IELTS) and institution-offered course-based routes. Standardized tests serve a gatekeeping role, reflecting an institutional belief that these assessments accurately gauge students' linguistic readiness for higher education, especially when combined with other academic metrics such as high school grades (Murray, 2016). Additionally, institutions often assume that different ELP tests are equivalent, despite evidence to the contrary (e.g., Schoepp, 2018; Tweedie & Chu, 2019). Davies et al. (1999) have argued explicitly that equating distinct ELP tests is fundamentally

unjustifiable, as each test targets different purposes and populations. IELTS and TOEFL, developed using differing frameworks, should therefore not be equated because doing so violates essential validity principles, leading to potential misinterpretation and misuse of their scores (Tweedie & Chu, 2019).

Given these validity issues, the institutional practice of equating scores from IELTS and TOEFL as thresholds for admission necessitates closer analysis. These tests effectively function as "selection, entrance, or readiness tests" (Bachman, 1990, p. 58). However, admission score thresholds vary across institutions—and even across faculties within the same institution—further complicating consistent interpretation and application (Tweedie & Chu, 2019).

Moreover, colleges and universities have increasingly introduced course-based alternatives, prominently English for Academic Purposes (EAP) programs, designed explicitly to simulate and prepare students for academic tasks, such as writing academic essays, formal presentations, and critical textual analysis (Douglas & Rosvold, 2018; Cheng et al., 2004; Pearson, 2020). EAP programs aim to develop proficiency within an academic linguistic register, equipping students with essential language skills for academic success (Jordan, 2004). Despite their proliferation, critics have raised concerns about potential compromises in linguistic standards, suggesting universities may prioritize enrollment growth and revenue generation over academic preparedness, thus calling for careful monitoring of outcomes from different ELP pathways (Fedorova, 2021; Floyd, 2015).

This study addresses critical gaps in existing literature by examining and comparing the relationship between these distinct ELP pathways (standardized test versus course-based approaches) and first-year academic achievement, measured through cumulative GPA and credits earned. Despite extensive correlational research between standardized test performance

and academic achievement (e.g., Cho & Bridgeman, 2012; Harsh et al., 2017; O'Dwyer et al., 2018; Al-Malki, 2014; Feast, 2002), there remains a significant deficiency of comparative analyses examining how standardized pathways directly compare to course-based pathways in predicting academic outcomes (Floyd, 2015; Tweedie & Chu, 2019; Johnson & Tweedie, 2021). Furthermore, comparative research across diverse institutional contexts is exceedingly rare, hindering broader generalization of findings.

This dissertation addresses notable gaps in the literature, contributing timely and essential insights not only for researchers but also for educational institutions, ELP test developers, policymakers, and other stakeholders. Institutions gain valuable data to refine admission standards and support mechanisms. Test developers benefit from evidence informing test validity and alignment with academic demands. Policymakers obtain crucial data to shape equitable and effective language admission policies. Ultimately, by examining comprehensive datasets from two Canadian universities (dubbed: University A and University B), this dissertation uniquely explores the robustness and transferability of proficiency-achievement relationships across distinct academic environments. Most research investigating the relationship between ELP pathways and academic achievement has relied heavily on single-site research designs. This limitation restricts the broader applicability and generalizability of findings, as institutional contexts vary significantly. Therefore, a key strength and significance of this dissertation is its comparative multi-site design, analyzing data from two distinct institutional settings. This approach addresses previous limitations by potentially enhancing the external validity and reliability of the findings, offering more robust insights into the relationship between ELP pathways and academic success. Such analyses thereby enhance our collective understanding of

international student success and may contribute to improved policy and practice in higher education.

Research Questions

This study asks the following research questions:

How do international undergraduate students who are required to provide evidence of ELP and are admitted to university through different pathways compare in terms of their academic achievement at the end of their first year of study?

This study explores this overarching research question through two sub-questions:

1. What is the relationship between English language proficiency pathway and academic achievement measured by year one cumulative GPA?
2. What is the relationship between English proficiency pathway and academic achievement measured by number of credits earned?
3. How do these relationships compare between two Canadian universities?

Overview of Thesis Organization

This thesis is organized into five chapters: introduction, literature review, methods, findings, and discussion and implications. The literature review chapter expands on the terminology listed above and examines how scholars have conceptualized the complex constructs of English language proficiency and academic or student success. Furthermore, it explores in detail the literature on the relationship between performance on standardized English proficiency tests and various post-secondary success metrics such as GPA. It also contains a synthesis of the limited literature that is directly related to the research questions of this study.

The methods chapter describes the data and variables of the study, data cleaning procedures, descriptive statistics of the sample, and the statistical analysis methods used. The findings chapter presents the results of data analyses on all relevant variables in this dissertation and provides statistical models for comparison. The last chapter discusses the findings and their implications for policy, practice and research, and discusses limitations and directions for future research.

Chapter 2: Literature Review

This chapter presents the relevant literature that directly informs the current study, which examines the relationship between English proficiency pathways and academic achievement. The review begins by exploring how academic and student success have been conceptualized, focusing on the challenges posed by inconsistent terminology in the literature. Academic achievement, often operationalized as grade point average (GPA), cumulative GPA (CGPA), or less frequently as credits earned, is central to the study. Additionally, this review also addresses the broader constructs of academic and student success, emphasizing the distinctions and overlaps between these terms. The review further delves into the range of factors that influence GPA, which is the primary measure of academic achievement in this study. These predictors include demographic variables such as gender, age, and country of origin. Next, the chapter provides a detailed exploration of English language proficiency (ELP), a critical component of this study. Understanding how ELP is conceptualized in higher education serves as the foundation for discussing the English proficiency pathways examined in this study. Following this, the chapter explores research on the predictive power of standardized ELP tests such as IELTS and TOEFL, highlighting the ongoing debates about their effectiveness in predicting academic success. Information on how these tests are widely used for admissions to English-speaking institutions is presented also. Finally, the chapter examines a significant gap in the literature: the comparison of different English proficiency pathways (EPPs), including English for Academic Purposes (EAP) programs, and their relationship with academic achievement. While substantial research exists on the predictive validity of individual ELP tests, there is a notable lack of studies comparing outcomes for students who follow different EPPs. This review synthesizes the limited research in this area and emphasizes the need for a more robust analysis,

which this dissertation aims to provide by utilizing a larger, multi-institutional dataset and by controlling for key moderator variables (e.g., discipline of study, gender).

Defining and Measuring Academic Success

When framing the research questions for this study, it became clear from the literature that terminology regarding outcome measurement of international students' success in post-secondary education is fraught with term inconsistencies. Relevant literature was found that used different terms, at times, to refer to the same outcome measurement (e.g., GPA). Particularly relevant to this dissertation were studies that explored the relationship between ELP and/or ELP pathways and some form of academic/student success. However, despite similarities in research goals, scholars have utilised terms such as *academic success* (e.g. Graham, 1987; Hill et al., 1999; Yuksel et al., 2021), *academic achievement* (e.g. Johnson & Tweedie, 2021; Wongtrirat, 2010; Yixin & Daller, 2014), *academic performance* (e.g. Cho & Bridgeman, 2012; Floyd, 2015; Kerstijens & Nery, 2000; Martirosyan et al., 2015), *study success* (e.g. Daller & Yixin, 2017) and more to explore very similar questions. This creates significant terminological inconsistencies which in turn leads to challenges when synthesizing research findings and drawing conclusions from across studies. This problem is particularly pertinent when examining the relationship between ELP pathways and academic outcomes, as this dissertation does, as variations in terminology obscure nuances in how outcomes are defined and measured, which outcomes are measured, and how results from studies are interpreted. As a result, this section aims to clarify these terminological discrepancies for adopting a consistent terminological framework for discussing student outcomes which ensures that research questions and findings are clearly communicated. Therefore, the following section presents key conceptual literature on how

student outcomes are understood in higher education and then presents terminology directly relevant to this dissertation.

Student Success versus Academic Success

Of the terms presented above, student and academic success are the broadest. The most comprehensive and cited work on student success is research by Kuh et al. (2006). The authors provided a synthesis of relevant literature regarding what they termed as student success in higher education (rather than academic success). Kuh et al. were concerned with poor graduation rates, lengthy time to degree completion, high attrition rates, increases in post-secondary costs and the lack of enrollment of certain minority groups and students with disabilities. To address these issues, the authors noted that it was necessary to understand that students do not come to post-secondary education as blank slates and, as a result, student success looks very different for different people. Therefore, a definition that encapsulates the various aspects of student success that can vary across diverse populations was deemed necessary. As a result, the authors defined student success in higher education as: “academic achievement, engagement in educationally purposeful activities, satisfaction, acquisition of desired knowledge, skills and competencies, persistence, attainment of educational objectives, and post college performance” (p. 7). The authors acknowledged that recent changes in the landscape of higher education (e.g. greater attention to diversity) required a multidimensional construct of student success and that indicators within each dimension may necessarily need to be broadened for different types of students. Therefore, a holistic and comprehensive definition of student success was required. Considering the proliferation of citations (3000+) that point to Kuh et al.’s comprehensive definition, student success has become an umbrella term that encompasses a wide range of outcomes and experiences. It is important to highlight that within the domain of student success,

key components such as academic achievement, persistence, or post college performance can be operationalized in various ways. For instance, academic achievement may be measured through indicators like GPA, grades, or credits earned (Allen, 2005; Stoyoff, 1997; Zheng & Mustapha, 2022), while persistence can be assessed through retention rates and continued enrollment (Hagedorn, 2005; Park et al., 2008; Reason, 2009; Soldner et al., 2016). These different operationalizations allow for a nuanced understanding of academic success, but they also present challenges, as variations in how these measures are defined and applied can lead to inconsistencies in research. This variability is compounded by terminological issues found throughout the student and academic success literature, where the use of terminology remains problematic.

Despite the clear definition of student success and its complexities, a meta-analysis by York et al. (2015) found that scholars often used student success and academic success interchangeably. This interchangeable use of terms further complicates efforts to synthesize research findings and draw consistent conclusions, as the lack of standardized language can obscure important distinctions between different aspects of student outcomes. The authors made the key distinction that academic success is intended to focus on "...the attainment of outcomes specific to educational experiences" (p. 1) and is, therefore, explicitly different from student success. This is because student success contains components such as engagement in educationally purposeful activities which accounts for various social and psychological factors present in the undergraduate experience (e.g. relationship building, mental health support). In other words, academic success typically focuses on the outcomes of education (e.g. GPA, labour market outcomes, graduation rates, skill development etc.) and while student success contains these outcomes of education, the term also includes experiential dimensions such as student

engagement. Engagement can mean everything from students' interest in educational topics to involvement in activities (York et al., 2015). Therefore, York et al. argued that engagement is not congruent with academic success, and that student success and academic success are unique constructs but with significant overlap. As a result, academic success is defined “as inclusive of academic achievement, attainment of learning objectives, acquisition of desired skills and competencies, satisfaction, persistence, and post-college performance” (p. 5). York et al. also found that of the various components of academic success, academic achievement was the most frequent explored in the literature – and “almost entirely measured with grades and GPA” (p. 7).

Despite these distinctions, many scholars simplify the complex constructs of academic or student success by operationalizing them with more straightforward metrics like grades or GPA (e.g., Angelkoska et al., 2016; Enikő & Szamosközi, 2017; Wharrad et al., 2003), or by hybridizing with credits earned (e.g., Mattern & Shaw, 2010; Trapmann et al., 2007). These are established measures of academic achievement rather than broad indicators of student or academic success. This common misconception is exemplified by Graham's (1987) observation that while GPA may be the most common indicator for judging academic success, it is not always the best indicator. The definition of student/academic success has evolved significantly since Graham's writing, now encompassing various aspects as noted above. Given the multidimensional nature of academic and student success, measuring either in its entirety is an extraordinarily complex task, as each component would require operationalization. For example, while persistence may be operationalized with quantitative measures such as retention rates (Park et al., 2008; Tinto, 1999, 2017), other components—such as the attainment of educational objectives, the acquisition of desired skills and competencies, and satisfaction—would require complex quantitative and qualitative student data. This complexity often leads researchers to

focus on academic achievement, typically operationalised by first semester GPA or end of the first year cumulative GPA (Van der Zanden et al., 2018).

The Operationalization of Academic Achievement

Academic achievement, as a specific component of both academic and student success, represents a crucial but narrow slice of the broader concepts of academic/student success. Although academic achievement forms one component of academic success, it is a prolific term throughout different bodies of literature. Despite its widespread usage, even highly cited meta-analytic studies that examine various predictors of academic achievement do not provide a definition of the construct (e.g. Valentine et al., 2004; Wilder, 2023). Fortunately, Kuh et al. (2006) suggested operationalizing academic achievement using traditional metrics such as college grades, entry exam performance, and credit hours earned, which represent progress toward degree completion. However, Steinmayr et al. (2014) offer a more comprehensive definition, suggesting that academic achievement is a multifaceted construct which encompasses different domains of learning. These domains include general indicators such as declarative knowledge, curricular-based criteria such as grades or performance on tests, and cumulative indicators such as educational degrees. According to Steinmayr et al. “academic achievement represents performance outcomes that indicate the extent to which a person has accomplished specific goals that were the focus of activities in instructional environments” (p. 1). As a result of definitions such as this, GPA is a commonly utilized operationalization of academic achievement.

Grade point average

GPA (measured by semester or cumulatively over various academic timeframes (e.g., one year, full length of a degree, etc.) has long been utilized as the primary operationalization of

academic achievement, making it a frequently used dependent variable in studies which explore the wide range of factors that influence academic and/or student success. It is not the purpose of this literature review to extensively cover all the fields of study which utilize GPA as the main operationalization of academic achievement; however, it is important to acknowledge that the scope of research in this area is vast and well established, thereby justifying the use of GPA as the key dependent variable and operationalization of academic achievement for this dissertation. Such research, spanning disciplines such as education, sociology, economics, psychology, and more, explores a diverse array of predictors which may correlate with GPA, such as socioeconomic status (Sirin, 2005), personality traits (Nofle & Robins, 2007; O'Connor & Paunonen, 2007; Poropat, 2009), resilience, self-efficacy, and motivation (Chemers et al., 2001; DeFreitas, 2011; Hannon, 2014; Richardson et al., 2012; Robbins et al., 2004; Wilson et al., 2019), educational activities (Fauria & Fuller, 2015), class attendance (Credé et al., 2010), parental involvement (Wilder, 2023), communication skills (Aydin, 2017), high school GPA and standardized test (e.g. SAT/ACT) performance (Noble & Sawyer, 2002; Saunders-Scott et al., 2018; Schwager et al., 2015; Vulperhorst et al., 2018; Westrick et al., 2015; Zwick, 2013), and more. Detailed meta-analyses, such as those done by Van der Zanden et al. (2018), have outlined a significant number of predictors specifically related to academic achievement which ranges in its operationalization from first semester GPA, end of the first year of cumulative GPA, to credits earned. The authors found from their review of 36 studies which explored predictors of academic achievement operationalized primarily by GPA “that students' previous academic achievement, effort, academic self-efficacy, study skills, and participation in special first-year programs related most strongly to students' first-year university academic achievement with small to medium effect sizes” (p. 65).

As demonstrated, GPA is a very common measure of academic achievement, used across disciplines to explore questions related to student/academic success. The appeal of using GPA perhaps lies in its quantitative simplicity, accessibility, and the standardized method by which it is calculated across courses, programs, and even across schools or countries. Furthermore, GPA is also an effective indicator or predictor of other aspects of academic success and post-graduation outcomes which has long been explored and debated (Bretz, 1989; Dye & Reck, 1989). Therefore, GPA serves as a useful and versatile tool for researchers and institutions to assess broader outcomes such as persistence, retention, and post-graduation success. For example, in their highly influential work, Pascarella and Terenzini (2005) noted that grades are most likely the best predictor of degree completion, student persistence and graduate school enrolment. Such claims are supported by numerous influential works in the field of student success (e.g. Adelman, 1999; Astin, 1993; Chen, 2005; Korobova, 2012). Although not a perfect indicator, Pascarella and Terenzini (2005) viewed GPA specifically as “...the lingua franca of the academic instructional world” and the key “to students’ standing and continued enrollment, to admission, to majors and enrollment caps, to program and degree completion, to admission to graduate and professional schools, and to employment opportunities” (p. 397). In other words, as noted by Kuh et al. (2006), “[in] Pascarella and Terenzini’s 1991 and 2005 syntheses...good grades in the first year appear to be particularly important to ensuing academic success and degree acquisition” (p. 76).

As a result of research utilizing GPA, it is widely recognized as a strong indicator of academic achievement due to its ability to predict various aspects of student success, both during and beyond university education. Numerous studies have explored the predictive power of GPA for future academic performance and broader outcomes. For instance, Gershensfeld et al. (2015)

found that first-semester GPA strongly predicted college graduation rates among underrepresented students, with those earning a GPA of 2.33 or lower at significant risk of not completing their degrees. Similarly, Letkiewicz et al. (2015) highlighted that students with a GPA below 3.0 were 78% more likely to take longer than four years to graduate compared to their peers with higher GPAs. Other studies have shown that early GPA performance can forecast later success in undergraduate programs (Adekitan & Salau, 2019; Aluko et al., 2018; Asif et al., 2017; Oguntunde et al., 2018), while GPA from undergraduate education can predict graduate school performance (Zimmermann et al., 2015). Additionally, GPA has been found to be a significant predictor of university retention rates (Aulck et al., 2016), emphasizing its utility across multiple aspects of academic success.

With such widespread adoption of GPA as an outcome variable in predictive studies due to its versatility and as a predictor of other types of student success, it is not surprising that studies vary in their choice of first-term GPA or cumulative GPA (CGPA) (either calculated at the end of the first year or a longer time frame) as the outcome measure. However, many studies that use various calculations of GPA do not clearly state why they have chosen one calculation of GPA over another. For example, some studies may use cumulative GPA (typically calculated at the end of a specific year of study) as their primary outcome variable but often do not offer explicit rationale as to why this specific measure was selected (e.g. Arco-Tirado et al., 2011; Bowman, 2014; De Wit et al., 2012; Shaw et al., 2011). There are, of course, various reasons why one form of GPA may be more appropriate as an outcome variable than another. For example, first-semester grade point average is a key predictor of persistence in university; therefore, studies that may wish to demonstrate that a particular educational intervention is effective may use first-semester GPA as an outcome variable to infer that the educational

intervention has longer term benefits than simply raising GPA – namely, also benefiting longer term academic success components such as persistence (e.g., Gibbison et al., 2011). Conversely, cumulative GPA is often used to evaluate longer-term academic success in college, graduate school, the workplace, or otherwise, offering a broader picture of a student’s sustained performance over time (e.g. D’Agostino & Powers, 2009; Zhou et al., 2015). Therefore, studies that utilize CGPA may be more focused on the lasting effects of interventions or various predictors (e.g., Friedman & Mandel, 2011), or the overall impact of factors such as English language proficiency (e.g. Sadeghi et al., 2013), or socio-economic background on long-term success (e.g. Sackett et al., 2009). This makes CGPA an attractive measure for studies looking to understand how initial conditions may influence trajectories over several semesters or years.

However, the inconsistency in justifying the choice between first-term GPA and cumulative GPA complicates the comparability of results across studies. Without clear rationale, it becomes difficult to discern whether variations in findings are due to differences in the populations studied or the specific GPA measure employed. As GPA is such a common metric for academic achievement, this gap in justification highlights the need for more explicit discussions regarding the selection of different GPA measures. By clearly articulating why a specific GPA measurement is chosen, researchers can ensure that their findings are better understood in context, and that comparisons across studies are more meaningful. This dissertation specifically used CGPA as the key operationalization of academic achievement, the rationale for this selection is provided in the section on CGPA in the methods chapter .

Overall, these findings collectively establish GPA as a key indicator of both academic achievement and future success in higher education, making it a strong and appropriate outcome variable for this dissertation when exploring the relationship between English language

proficiency pathways and academic achievement. While GPA has been widely accepted as the primary operationalization of academic achievement due to its ability to predict various aspects of student success, it is not the only measure available. In addition to GPA, credits earned provide another way to operationalize academic achievement. Credits earned reflect the progress students make toward degree completion and are often used in conjunction with GPA to offer a more comprehensive view of student success (e.g., Kuh et al., 2008). Therefore, the following section will explore the use of credits earned as an additional operationalization of academic achievement and discuss its relevance to this dissertation's analysis.

Credits earned

While GPA is widely used as the key operationalization of academic achievement in research on academic/student success, a smaller subset of literature incorporates credits earned as an alternative or additional operationalization, typically as a proxy for study progress (e.g., Arnold & Rowaan, 2014; Jansen & Suhre, 2010; Malm et al., 2012; Severiens & Schmidt, 2009; Torenbeek et al., 2010). This operationalization may be underutilized, partly due to the relative ease of access to GPA data and the standardized nature of GPA calculations across institutions. In their meta-analyses of 32 studies and 132 effect sizes exploring validity evidence of the ability of measures of English proficiency to predict academic achievement, Ihlenfeldt and Rios (2022) noted that GPA is “the most accessible academic outcome metric” (p. 7) and included no studies that measure academic achievement using any measure of credits. Conversely, Zanden et al. (2018), in their meta-analyses of predictors of first-year students’ academic achievement found that out of 36 studies, 30 used first-semester GPA or end of the first year cumulative GPA as outcome variables, while only 5 incorporated credits, and only one study used both CGPA and credits. In this sole study that used both measures, Shao et al. (2010) explored the effect of a

multi-year student success program and measured its impact on student outcomes. In addition to GPA, the authors analyzed whether students attempted and earned more credits compared to those who did not participate in specific program components, finding that students who completed the programming attempted and earned significantly more credits than those who did not (Cohen's $d = .53$).

This gap underscores the importance of incorporating both GPA and credits earned as measures of academic achievement, especially in studies related to the focus of this dissertation. Notably, only two relevant studies—Floyd (2015) and Light et al. (1987)—were found to include both metrics, further emphasizing the limited exploration of this dual operationalization in the context of English language proficiency pathways and their impact on academic outcomes. Light et al. in addition to exploring the relationship between language proficiency demonstrated on the TOEFL and GPA, investigated the relationship between language proficiency and credit hours earned by international students. The authors found that TOEFL scores correlated significantly with credit hours earned in the first semester ($r = .19, p < .01$). They also conducted an ANOVA and found a significant difference in the mean number of credit hours earned across different TOEFL score ranges ($F(4, 371) = 3.76, p < .005$). Students with higher TOEFL scores generally earned more credit hours, with the most notable differences between the 530-549 and 550-599 score ranges. However, this increasing pattern leveled off for scores above 599.

Floyd's (2015) work explored the differences in academic achievement between testing and EAP pathways. While Light et al utilized credits attempted and earned as separate measures, Floyd created a credit ratio (the number of credits earned divided by credits attempted) as an additional operationalization of academic achievement (in addition to GPA). However, the

author found no significant difference between the testing and EAP group for this variable when comparing the two pathway groups.

While GPA is the most used operationalization of academic achievement in studies of student success, the literature demonstrates that credits earned may serve as a valuable supplementary measure or indicator of study progress. Credits earned reflect a student's ability to complete coursework and by examining credits earned, researchers can gain insights into not only a student's academic performance but also their persistence and engagement with their studies, which GPA alone may not fully capture.

The limited use of credits earned as an outcome variable may stem from the ease of access to GPA data and the standardized method of GPA calculation across institutions. However, studies like Light et al. (1987) demonstrate the utility of incorporating credits into analyses, as they found significant differences in credit hours earned across different TOEFL score ranges. These findings suggest that students' English proficiency may not only affect their academic performance as measured by GPA but also their ability to complete coursework, further emphasizing the value of including credits as an additional measure of academic achievement.

Considering these findings, this dissertation includes credits earned as additional dependent variables, alongside CGPA, to offer a more comprehensive view of the relationship between English proficiency pathways and academic achievement. This multi-faceted approach aims to address gaps in the literature and provide a fuller understanding of how different pathways influence student success in a Canadian higher education context.

Incorporating multiple measures of academic achievement, such as GPA and credits earned provides a broader perspective on student success and allows for a more nuanced analysis of how factors included in this dissertation impact different academic outcomes. With this foundation established, it is now important to examine English Language Proficiency (ELP) and the pathways through which international students meet the ELP requirements for university admission. By defining and operationalizing ELP and its various pathways, it is possible to understand how these mechanisms influence academic achievement and study progress, as measured by GPA and credits.

Defining and Operationalizing English Proficiency and Pathways

This study examines the relationship between English language proficiency pathways (i.e., IELTS, TOEFL, university coursework) and academic achievement. As such, how language proficiency has been conceptualized is a necessary component of this review. By establishing the conceptual framing of this construct, it will be possible to discuss how various tools (i.e., TOEFL/IELTS) seek to measure language proficiency – a key component of the university admission process and, therefore, an important aspect of English proficiency pathway analysis.

What is Language Proficiency?

Language proficiency is a difficult construct to define. It has been studied for some time with debate on “what constitutes ‘language proficiency’ and the nature of its cross-lingual dimensions” (Cummins & Oller, Jr., 1983, p. 138). Famously, Chomsky (1965) described a competent language user as one who functions homogeneously within a language community, unaffected by irrelevant performance conditions and can instead apply the language when called upon. However, where Chomsky placed emphasis on language competence being synonymous with grammar, Hymes (1972) developed the concept of communicative competence as an

extension of Chomsky's definition but recognized the sociocultural/situational nature of language use and the need to account for appropriateness and context (Murray, 2016). In other words, Hymes extended language competence to include more than just grammatical and sociolinguistic rules; it also includes the ability to recognize context given environmental and cognitive conditions (Murray, 2016). Other scholars throughout the remainder of the 20th century would build from this foundational conceptualization of communicative competence (e.g. Bachman, 1990; Canale & Swain, 1980), however the work of Cummins (1979, 1981, 2008) is of particular use for the current project as his conceptualization of communicative competence is specific to the educational context. As a result, further theoretical frameworks were established that more directly addressed second language (L2) concerns within the educational space.

The work of Cummins (1979, 1981) developed theoretical constructs to draw the attention of educators to specific challenges that students may face within an educational context. Cummins (1979) argued against a single, all-encompassing language proficiency dimension such as communicative competence, due to the danger of conflating L2 conversational fluency and L2 academic proficiency. He suggested these two aspects of L2 were very different and as such "proficiency cannot be considered to reflect just one unitary proficiency dimension" (Cummins, 2008, p. 73). Through a critique of communicative competence, Cummins established the basic interpersonal communicative skills (BICS) and the cognitive/academic language proficiency (CALP) distinction to emphasize the difficulty L2 students have in mastering English academic skills in comparison to English face-to-face communication skills (Cummins & Oller, Jr., 1983). This framework describes the gradual development and demands of learning daily English coupled with the demands of the academic and literacy requirements of the curriculum and

content at various levels of education (Roessingh & Douglas, 2012). Cummins (1984) further argued that an individual can become reasonably proficient in conversational English in about two years, but that they can take five to seven years to become competent in verbal-academic skills. Since Cummins' work, many studies have reported that nonnative speakers of English struggle with academic success due to perceived deficiencies in their academic ELP (Andrade, 2006; 2009). Furthermore, numerous studies have sought to examine the correlations between ELP and academic success, which are presented in detail below. (Stephan et al., 2004). It is difficult to estimate the influence of Cummins' theory within the educational field. Various scholars such as Scarcella (2003) have critiqued the framework, arguing for example, that the distinction does not assist teachers in the creation of tasks that may help students develop their lagging academic English; while others, including Cummins, have evolved the framework to include various other aspects (such as discrete language skills, in the case of Cummins, 2001). Another framework that builds off the work noted thus far and is useful for the current study is Murry's (2016) tripartite model.

Murray (2016) builds off the BICS and CALP distinction set by Cummins to present a tripartite model that seeks to give clarity to the somewhat problematic way language proficiency is discussed and addressed in the context of higher education. The three components Murray proposes are:

(1) *General proficiency*: which is similar to language knowledge and ability presented in communicative competency frameworks such as a la Hymes and Cummins' BICS. This is a starting point and refers to the generic skills and abilities of language learning such as grammatical structures, syntax, general listening, speaking, reading, and writing skills, and the various pragmatics of communication (e.g. politeness). Non-Native English Students (NNES)

require mastering these various aspects to be able to have basic engagement with higher education.

(2) *Academic literacies*: refers to discipline specific social and cultural knowledge and language skills. This distinction is close to Cummins' CALP concept; however, as international student enrolment has increased significantly since Cummins' initial conception, increased diversity and the lack of a shared background make any unifying assumptions regarding academic literacy problematic. Therefore, Murray argues that academic literacy development needs to be incorporated into and regarded as core or essential in higher education. He argues that EAP programs focus on study skills and other generic learning needs and do not promote enough academic literacy development to achieve the level of incorporation he deems necessary.

Lastly, (3) *Professional communication skills*: refers to interpersonal skills and knowledge of the changing contexts of professional workplaces. Due to the nature of an ever changing and culturally and linguistically diverse workforce, this third aspect of Murray's framework may be the most relevant to students who seek employment post-graduation. Professional communication skills are subdivided into six competences: intercultural competence, cultural relativism, interpersonal skills, conversancy in discourse and behaviours of specific domains, non-verbal skills, and group/leadership skills (Murray, 2016, pp. 93-97). While it is not necessary here to delve into the specifics of each of these sub-categories, or more details and examples of Academic literacy or general proficiency, it must be noted that these components of ELP are presented as key to the student experience in higher education and should be incorporated into learning for students to be successful.

This definitional and conceptual framework establishes that English language proficiency is far more than just the ability to listen and speak or to read and write, but instead presents post-

secondary language requirements as a multifaceted often contextual set of proficiencies, competencies, skills, and experiences. For a student to be successful in an English speaking post-secondary context, numerous language and literacy skills must be either present at admission (e.g. as evidenced through a standardized test score) or developed (e.g. through coursework) if the student is not proficient. As scholars have come to understand the construct of language proficiency, there has been significant research on the measurement of ELP and various ways to operationalize it.

The Measurement of English Language Proficiency

The continued internationalization of higher education has resulted in significant student mobility. Recent figures place the number of worldwide post-secondary international students at over six million in 2021, with over half of all international students receiving education in only five Anglophone countries (USA, UK, Australia, Canada, and New Zealand) (International Organization for Migration, 2023). Since most international students originate from countries where English is not the primary language (e.g., China,), English as a lingua franca in academic settings has propagated (Jenkins, 2013). The proliferation of international students and the demand for academic English has resulted in relatively standardized ELP requirements at universities and colleges for international students - with a high reliance on English language testing scores and English for academic purposes programming. For the testing route, institutions have adopted a wide range of approved language tests but by far the most common and widely accepted are the International English Language Testing System (IELTS) and the Test of English as a Foreign Language (TOEFL) (Hill et al., 1999; Hoekje et al., 2021; Oliver et al., 2012).

Due to the widespread adoption of IELTS and TOEFL, test score requirements for admission purposes are consistently reviewed by various stakeholders within the higher

education sphere with ongoing debate and research on these tests (Fox, 2005). While it is not the purpose of this dissertation to assess the psychometric characteristics of the IELTS or TOEFL, it is important to acknowledge debates regarding the tests due to the high reliance on language test scores for admission purposes due to their “supposed...objectivity and the fair and equal treatment of all test takers” (Piller & Bodis, 2024, p. 4). Furthermore, since ELP has been shown to be a significant factor in predicting the academic success of international students in English-speaking higher education institutions (Bo et al., 2023; Ghenghesh, 2016; Martirosyan et al., 2015; Neumann et al., 2019; Rose et al., 2020; Waluyo & Penmei, 2021), and ELP is often measured by IELTS and TOEFL, acknowledgement of key debates surrounding this measurement is key.

Research and discussion surrounding the IELTS and TOEFL take many forms and significant work occurred in the latter portion of the 20th century to understand the theoretical and conceptual design, measurement, usage, and validity and reliability of language testing (e.g., Bachman, 1990, Bachman & Palmer, 1996). This was perhaps due to the broad adoption of standardized language assessments (like the IELTS) for admission to English-medium higher education institutions, for migration purposes, and for employment and professional registration in English speaking contexts (Read, 2022). Such widespread adoption of these tests concerned (and continues to concern) researchers for a variety of reasons.

For example, debates continue about the subjective/objective view of language, or discussion surrounding whether language proficiency can be measured objectively through standardized tests or if language understanding and use are more subjective and influenced by context and individual experience. These debates call into question the ability of IELTS and TOEFL to effectively measure language ability (e.g., Pilcher & Richards, 2017; Shohamy, 2007).

Additionally, a vast body of literature has explored test validity, or the extent to which a test accurately assesses what it is designed or intended to measure (Brown, 2011). Much of this work centers on IELTS and TOEFL adoption into domains they were not originally designed for, such as immigration (Frost & McNamara, 2018; Hoang, 2019; Yao & Wallace, 2021). Results in this area are varied with scholars often debating the purposes of testing immigrants per se, but not necessarily demonstrating validity concerns (Pochon-Berger & Lenz, 2014). Further research has explored sub-section (e.g. reading, listening) validity (i.e. construct validity) or reliability - that is, the extent to which a test's results can be recognized as consistent or stable (Hashemi & Daneshfar, 2018). This subfield is extensive as each sub-section has a body of literature that may suggest various improvements to the IELTS or TOEFL such as implementation changes (Karim & Haq, 2014, Li, 2019), scoring changes (Alshammari, 2016), specific task changes (Moore et al., 2012; Peltekov, 2021), and more.

University Admission: English language proficiency challenges

Regardless of ongoing debates regarding the appropriateness of the measurement of English language proficiency or ongoing academic discussions about how to best implement such measurement, ELP requirements are in place across English speaking institutions of higher education to ensure that potential students are prepared for their academic journey (Wonnacott, 2024). In practice, there are many issues with the use of commercially available standardized tests as gatekeeping tools to university. While the tests themselves are typically “ready-to-use”, institutions must make critical policy decisions about cut scores (i.e. the minimum scores required to be admitted to the institution generally or to a program, specifically) and manage (or ignore) any validity concerns such as those noted above (Hamid et al., 2019, p. 230).

Additionally, there is a significant amount of academic discourse on the commodification,

marketization/ and neo-liberalization of higher education and the adoption of standardized English requirements within these processes (Heller, 2003; Dewey, 2007; Murray, 2018; Soto & Pérez-Milans, 2021). This discourse raises concerns regarding the competition of institutions for international students in an ever-expanding market whilst simultaneously seeking new and expansive funding sources in the form of international student tuition. Such English requirements require ELP admission policy which may be viewed as quite arbitrary, as minimum entry scores vary widely and rationale for their selection are not easily located. For example, this dissertation identified slightly differing entry scores at two institutions and no rationale for minimum score selection criteria were located. Therefore, while some may argue that institutions set English language requirements to ensure that students have enough proficiency so that low proficiency does not prevent success, students still struggle, which influences lecturers, and other academic and student support staff (Murray, 2013). Due to these challenges and concerns, both academic and government bodies have pressured institutions to provide adequate support for students and base their ELP decisions on sound theoretical grounds—particularly because poor performance can have significant financial and personal consequences for students (Murray & Hicks, 2016). Some countries have gone as far as implementing monitoring, compliance, and inspection systems to ensure that institutions have collected proper evidence that students have sufficient ELP to be admitted to study in English medium institutions (Pearson, 2021).

Since ELP has been shown to be a significant factor in predicting the academic success of international students, particularly in English-speaking higher education institutions (Bo et al., 2023; Ghenghesh, 2016; Martirosyan et al., 2015; Neumann et al., 2019; Rose et al., 2020; Waluyo & Penmei, 2021), it stands to reason that difficulties stemming from inadequate ELP can have substantial impacts on students. There has been significant research on how language

proficiency, as assessed through various standardized tests, can serve as a predictor of academic achievement within higher education. This literature is of high importance to this dissertation, as it lays the foundation for understanding the role of one ELP pathway in predicting academic achievement.

The Relationship Between English Language Proficiency and Academic Achievement

Since the 1970s, scholars have investigated the relationship between English language proficiency (usually demonstrated as scores on various standardized language tests) and academic achievement/performance typically indicated by GPA (e.g., Burgess & Greis, 1970; Hwang & Dizney, 1970). Often, these studies are termed as *predictive validity* research which is defined as studies that explore the extent to which test performance predicts some specified non-test language performance (Bachman, 1990) or studies that explore the extent to which a test can predict the test takers future language performance (Woodrow, 2006). Typically, predictive validity studies “calculate correlation coefficients between English proficiency test scores and first-semester grade point average” (Graham, 1987, p. 506). In this manner, GPA is an operationalization of academic achievement.

Scholarly interest in predictive validity research, which calculates correlation coefficients between ELP test scores and academic achievement, is based on a theoretical understanding that language proficiency is fundamentally linked to academic success. For example, the work of Cummins (1979) and the basic interpersonal communicative skills (BICS) and the cognitive/academic language proficiency (CALP) distinction explained above suggested that higher levels of language proficiency enable students to better comprehend academic material and engage in complex cognitive tasks. CALP in particular is essential for understanding academic content, analyzing information, and articulating critical thinking in both written and

oral form – skills necessary to succeed in the higher education context. Other theories such as Vygotsky's (1978) sociocultural theory, highlight the role of language as a critical tool for learning and cognitive development. Therefore, in an academic context, if students cannot fully comprehend lectures, readings, or discussions due to limited English proficiency, their ability to learn and perform academically is impeded. As a result of the foundational theoretical work on the significance of language proficiency to performance on academic tasks, it is theoretically assumed that students with higher English language proficiency will struggle less academically due to language-related challenges than students with poorer ELP. This assumption prompts the exploration of predictive validity studies to explore the relationship between ELP and academic outcomes.

Building from these foundational theoretical frameworks, predictive validity research has aimed to empirically assess the relationship between English language proficiency and academic achievement. Graham's work (1987), while relatively early in the field, provided a review of predictive validity studies and noted a lack of consistently high correlations or any "clear cut answers" (p. 513). Most of the studies reviewed by Graham explored the TOEFL, specifically. Graham posited many possible explanations for the lack of consistency in results up to this point. These included the fact that foreign students with low levels of proficiency are generally not admitted to university, a lack of sociolinguistic factors included in the analyses (e.g., individual attitudes to specific tasks on tests), as well as other non-academic and academic factors that are difficult or impossible to quantify and include in modelling (e.g., student personality, attitudes, motivation, math aptitude, high-school GPA and more). Despite these early concerns and limitations highlighted by Graham, the theoretical significance of English language proficiency in academic achievement has continued to encourage scholarly investigation. This may be due to

the proliferation of international student applicants (whose first language is not English) in English institutions of higher education; therefore, admissions departments wish to know how ELP affects academic success (Wongtrirat, 2010).

As a result of this continued interest, significant research has been conducted on the relationship between ELP typically demonstrated through standardized testing and academic achievement. To align with the purpose of this dissertation, the following sections of this literature review have been organized to discuss research on the predictive validity of TOEFL and IELTS in relation to academic achievement. Given that the focus of this dissertation is on the relationship between English proficiency pathways and academic achievement, this structure allows for a detailed exploration of the most widely used ELP tests in higher education (which were included in this dissertation) and provides a synthesis of established correlations with academic outcomes like GPA. Overall, this detailed analysis of standardized ELP tests supports this dissertation's investigation of English proficiency pathways, enabling a nuanced understanding of how different testing pathways impact academic achievement.

The relationship between TOEFL performance and academic achievement

Research on the predictive validity of TOEFL has been explored for some time. Light et al. (1987) examined the relationship between TOEFL scores, GPA, credits earned, and academic major for $n = 376$ international graduate students at a U.S. university. The authors found a correlation of $r = .14$ ($p < .05$) between TOEFL scores and GPA; however, they noted that this correlation was “too low to have any practical significance” (p. 225). Interestingly, when grouping students into subject areas, they found higher correlations for humanities/fine arts/social science students ($r = .24$, $p < .001$) than for science/math/business students (not statistically significant). Each academic major was also explored individually, revealing that

social science ($r = .22$), education ($r = .30$), and public affairs ($r = .30$) all produced correlations significantly higher than the aggregated sample ($r = .14$). The authors also investigated the GPA of students admitted below the university cut score and found that these students did not perform significantly worse than those with TOEFL scores above the cut score. This suggests that factors other than English language proficiency supported student success and highlights the importance of including discipline of study as a significant moderator variable in future modelling.

Recognizing the importance of Light et al.'s contribution, Wongtrirat (2010) conducted a meta-analysis examining studies beginning in 1987. The author synthesized existing research to 2009 which explored the relationship between TOEFL scores, GPA, and course completion. The meta-analysis consisted of 22 studies, with criteria including the use of TOEFL scores as a predictor and GPA or course completion as outcome variables. Wongtrirat's analysis included a total of 3,937 international students at both the undergraduate and graduate levels and reported an average correlation at all study levels of $r = .18$ between TOEFL scores and GPA. Wongtrirat noted that this finding indicates that TOEFL scores have "a small predictive ability (.181) on GPA of international students" (p. 34) in general. In comparison, and relevant to the current study, the author explored the average effect size on GPA for undergraduate international students only and found an average of .188 – which is not significantly different than the effect size noted for all study levels. Also relevant to this study, the author reported an effect size of .250 for TOEFL's ability to predict course completion for undergraduate international students. Wongtrirat's concludes that these findings demonstrate that "TOEFL has a small predictive ability on academic achievement of international students whether measured by GPA or course completion" (p. 45) According to Bakker et al. (2019), Wongtrirat's findings would be considered a small effect size (.20 when rounded), while Plonsky and Oswald (2014) would

categorize these results as under the .25 threshold to meet the small effect size cutoff. Therefore, the findings from Wongtrirat's meta-analysis indicate that the predictive ability of TOEFL scores on GPA is limited. The effect size is small or possibly negligible, depending on the criteria used. This suggests that while there is a statistically significant relationship between TOEFL scores and academic achievement operationalized by GPA, the practical significance is likely minimal.

While Wongtrirat's meta-analysis is extremely useful for highlighting the findings of over a decade of research into the predictive validity of the TOEFL, individual studies are still valuable if they provide detailed insights into specific contexts, populations, or variables that may not be fully captured in a meta-analysis. Cho and Bridgeman (2012) conducted one of the most thorough analyses on the relationship between TOEFL scores and academic performance (operationalized by GPA), utilizing correlation and hierarchical multiple regression analyses. The authors collected data from $n = 2,594$ undergraduate and graduate students from ten U.S. universities. They grouped students into business, humanities and arts, science and engineering, and social sciences and reported a correlation of $r = .18$ between TOEFL scores and GPA for undergraduates and $r = .20$ for graduates, while controlling for discipline of study. These values equated to 3% and 4% of the variance in GPA explained by TOEFL scores, respectively. At the discipline-specific level, correlations ranged from $r = .12$ for business students to $r = .25$ for social science students. Cho and Bridgeman complemented these analyses with expectancy graphs, a visual tool representing the relationship between predictor scores and outcomes. The authors reported that “the chance of being in the top 25% GPA category doubled for both undergraduate and graduate students when their TOEFL iBT scores were in the top 25%, in comparison with those in the bottom 25% TOEFL group” (p. 438). Therefore, higher TOEFL scores were associated with a lower likelihood of being in the bottom 25% GPA range. This

study, notable for its large sample size and number of institutions, underscores the importance of considering academic discipline in analyses and further helps set expectations regarding the effect of TOEFL performance on GPA.

Recognizing that the 21st century saw significant interest in research on the TOEFL and the introduction of new test versions, Abunawas's (2014) dissertation sought to provide "a meta-synthesis of studies examining the relationship between students' TOEFL scores and academic performance" (p. 3). As such, studies were included from 2000 to 2014. The author used GPA as the operationalization of academic performance. The study aimed to present the variance in effect sizes across studies and examine how factors like degree level, publication status, test version (computer vs. paper-based), and research setting (U.S. vs. international) moderated the TOEFL-GPA relationship. The meta-analysis included 47 independent effect sizes from 41 studies, encompassing a total sample size of 17,495 international students. Utilizing Comprehensive Meta-Analysis (CMA) software, Abunawas calculated an overall effect size of $r = .21$ (95% CI = .16 - .26) or only $r = .03$ higher than Wongtrirat's meta-analytic finding. This suggests that, despite the introduction of new test versions and the increased research interest in the 21st century, the predictive power of TOEFL scores on academic performance remains relatively stable and modest. The slight increase in the overall effect size—from $r = .18$ in Wongtrirat's (2010) meta-analysis to $r = .21$ in Abunawas's (2014) study—is not substantial enough to indicate a significant improvement in TOEFL's ability to predict GPA. While this slightly higher effect size does meet Bakker et al.'s (2019) criteria for a small effect size (.20), it does not reach the .25 threshold suggested by Plonsky and Oswald (2014). Furthermore, of the moderator variables examined, only the research setting emerged as a potentially significant, with studies outside the U.S. reporting higher effect sizes. Lastly, this study did not examine

other student predictors like age or gender due to insufficient data, which suggests a gap that this dissertation could contribute to by including these moderator variables in the analyses. The findings from this additional meta-analysis reinforce the conclusion that while TOEFL scores are statistically related to academic achievement, the practical significance of this relationship is limited, and other factors likely play a more substantial role in predicting students' academic outcomes.

Building on their previous work, Bridgeman et al. (2016) sought “...to peel back the layers of the onion to uncover variations in predictive validity results...”(p. 310) in the study of the TOEFL. Using a sample of $n = 787$ undergraduate international students who had taken the TOEFL iBT, the authors calculated correlations between TOEFL scores and combined grades from the first three terms of academic study. The authors reported a weak correlation ($r = .18$, $p = 0.05$) between TOEFL overall scores and GPA. Subsequently, the authors desegregated the sample to explore the relationship between TOEFL scores and GPA by study major (e.g., business or engineering), student nationality (e.g., China or India), and major and nationality together (e.g., Chinese vs Indian students in engineering). Depending on how groups were split, some correlations were not significant at one level but were at another. For example, speaking scores were weakly correlated with engineering students' academic achievement; however, when examining Chinese and Indian engineering students separately, a correlation of $r = .55$ was found for both groups. Such differences in results depending on the subsamples used suggested to the authors that simply looking at correlation in an aggregated sample of students “can lead to limited and often erroneous conclusions” (p. 317). Thus, this study provides valuable perspective on how data can be examined at various levels of analysis to produce more nuanced and complex results.

Other more recent studies have explored the predictive validity of TOEFL on academic achievement operationalized by GPA while also including potential moderators. Harsch et al., (2017) asked what predictive power TOEFL overall scores had on students' final academic grades, as well as if variables such as academic discipline and nationality influenced the relationship between TOEFL scores and academic success. Similarly to earlier TOEFL studies presented in this section (e.g., Cho & Bridgeman, 2012; Wongtrirat, 2010), the authors reported a small correlation ($r = .20, p = 0.01$) between overall TOEFL score and academic achievement measured by GPA. When exploring academic discipline, the authors reported a much higher correlation between TOEFL overall scores and academic achievement for Arts ($r = .46, p = 0.01$), followed by a moderate correlation for Science ($r = .25, p = 0.01$), and a weak one for Social Sciences ($r = .14, p = 0.05$). By nationality, the authors found the Chinese student subgroup to have the highest overall correlation between TOEFL scores and overall academic achievement ($r = .31, p = 0.01$) - while they did not find any statistically significant correlations between overall TOEFL score and GPA for the Indian and German subgroups. Ordinal logistic regression modeling revealed that TOEFL overall scores and nationality had the most predictive power, whereas section scores, faculties, and additional language support did not significantly enhance the models. The use of expectancy graphs demonstrated that students who scored higher in TOEFL tended to achieve higher academic grades, while students who failed or achieved lower grades tended to have lower TOEFL scores. The authors conclude that their work supports the conclusion that there is a meaningful relationship between ELP test scores and academic outcomes and that it is appropriate to use ELP test scores for admissions and placement decisions at English medium college and universities. This study highlights the importance of controlling for the effect of nationality as well.

The studies cited above explored, in various ways, the relationship between TOEFL scores and academic performance typically defined by GPA. Correlations varied somewhat, but in general they were found to be weak with small levels of variation attributed to the inclusion of specific moderator variables such as discipline of study or student nationality. Furthermore, meta-analyses of TOEFL studies have synthesized the effect sizes from numerous research efforts in this field from 1987 to 2014. Within the field, two of the most frequently cited meta-analyses are those by Abunawas (2014) and Wongtrirat (2010). Such meta-analyses are valuable because they provide insights into the expected strength of correlations for the TOEFL pathway in this dissertation and they offered perspectives on which variables have been deemed most significant. (e.g. discipline of study, student nationality).

This section on the predictive validity of TOEFL has shown that correlations between TOEFL test performance and academic achievement are generally small or weak. Some studies improved the strength of the correlation by controlling for moderator variables such as subject/discipline of study. This research informed the data collection of this dissertation and highlighted several limitations in the literature such as small sample sizes, data collection from single cohorts or single institutions, a general lack of additional moderator variables controlled for in analyses, and a lack of comparisons between performance on different English Language Proficiency tests and their associated outcomes. Like TOEFL, studies on IELTS have examined the relationship between test performance and academic success, with varying results. The next section will present this literature, providing a comparative perspective and further highlighting gaps in research.

The relationship between IELTS performance and academic achievement

Given its widespread use and significance, academic interest in the IELTS is both expected and justified. According to the British Council, a public corporation and joint owner of the IELTS, the test is the world's most popular English language proficiency exam used for higher education and global migration (British Council, 2017). The council claims that the IELTS is recognized by more than 11,500 organizations worldwide and that over three million people take the test annually. Such extensive adoption of a psychometric tool naturally attracts scholarly attention across different disciplines and research on the IELTS has been explored for some time (e.g., Davies & Criper, 1988; Kerstjens et al., 2000). As a result, numerous studies have investigated the predictive power of the IELTS, producing varying results over the years.

Many studies report a positive correlation between IELTS performance and quantitative measures of academic achievement – typically GPA (Al-Malki, 2014; Arrigoni & Clark, 2015; Feast, 2002; Hu & Trenkic, 2021; Müller & Daller, 2019; Schoepp & Garinger, 2016; Schoepp, 2018; Shakibaei & Memari, 2019; Thorpe et al., 2017; Woodrow, 2006; Yen & Kuzma, 2009; Yixin & Daller, 2014). Researchers have employed a wide range of analytical tools, sample sizes, and variables to explore this relationship. It is therefore pertinent to examine how these studies have investigated the correlation between IELTS performance and GPA.

Among the key studies exploring this relationship, Feast (2002) utilized a multiple regression technique called multilevel analysis to examine the link between IELTS scores and GPA ($n = 101$). The study found a statistically significant positive relationship, demonstrating that a one-unit increase in IELTS score resulted in a mean GPA increase of 0.39, assuming all other variables remained constant. The author noted that this relationship was relatively weak.

However, Feast found that certain moderator variables had more significant impacts. Specifically, level of study (a 0.79 GPA increase for postgraduate students compared to undergraduate students), discipline of study (a 1.2 GPA increase for students in Health Sciences compared to non-Health Science disciplines), and home country (a 0.99 increase in GPA for Chinese students compared to non-Chinese students) were statistically significant. In contrast, age, semester of entry, and gender were not significant factors. Unlike most TOEFL studies discussed above, this study included many different moderator variables and demonstrated their significance to statistical modelling. Thus, this study, although limited by its sample size, suggested the importance of controlling for various demographic variables—particularly the inclusion of study discipline and country of origin in statistical modeling.

Woodrow (2006) also found evidence for the predictive validity of IELTS, but as with the studies above, also noted weak correlations despite statistical significance. Using a sample of $n = 62$ international students from mainly Asian countries in the Faculty of Education and Social Work at the University of Sydney, Woodrow found a correlation (Pearson's $r = .40, p = <.01$) between all IELTS scores and first semester GPA. Additional analyses found that when the sample was split into sub-groups by IELTS score such as band 6.5 and below (those who just met entry requirements) and band 7.0 and above (those who had exceeded entry requirements)), only significant positive correlations were found for those scoring 6.5 and under ($r = .52, p = <0.01, n = 26$). The author also collected data for other variables that may be related to academic success for the study population, such as years learning English and time in Australia; however, Woodrow found no statistical significance through bivariate correlation analyses. This study was useful as it further suggested exploring IELTS scores by band while also including other

variables. However, limitations included a small sample size and a lack of demographic moderator variables.

Similarly, Yen and Kuzma (2009) investigated the predictive validity of IELTS scores on the academic performance of Chinese students at the University of Worcester. Focusing on a homogeneous sample of 61 Chinese Higher National Diploma (HND) students enrolled in business courses during the 2007/08 and 2008/09 academic years, they aimed to determine whether IELTS scores could reliably predict academic performance which was operationalized with first and second semester GPA. Pearson correlation analyses were conducted and a statistically significant result ($r = .46, p < .01, n = 61$) for first-semester GPA and a weaker relationship ($r = .26, p < .05$) for the second semester was found. Yen and Kuzma concluded that while IELTS scores have moderate predictive validity for initial academic performance, their effectiveness as a predictor decreases in subsequent semesters. They suggested that factors beyond language proficiency, such as cultural adaptation and familiarity with the UK educational system, might increasingly influence academic success over time. While this study is useful in suggesting decreasing correlation over time, it lacked the inclusion of various moderator variables, included quite a small sample, and did not employ more robust statistical modelling methods.

While the studies presented thus far are limited by small samples, this was not the case for Oliver et al. (2012) who collected data for 5,675 international students. The authors investigated the relationship between different types of evidence of English Language Proficiency (ELP) and the academic achievement of Non-English Speaking Background (NESB) students at an Australian university. The authors operationalized academic achievement by weighted average marks (WAM) which, they claimed, accounts for both the marks obtained in

each course and the credit value of each course. The authors used WAM to identify students who have achieved a score of 50% or higher, which they suggested indicated that students had met the minimum passing standards and had, therefore, completed their degree requirements. Using this measure, the authors conducted Pearson's Product Moment Correlation Coefficient analyses for undergraduate and postgraduate cohorts separately. They found weak but significant positive correlations ($r = 0.275$, $p = 0.000$) between IELTS scores and WAM ($n = 291$), particularly for Reading sub-scores. The authors also found that students admitted with standardized test scores generally performed better academically than those who were admitted through English foundation programs. Due to this finding, the authors suggested that standardized ELP tests like IELTS and TOEFL are more effective in predicting academic success than alternative forms of ELP evidence. This study provided additional information about the relatively weak relationship between IELTS performance and academic achievement while offering some insight into ELP comparisons. However, while the sample for this study was quite large at the aggregate level, when students were grouped into their respective ELP pathways, this resulted in some small sample sizes (e.g. 291 students for IELTS). Furthermore, as with most studies reviewed in this dissertation, this study focused on a single institution and lacked the inclusion of moderator variables. Lastly, the authors acknowledged that the use of 50% WAM thresholds may be disputed by scholars as it breaks from the typical norms of using GPA or credits separately as operationalization's of academic achievement.

Another study with a significant contribution to understanding the predictive validity of IELTS, as well as providing insight into including additional measures of academic achievement in predictive validity research was the work of Arrigoni and Clark (2015). This study investigated the appropriateness of IELTS cut-off scores for admissions and placement decisions

at the American University in Cairo (AUC), an English-medium university in Egypt. The primary aim was to determine whether IELTS scores could predict the academic achievement of non-native English-speaking undergraduate students in their English language support programs and writing courses. The authors collected data from over 1,100 undergraduate students who entered AUC between Fall 2010 and Spring 2013 with IELTS scores. The data included students' overall IELTS scores, sub-scores (Listening, Reading, Writing, Speaking), course placements in the Department of English Language Instruction (ELI) and the Rhetoric and Composition Department (RHET), course outcomes (i.e., grades) and GPAs. In general, the authors found varying results of weak to moderate correlations between overall IELTS scores and both first semester GPA and overall GPA. Results varied depending upon the specific course (i.e., ELI 98, 99, 100, or RHET 101, 102). For example, the authors reported correlations of .14 ($p < 0.05$) and .23 ($p < 0.05$) for ELI 98 and REHT 102, respectively between overall IELTS scores and overall GPA. Overall, the authors found very little correlational difference between first semester GPA or total GPA and overall IELTS performance. The authors conclude that weak correlations are typically the norm in the literature; however, this lack of correlation may be due to “the measures of academic achievement used” (p. 24). While this study does have limitations such as the absence of detailed statistical reporting, limited control over moderator variables, and the lack of advanced statistical methods, the authors recommendation of using more than one measure of academic achievement when conducting predictive validity research was particularly valuable to this dissertation.

Similarly to research cited thus far, Schoepp and Garinger (2016) found weak/low correlations in their study that gathered IELTS scores and GPA from students at a federal university in the United Arab Emirates where English is the language of instruction. Using

Pearson's correlation coefficient, the authors demonstrated that overall GPA was weakly correlated with higher IELTS scores ($r = .21, p = 0.01, n = 241$). When splitting the sample into three score groups, students with an IELTS score of seven or higher outperformed groups with scores of 6 and 6.5 in mean GPAs. Statistically significant differences in GPA by IELTS group and different academic streams (composition, global studies, science, math, and technology) further suggested the importance of including discipline of study in statistical modeling and further highlighted the importance of not treating all student with IELTS scores as a single homogenous group, but that splitting scores and comparing could be valuable (as demonstrated by Woodrow, 2006). As with many studies explored for this dissertation, there was a lack of moderator variables controlled for (age, gender, etc.) in this study too.

Hu and Trenkic (2019) also conducted a study that examined the predictive validity of IELTS scores on academic achievement while considering how discipline of study moderates this relationship. The study involved 153 Chinese master's students at a UK university, all of whom had taken the IELTS exam to meet language entry requirements. Using hierarchical regression modelling, the authors demonstrated that IELTS scores submitted at university admission and the linguistic demand of the program in which they were enrolled were both significant predictors of average mark achieved. Modeling predicted that "an increase in IELTS score of one band (in the range 5.5 – 8.0) results in an average improvement of the weighted average mark of 4.76 points." (p. 17). Students were further categorised into more linguistically demanding programs (e.g., TESOL, Applied linguistics, English literature) or less linguistically demanding programs (e.g., business, management or marketing programs). When students were grouped together without accounting for study program, IELTS was found not to be predictive of academic achievement; however, when program linguistic demand was

accounted for, results were statistically significant and varied by programme classification. For more linguistically demanding programs, "...IELTS scores accounted for 14% of the variance in academic performance ($F(1,78)=12.93, p=.000$), with each band increase in IELTS overall scores leading to an increase in the weighted average mark of about 8.85 points" (p. 17-18). In less linguistically demanding programs, IELTS scores were found to account for comparatively only 6% of the variance in academic performance ($F(1,71)=4.13, p=.046$), "...with each increase of one band in IELTS scores leading to an increase in the weighted average mark of about 3.69 points" (p. 18). These results demonstrate that the predictive validity of IELTS scores is not uniform across disciplines and is influenced by the linguistic demands of the field of study. Hu and Trenkic's (2019) findings align with earlier research emphasizing the importance of considering the discipline of study when evaluating the relationship between English proficiency and academic achievement.

Up to this point, this review has focused on individual studies that examine the relationship between IELTS performance and academic achievement, highlighting their singular findings, contributions, and limitations. However, similar to the TOEFL section above, meta-analyses are instrumental in synthesizing these individual studies to provide a broader perspective on the predictive validity of IELTS scores. Relatively recently, Gagen (2019) conducted a meta-analysis of studies on the correlations between IELTS scores and GPA published between 1998 and 2019. The author conducted a comprehensive literature review, coded 28 studies and included 18 quantitative studies in their final analysis. Gagen reported an overall effect size of $r = .227$ from 29 effect sizes across 17 studies with a total sample size of 2,432 students. The author also identified one possibly problematic study with the lowest effect size and the smallest sample and thus removed this study from the sample to further test the

overall effect size of the remainder. Doing so increased the overall effect size to $r = .236$; however, this increase is not particularly significant. Gagen notes that many predictive validity IELTS studies did not include moderators. But among the moderators examined, the IELTS reading subscore and the influence of top-up programs (i.e. the effect of additional English language preparation courses) stood out as significant factors affecting the predictive validity of IELTS scores on academic success. The other moderators —field of study, level of study, study location, and timepoint of GPA measurement — generally did not show significant differences. Gagen’s findings align with the general consensus in the literature that, while IELTS scores are statistically significant predictors of academic achievement, there is “an overall approaching-small effect size between IELTS scores and post-secondary GPA” (p. 49).

When reviewing the literature on the predictive validity of IELTS, it is clear that while there is a relationship between IELTS performance and academic achievement, typically operationalized by GPA, this relationship is weak. For instance, Müller and Daller (2019) suggest that an effect size of $r = 0.30$ is common in the literature. In general, an effect size of 0.2 is considered small, while 0.5 is considered medium, with 0.8 considered as a large effect (Bakker et al., 2019). However, in the field of language testing, Plonsky and Oswald (2014) analyzed 175 correlational studies and concluded that effect sizes of 0.25 are small, 0.40 are medium and any above 0.60 are large. While some predictive validity IELTS studies have reported higher correlations, implying a medium effect (e.g., Dang & Dang (2023) found $r = 0.44, p < .001, n = 80$), these findings deviate from the norm as discussed in this dissertation. Generally speaking, many predictive validity IELTS studies have been limited by small sample sizes, the lack of multi-institutional data, the absence of complex statistical methods, and the lack

of consideration for important moderator variables such as age, gender, discipline of study, and cultural/national background.

Given these findings and constraints, it becomes essential to explore the literature on the comparison of academic outcomes based on different English proficiency pathways through which students gain admission to higher education institutions – as this topic is directly relevant to this dissertation. This aspect involves examining whether students who meet English language requirements through standardized tests like IELTS perform significantly academically different compared to those who enter via alternative pathways, such as completion of English preparatory programs or foundation courses. Therefore, the next section of this review delves into studies that compare academic achievement based on English proficiency pathways.

English Proficiency Pathways and Academic Achievement

As demonstrated in the previous section, there is a significant amount of research on the predictive validity of individual English language proficiency assessments (i.e. TOEFL and IELTS). Conversely, there is a distinct lack of research which examines and compares the academic outcomes of international students who are admitted through different English proficiency pathways. Therefore, this section will provide a synthesis of relevant literature and demonstrate the knowledge gap that this dissertation contributes to.

Firstly, Schoepp (2018) examined differences between two groups of undergraduate students—Direct Entry (DE) and Foundation (FO) students—to understand how English language proficiency impacts their academic performance. Within an English as a Medium of Instruction (EMI) environment in the United Arab Emirates (UAE), the author explored differences in GPAs between 417 DE students (i.e. students who met minimum IELTS

requirements to enter bachelor studies directly) and 536 FO students (i.e. students who did not meet the IELTS requirements and spent at least one semester in an English preparatory program before entering bachelor programs). Utilizing independent T-tests as well as Pearson Product-Moment Correlations, the study revealed that DE students had a significantly higher mean IELTS overall score of 6.1 compared to 5.3 for FO students ($p < .01$), indicating a stronger initial English proficiency. Correspondingly, DE students achieved higher cumulative GPAs at the end of their general education courses, with a mean CGPA of 2.79 versus 2.41 for FO students ($p < .01$). Notably, positive correlations were identified between IELTS scores and academic performance across the entire cohort, with an overall correlation coefficient of $r = .307$ between IELTS scores and CGPA ($p < .01$). Schoepp's (2018) study highlights significant differences between the two educational pathways, DE and FO. The study suggests that the pathway through which students enter university has an impact on their academic success due to differences in initial English language proficiency and the effectiveness of preparatory programs. Furthermore, these findings suggested to the author that lower IELTS admission requirements bolstered with additional language support may be insufficient in supporting students and ensuring their academic success. Schoepp's work was limited by its single institutional context, the use of GPA as the sole measure of academic achievement, the lack of more advanced statistical techniques, and the exclusion of moderator variables; however, this study contributes to the relatively small body of literature which compares pathways.

Exploring equivalency through a sample of student data, Tweedie and Chu (2019) compared the academic achievement of international students in a Canadian research institution's engineering program who met university admission ELP requirements via IELTS, TOEFL, or completion of an EAP program. With a sample of $n = 254$, the authors found statistically

significant differences in student completion rate and academic performance “depending on whether IELTS, TOEFL, or EAP was used for admission” (p. 692). EAP students specifically had lower completion rates than their test taking peers. For example, “[w]hile 30.7% of students who were admitted using the TOEFL completed all 10 courses...only 13.8% of the students admitted using EAP completed the same courses” (p.691). Also, the performance of TOEFL students was noted as higher than EAP students in a specific course, even though the EAP program was designed to support students within that course. Thus, the researchers found that students submitting test scores performed better in this specific course despite the preparation of EAP students. IELTS students were also noted to have outperformed TOEFL students in specific courses. The authors found that such inconsistencies between different measures in terms of success outcomes emphasize the importance to reassess the assumption that such measures are indeed equivalent. This study was one of the first that addressed questions similar to the current study and it demonstrated the importance of examining the equivalency assumption between IELTS, TOEFL, and EAP pathways. While the authors work is limited by the sample of a single cohort in a single program, their work served as a foundational paper for this dissertation.

In another Canadian study, Johnson and Tweedie (2021) explored nine different measures of ELP (TOEFL iBT, TOEFL PBT, IELTS Academic, CAEL, MELAB, PTE, CAE, PTE, and EAP) and examined how they may predict post-secondary academic achievement (measured by GPA) across different programs over a seven-year period. The authors collected data from a large sample ($n = 1918$) of English as an additional language (EAL) students which represented 107 different nationalities and 19 different academic programs. They also collected first semester GPA and final semester GPA and calculated Pearson correlation coefficients to estimate “the capacity of the different ELP tests and EAP courses [to predict] student success in academic

programs” (p. 107). Results indicated that entering through the IELTS pathway did not predict first semester student performance significantly ($r = .054, p = .495, n = 163$), and only demonstrated a weak relationship in the final semester of study ($r = .199, p = .011, n = 162$). Conversely, TOEFL showed significant but weak associations for both first ($r = .246, p = .000, n = 341$) and final semester GPA ($r = .218, p = .000, n = 340$). Conversely, results from students who completed EAP courses in each of the four skill areas varied but overall demonstrated stronger performance than students who were admitted through standardized test results. Specifically, EAP reading outcomes demonstrated a weak association with first semester performance ($r = .246, p = .005, n = 127$) but improved to a moderate association with final semester performance ($r = .386, p = .000, n = 127$). EAP listening and speaking course results were presented together, and similarly to reading, demonstrated a weak association in first semester performance ($r = .205, p = .021, n = 126$), but a moderate relationship with final semester performance ($r = .326, p = .000, n = 125$). EAP writing did not significantly predict first semester grades and only weakly predicted final semester GPA ($r = .280, p = .002, n = 125$). The authors noted that EAP courses, despite their lack of discipline specific content, substantially outperformed all the standardized measures of ELP in terms of predictive capability in relation to academic achievement. Overall, the authors found that IELTS and TOEFL weakly predicted student success (explained 4% to 6% of variance in GPA,) as measured by GPA in both first and last semester of study. Conversely, the EAP courses predicted between 2 to 6% variance for first semester performance and 8 to 16% for final semester performance. This study is significant in that it has one of the largest sample sizes of any single study reviewed, and for the purposes of exploring pathway outcomes, this study directly compares the strength of each pathway in predicting academic success. The authors argue that their findings raise “unsettling

questions about institutional practices for benchmarking English Language Proficiency of prospective students” (p. 111), and their findings do suggest an issue with equivalency between pathways. Furthermore, while EAP programs may be cost prohibitive to some students, results from this study suggest that such programming may result in better outcomes than students who take the standardized test route. This study is significant in that it provides direct analyses of EPP in terms of student success and displays how strong each pathway is in predicting outcomes. It also has a significantly large sample – although, as the authors noted, when divided among the 9 ELP measures some sample sizes became unusable (<25 cases). This study was limited in that it did not control for various demographic factors in statistical modelling, nor did it present predictions of GPA increases/decreases between different EPP. However, the authors convincingly demonstrated the need for future research in the area and provided key insights into areas of future research.

While studies thus far have been limited by their lack of moderator variables included in more complex statistical modeling, this was not the case in recent work by Clark et al. (2021). The authors examined the impact of Intensive English Language Program (IEP) attendance on the academic achievement of international students at a large research-intensive university in the southwestern U.S. The study compared two groups of students: those admitted based on their English proficiency test scores (TOEFL or IELTS) and those admitted through IEP completion. Utilizing OLS regression, and operationalizing academic outcomes as first-semester GPA, credits earned, and academic standing (i.e., academic probation status), the authors also controlled for several demographic variables, such as nationality, age, gender, and academic major. In their initial model ($n = 4,888$), the authors found that IEP students had a 0.126 lower GPA ($p < .001$) compared to non-IEP students. Additionally, students from Asia and the Middle

East had GPAs that were 0.182 and 0.482 lower, respectively, than their counterparts. Majors such as business and engineering were associated with higher GPAs, while liberal arts majors had lower GPAs. Interestingly, when the sample was restricted to students with TOEFL scores within ± 1 of the cutoff (around 61), the differences between IEP and non-IEP students became statistically insignificant. In additional statistical models, IEP students had a slightly higher GPA (0.029), earned more credit hours (0.058), and were more likely to be in good academic standing, though none of these differences were statistically significant. Additionally, as the bandwidths around the TOEFL cut score were tightened (e.g., ± 1 to ± 10 from the TOEFL cutscore of 61), the differences between IEP and non-IEP students continued to narrow, reinforcing the similarity between the two groups. For this dissertation, this study is quite significant as it is one of the few found that control for various moderator variables and explores direct comparisons between the testing and the course-based pathway. While the results of this study demonstrated to the authors that IEP students performed similarly to their non-IEP counterparts, they raised concerns about the lack of long-term data, as well as the need for institutional comparisons.

In a recent study, Ruegg et al. (2020) further identified that there was a significant lack of research exploring the relationship between academic achievement of international undergraduate students and ELP requirements met through different pathways. The authors explored the relationship between English proficiency pathway and academic performance by collecting data from 747 international students in 2014-2017 in the New Zealand tertiary education context. They operationalized academic performance as number of credits attempted, number of credits achieved, and cumulative GPA. Regarding English proficiency pathways students were coded into four groups:

- Students who met ELP requirements because they had studied 2 years of secondary education in the country and, therefore, entered through the same admission procedures as domestic students.
- Students who met ELP requirements by achieving the admission standard on IELTS (overall band score of 6.0, no sub-score below 5.5)
- Students who met ELP requirements by demonstrating they had studied in English medium education for 1 year or more prior to admission.
- And students who met ELP requirements by enrolling and completing an English proficiency programme and passing an “in-house English proficiency test, which is considered equivalent to large scale ELP test requirements” (p. 1290).

Surprisingly, as the authors acknowledged, students who met ELP requirements by studying 2 years of secondary education in the country prior to admission performed, on average, the worst in comparison to the other three groups both in terms of credits achieved and in overall GPA ($p < 0.001$). These students, who studied in New Zealand prior to admission to university, were dubbed by the authors as being “at risk compared to other groups of students... in terms of successful completion [and]... academic success” (p. 1293). There is no clear explanation for this unexpected finding, although the authors suggest that such a result may point to deficiency in the secondary curriculum or its delivery. While Ruegg et al., (2020) provided an outline of how different pathways may be identified/coded, their work is limited by the lack of delineation between different ELP admission tests as unique pathways, and as stated by the authors, a limited amount of data impacted the type of analyses that were conducted. Also, no moderator variables were controlled for in the analysis, nor were any more complex analyses presented to

demonstrate prediction on how much GPA or course completion may be impacted by admission pathway. While the authors recommended that future research should collect qualitative data to understand factors that enable or inhibit the academic success of university students who meet admission requirements through different pathways, the limitations noted above also warrant expanded quantitative work that this study seeks to contribute.

Building upon an acknowledged gap in studies that compared the predictive power of different ELP assessments (e.g., Hill et al., 1999), Ihlenfeldt and Rios (2023) meta-analyzed 132 effect sizes from 32 studies and synthesized and presented the predictive evidence of both IELTS and TOEFL instead of focusing on one test as found in studies cited above (e.g. Abunawas, 2014; Gagen 2019). The authors examined the degree to which standardized ELP assessments for the purposes of admission predicted GPA in higher education. TOEFL and IELTS performance were not found to differentially predict success in either undergraduate or graduate school and the authors produced an overall correlation of $r = .23$ ($p < .001$) which aligned with previous meta-analyses on individual assessments cited above (e.g. Wongtrirat, 2010). Furthermore, the study considered if data were generated from private or public institutions, the first meta-analysis to include this distinction. However, this distinction in school type was not found to be significant. The authors argue that ELP assessment performance does positively correlate with GPA and that their data stem from “all available studies... analyzing the validity evidence of the TOEFL and IELTS in predicting some form of GPA” (p. 19). Therefore, the authors claim, admission decision makers do not need to differentially weigh TOEFL and IELTS assessments for the purposes of meeting ELP admission requirements. In other words, both assessments are equivalent in their predictive capabilities regarding academic achievement; and therefore, should be treated as equivalent for admission purposes. This claim of equivalency is

important, as this dissertation explores this assumption by examining academic outcomes for students who meet ELP requirements through different pathways.

Equivalency claims have been noted as particularly problematic. Pearson (2020), after a review of studies which explore the effectiveness of EAP programs, recommended that such programs should not be marketed as equivalent to direct entry (i.e. sufficient scores on standardized tests). The basis for this claim was that academic outcomes and students' perceived academic experience were found to differ (e.g. Thorpe et al., 2017) in a manner that was detrimental to EAP students when compared to direct entry students. Pearson recommended that non-native English speaking students receive a disclaimer on university pre-sessional webpages that undertaking EAP programming in the UK is "potentially more challenging, less beneficial, and has a higher rate of failure" (p. 442). Considering that universities have been criticized for using EAP programs as a significant revenue stream (e.g. Corcoran et al., 2022), it is unlikely that institutions would place such a disclaimer which may ward off potential student customers. However, it is not clear from the literature that EAP programs (or equivalent options) are indeed problematic at delivering the skills and competencies in language that are necessary for success in English university, as some studies have found "significant measurable benefits" of such programming (Goldsmith et al., 2022, p. 11). It is this lack of consensus noted in the literature thus far that this dissertation contributes to.

Existing pathway research has deepened our understanding of how IELTS, TOEFL, and course-based options relate to academic achievement, yet nearly all comparative work has been confined to single-institution samples (e.g., Schoepp 2018; Tweedie & Chu 2019; Clark et al. 2021; Ruegg et al. 2020). Recent meta-analyses (Gagen 2019; Ihlenfeldt & Rios 2023) confirm that little research has applied a common analytical framework across multiple universities,

leaving open the question of whether institutional context itself shapes pathway effects. To close this gap, the present dissertation purposefully contrasts cohorts from two Canadian universities—thereby providing the first cross-institutional test of pathway equivalence in the country.

Equally important, prior studies seldom control for moderators such as discipline, gender, age, or country of origin, factors repeatedly shown to influence GPA and credits earned (see below). Leveraging a larger multi-institutional dataset, this study incorporates those key predictors wherever the participating universities could supply reliable data, even though other potentially relevant variables (e.g., socio-economic background, prior academic performance, etc.) were not available. By integrating institutional comparisons with a richer set of controls, this dissertation offers a more robust assessment of how different English-language pathways translate into academic outcomes.

Moderators of Academic Achievement

Given the limitations noted above in the predictive validity literature, as well as the lack of research exploring the relationship between English proficiency pathways and academic achievement, it was imperative to attempt to include additional factors that may influence students' academic achievement, as this is a significant gap. However, the relationship between various factors and academic achievement (operationalized by GPA or otherwise) is complex. Factors such as gender, age, country of origin, and academic major have been identified in the literature as significant predictors of academic achievement. Incorporating these variables into statistical modeling for this dissertation can provide more accurate results. The following sections will delve into how each of these factors influences academic outcomes and the rationale for including them as moderator variables in this study.

Gender

There has been significant work and acknowledgement in the literature for at least the last twenty years about a gender gap in academic performance/achievement and more broadly regarding academic/student success. Studies typically try and control for gender effects due to broad acknowledgement of its association with academic success (Sheard, 2009; Van Herpen et al., 2017). Chee et al. (2005), based on survey data ($n = 659$) from a medium sized university in the USA found that women tended to have a higher GPA than men ($p < 0.001$) even though men were found to have higher SAT scores and more highly educated parents. Various other predictors were identified as positively influencing GPA that were specific to each gender. For example, participating in clubs was found to be a positive predictor of GPA for women, but not for men. The authors concluded that “factors associated with GPA do indeed differ by gender” (p. 611). Other studies have found that while some predictors are significant for both genders (e.g. self-efficacy), other predictors have been shown to be significant for men only (e.g., active learning strategies, performance goals), providing “...evidence for gender differences in motivation and performance... [and] gender-specific correlates of academic achievement” (Pirmohamed et al., 2017, p. 320).

Additionally, Conger and Long (2010) further examined the male disadvantage in GPA, credits earned, and persistence in college in the US. By using multiple administrative data sets from 11 public four-year universities in Florida, Florida state census data collected in high school, and census of enrollees in Texas universities, the authors were able to follow students as they moved through the educational system and produced a large data set ($>25,000$). Women were found to be more likely to enroll in college, to earn more credits (men earned 0.43 fewer credits than women in the first semester and 6.6 credits fewer by the end of the sixth), and to

earn higher grades than men (GPA gap of 0.20 in each semester favouring women). Men were found to be disadvantaged due to lower high school performance and by choosing majors with lower GPAs. Despite controlling "...for backgrounds, educational needs, high school quality and performance, and college quality and major..." men were still found to be underperforming women in university.

There are many possible reasons for why the gender gap in academic achievement persists. Sax and Harper (2005) found that gender differences in college outcomes are largely unrelated to the college experience itself, but more so related to pre-college development of "different values, confidences, aspirations, and patterns of behaviour" (p. 21). Additionally, research has examined personality differences between genders as a potential mediator, with Carvalho (2016) highlighting how these differences may influence academic performance. Other studies, such as those by Sonnert and Fox (2012), have explored the impact of institutional type, noting that the environment and resources available at different types of institutions (e.g., those classified by the Carnegie system) can also play a significant role in shaping academic outcomes by gender.

While it is not the goal of this dissertation to delve into the underlying causes of the gender gap in academic achievement, it is crucial to acknowledge and control for gender as a variable. Given the literature documenting gender differences in GPA and other academic outcomes, including gender as a control variable ensures a more accurate analysis of how English language proficiency pathways influence academic achievement. By doing so, this dissertation aims to provide a clearer understanding of how these pathways impact academic achievement when considering demographic groups such as gender.

Age

Student age is another predictor that has been explored as influencing academic achievement in higher education. For example, Cassidy (2012) reported age as a significant predictor of GPA with a R^2 change of .201 when included in regression modelling and a .414 increase in GPA when treating age as a continuous variable. Other research has found similar results that older/more mature students achieve a higher CGPA than younger students (Sheard, 2009). Age has long been included as a demographic predictor (along with gender and others) in research that sought to predict student outcomes (Hellas et al., 2018; Strayhorn, 2006). This research has generally found an effect of age as a meaningful predictor (Vella et al., 2016). Due to its acknowledged effect on academic achievement, age was included as a control variable in modelling for this dissertation.

Country of Origin

In addition to gender and age, country of origin or country of citizenship is another crucial demographic factor that can significantly influence international students' academic achievement in higher education. Research has shown that students from different cultural, linguistic, and ethno-racial backgrounds often face challenges that can impact their academic achievement (Chen et al., 2020; Lehmann & Meldrum, 2021; Makarova & Birman, 2015; Warikoo & Carter, 2009). For example, Grayson (2009), who drew on a sample of students at York University in Toronto, found "students who reported Arab/West Asian origin and Blacks achieved lower high school grades than students of European origin" (p. 53). Additionally, a 'country of origin effect' (Parasnis & Swan, 2020) and 'region-of-origin differences' (Vaquera & Kao, 2012) have been noted as moderately significant in relation to academic outcomes. For instance, Bista (2014) examined how learning outcomes differ in students from East Asia, South

and Central Asia, and Southeast Asia at the undergraduate and graduate level in the United States. Statistically significant differences were found among the three different groups of Asian students in terms of their college outcomes. The author found a significant multivariate effect by students' country of origin/citizenship suggesting that Asian students' college outcomes vary across regions of Asia and that they are culturally and academically different.

Furthermore, studies that have examined the relationship between ELP and academic success (frequently operationalized through GPA), reveal that country of origin/citizenship is a notable factor. For instance, foundational research by Graham (1987) and Light et al. (1987) demonstrated how correlations between TOEFL scores and GPA could vary across academic disciplines and regions of origin. Subsequent meta-analyses and empirical studies (e.g., Abunawas, 2014; Cho & Bridgeman, 2012; Harsch et al., 2017; Hu and Trenkic, 2019; Schoepp & Garinger, 2016; Wongtrirat, 2010) have further support the significance of including a moderator representing nationality or country of origin when predicting academic achievement. Analyses have shown that students from diverse backgrounds may experience varying levels of academic success based on ELP pathways, revealing how demographic factors like country of origin can moderate these relationships (Johnson & Tweedie, 2021; Ruegg et al., 2020; Yen & Kuzma, 2009). Meta-analyses by Gagen (2019) and Ihlenfeldt and Rios (2023) reaffirm the importance of controlling for citizenship/country of origin to fully understand the predictive validity of standardized ELP measures. Thus, country of origin remains essential in moderating the relationship between ELP and GPA, making it a critical factor to include in any analysis of international student success in higher education.

Given these findings, country of origin/citizenship was controlled for in this dissertation. By controlling for this variable, this study aims to provide a clearer understanding of the factors

that contribute to academic achievement among international students, particularly in the context of English language proficiency pathways. Therefore, including country of origin allows for a more nuanced analysis of how cultural, linguistic, educational, and regional differences impact academic achievement.

Academic Major

Academic major, while not a demographic factor, is another predictor that has been demonstrated to influence academic outcomes in higher education (Leppel, 2001). For example, when exploring the influence of first-generation status on cumulative GPA using hierarchical multiple regression techniques, Strayhorn (2006) included academic major as an independent predictor (along with others such as gender and race). Major was found to be a significant independent predictor when many other variables included in the modelling were not (e.g. average hours worked while enrolled, job on/off campus). More recently, Gipson (2018) contributed one of the most comprehensive analyses of the contribution that academic major makes to predictive modelling of GPA. Gipson sought to predict cumulative GPA (as well as four-year and six-year graduation) while accounting for various pre-college characteristics (e.g. high school GPA), select during college characteristics (e.g. first year GPA), and the interrelationship between student-level and major-level predictors using multi-level modelling (MLM) techniques. Average CGPA was found to significantly vary across majors with 30% of the variance in CGPA attributed to differences across academic majors. Also, academic majors were found to account for a large proportion of variance in four-year and six-year graduation. As Gipson suggests, "...academic environments play a critical role in academic outcomes" (p. 69).

Studies exploring the relationship between ELP and GPA have shown that correlations often vary by academic discipline, emphasizing the need to control for study major in analyses of

academic outcomes. For instance, Light et al. (1987) found stronger correlations between TOEFL scores and GPA for students in humanities and social sciences compared to those in science, math, and business fields. Similarly, Cho and Bridgeman (2012) identified higher predictive validity for TOEFL scores within social sciences compared to other disciplines, while Bridgeman et al. (2016) highlighted how this effect can shift when accounting for both discipline and nationality. The impact of study major extends to IELTS as well. Schoepp and Garinger (2016) showed that academic discipline influenced GPA outcomes across IELTS score groups. Moreover, Hu and Trenkic (2019) demonstrated that the predictive validity of IELTS scores was significantly higher for linguistically demanding programs. Recent meta-analytic research by Gagen (2019) also confirmed discipline as a moderating factor for IELTS-GPA relationships. Large-scale studies by Johnson and Tweedie (2021) and Ruegg et al. (2020) further reinforced these findings, indicating that academic programs strongly mediate GPA outcomes across various ELP pathways. Given the substantial evidence linking study major with academic achievement for students meeting ELP requirements, it was essential to include academic major as a control in this dissertation.

To conclude this section on additional predictors of GPA, it is clear that a wide range of factors, from demographic characteristics like gender, age, and country of origin, to academic and institutional variables such as study major, can influence academic achievement as operationalized by GPA. Understanding research on these predictors is essential for modeling and interpreting the relationship between English language proficiency pathways and academic achievement. By controlling for these variables, this dissertation aims to provide a more precise and nuanced analysis of how English language proficiency pathways impact academic outcomes.

Conclusion

This literature review has provided a comprehensive examination of the key areas relevant to this dissertation, which investigates the relationship between English proficiency pathways and academic achievement. The chapter began by defining and operationalizing academic achievement, highlighting the challenges posed by terminological inconsistencies in the literature. Terms like academic success, academic achievement, academic performance, and student success are often used interchangeably, leading to difficulties in synthesizing research findings and drawing clear conclusions. By clarifying these terms, the review established that academic achievement is typically operationalized through measurable outcomes such as Grade Point Average (GPA) and credits earned. GPA emerged as a widely accepted and utilized measure due to its predictive power for various aspects of student success, including persistence, retention, and post-graduation outcomes. Credits earned, although less commonly used, was also identified as a valuable supplementary measure, providing insights into students' progress toward degree completion.

The chapter then moved on to defining and operationalizing English Language Proficiency (ELP), exploring how ELP has been conceptualized and measured in higher education contexts. The evolution of ELP as a construct was traced through theoretical frameworks such as Cummins' Basic Interpersonal Communicative Skills (BICS) and Cognitive Academic Language Proficiency (CALP) distinction, which differentiates between conversational fluency and academic language proficiency. Murray's tripartite model further expanded this understanding by introducing general proficiency, academic literacies, and professional communication skills as integral components of ELP in higher education. The measurement of ELP was discussed in the context of standardized tests like IELTS and TOEFL,

which are widely used for admissions purposes despite ongoing debates about their validity and effectiveness. The challenges institutions face in setting appropriate cut scores and the implications for international students were also highlighted.

In examining the relationship between ELP and academic achievement, the review delved into predictive validity studies that assess the extent to which performance on standardized ELP tests predicts academic outcomes like GPA. Studies on both IELTS and TOEFL were analyzed, revealing that while studies typically agree that there is a statistically significant correlation between ELP test scores and academic achievement, the relationship is generally weak. Meta-analyses on the predictive validity of both IELTS and TOEFL confirmed these findings, showing small effect sizes in relation to academic achievement. The review also noted that studies often lacked the inclusion of moderator variables such as age, gender, discipline of study, and country of origin, which is a gap this dissertation seeks to contribute to.

Subsequently, this chapter highlighted a significant gap in the literature concerning the comparison of different English proficiency pathways and their relationship with academic achievement. While substantial research exists on the predictive validity of individual ELP tests like IELTS and TOEFL, there is a notable lack of studies comparing outcomes for students who follow different pathways, such as standardized testing versus completion of English for Academic Purposes (EAP) programs. Due to the limited number of studies and various methodological limitations—such as small sample sizes, single-institution focus, and lack of control for important moderator variables—it is challenging to arrive at a clear consensus regarding the impact of different English proficiency pathways on academic achievement. While some studies suggested that standardized test pathways (IELTS and TOEFL) may be associated with better academic achievement (e.g., Tweedie & Chu, 2019; Schoepp, 2018), others observed

that the EAP pathway has a better association with achievement (e.g., Johnson & Tweedie, 2021). Therefore, the existing literature does not provide enough evidence to conclusively determine which English proficiency pathway generally results in better academic outcomes. This underscores the need for this dissertation which provides more comprehensive research that compares different pathways while controlling for key variables—a notable gap in the literature.

Additionally, much of the extant research exploring ELP pathways and student outcomes is constrained by single-institution datasets. Such designs limit the generalizability and robustness of conclusions across diverse educational environments. Consequently, a multi-site comparative approach—such as the one adopted in this dissertation—is crucial for providing greater validity and reliability of the findings and, therefore, contributes to addressing this significant gap in the literature.

Lastly, recognizing the limitations in the existing literature, the chapter then explored other factors that can influence academic achievement and demonstrated the necessity of controlling for these variables when examining the relationship between ELP and academic outcomes. Factors such as gender, age, country of origin, and academic major were identified as significant predictors of GPA. Research has consistently shown gender differences in academic achievement, with women often achieving higher GPAs than men. Age was also found to be a meaningful predictor, with older students generally outperforming younger ones. Country of origin can impact academic achievement due to cultural, linguistic, and regional differences, while academic major/discipline of study has been shown to account for significant variance in GPA and graduation rates. Including these factors as control variables is essential for a more accurate and nuanced analysis of how English proficiency pathways impact academic achievement.

Overall, this review has established that this dissertation seeks to contribute to the academic discussion found in the literature in three key ways:

(1) it tests the claimed equivalence of IELTS and TOEFL and contrasts both with a course-based EAP pathway;

(2) it does so with cohorts drawn from two different Canadian universities, allowing for the first cross-institutional analysis of pathway effects; and

(3) it incorporates key moderators—gender, age, discipline, and country of origin—so that any pathway differences (or similarities) can be interpreted with greater confidence. In doing so, the study responds directly to Ihlenfeldt and Rios’s (2023) call for broader, multi-context evidence and fills a critical gap left by prior single-institution, single-test research. The next chapter will outline the methods used to conduct this analysis, including data collection, data cleaning, and the selection of appropriate statistical techniques.

Chapter 3: Method

This chapter provides a detailed overview of the methodology and data sources used for this dissertation, which investigates the relationship between English language proficiency (ELP) pathways and academic achievement among international students at two large Canadian universities: University A and University B.

The chapter begins by examining the institutional contexts of University A and University B, outlining their distinct admission policies, ELP requirements, and English for Academic Purposes (EAP) programs. Following this, the chapter describes the process of data collection, detailing the data received from both universities, and outlining the key variables. Subsequently, a detailed overview of the data cleaning process is included. Next, the creation of the variables for this dissertation are discussed. Descriptive statistics and an analysis of emerged patterns for each variable are presented. In addition, the chapter presents the regression equations used to explore the study's research questions. These equations provide the framework for examining the relationship between ELP pathways and academic achievement, while controlling relevant moderator variables. Finally, the chapter presents statistical assumption testing, which is key to ensuring the validity of the regression analyses used to explore the research questions.

Institutional Context

Data for this study was obtained from two large universities in Canada. As such, this dissertation employs a multi-site comparative research design by examining data from two distinct Canadian institutions, University A and University B. This methodological decision explicitly addresses limitations prevalent in earlier single-site studies, enhancing both the validity and reliability of the study. By comparing outcomes across multiple institutional contexts, this

dissertation provides a stronger basis for generalizable findings to broader populations and educational settings, thus increasing the robustness and applicability of the results.

The first institution, anonymized as University A for the remainder of this study, is a large institution in southern Ontario. The second, University B, is a large institution in western Canada. From the outset, this dissertation was designed as a multi-institutional study to test whether pathway achievement patterns replicate across diverse Canadian university contexts. This was done to move beyond the single-site limitation noted in earlier research (Schoepp 2018; Tweedie & Chu 2019). As such, four large, English-medium universities in Canada were approached - representing different provinces and institutional types (e.g., research, teaching, comprehensive). Formal data-sharing requests—backed by Research Ethics Board (REB) approval—were submitted all four between June 2020 and December 2020. Ethical approval was approved as follows:

University A – approval # STU 2020-093 on September 09, 2020

University B – approval # Pro00103209 on October 30, 2020

University C - approval # 20-11-068 on December 24, 2020

University D – approval H20-03168 on October 13, 2020

However, after contacting each universities data or admission office, only University A (Comprehensive/Teaching - Ontario) and University B (Research – Western Canada) provided full undergraduate datasets that included admissions pathway indicators, test scores, and first-year academic records.

University C (Research – Central Canada) did supply a large longitudinal data file; however, after inspection, lacked a reliable indicator distinguishing whether English-language proficiency was met via standardized testing, EAP completion, or exemption. This institution was contacted for insight into this lack of data and research officers were unable to locate or construct the appropriate variable from the institutional data sets - citing a lack of data collection on metrics related to how students met ELP admission requirements. Because pathway identification could not be constructed from the raw fields, this institutional site could not be entered into the present analyses without drastically expanding the scope of the research questions and/or jeopardizing validity. This dataset, however, offers a fertile context for future work: it displays numerous other pathway categories (e.g., International Baccalaureate applicant, Domestic university/college transfer applicant, USA university/college student applicant, High school applicant, etc.) of international students. Future research could utilize these data to examine the relationship between admission pathways and first-year academic achievement within the international student population, thereby filling a contributing to a gap in the higher-education literature, where systematic, longitudinal, multi-pathway comparisons international student pathways remain rare (Davis & Pizarro Milian, 2020; Diamond & O'Brien-Malone, 2018; Ruegg et al., 2020; Tan, 2021). Lastly, the registrar at University D (Research-Western Canada) declined to release student-level data citing staff capacity concerns.

University A and University B—provide a solid basis for comparing pathway effects in different institutional settings. The following sections describe each university's context, covering admission policies, ESL programs, and the English-language-proficiency tests they accept.

University A

As one of the largest institutions by enrollment in Canada and by its location in the population-dense area of southern Ontario, University A has a significant international student body. According to publicly available information at the time of writing, 28% of the total student body (undergraduate and graduate) were identified as international students. Other information indicates that international undergraduate students are ~15% of the total first year undergraduate population. According to The Academic Ranking of World Universities as presented by the Shanghai Ranking, University A is ranked within the 401-500 range. When exploring by subject, University A has one subject area ranking in the top 100 with all other subjects ranging in rank between 100 – 500.

As with all English-medium universities in Canada, University A requires that if a student's first language is not English, they must show an acceptable level of English proficiency by achieving a specific score on an English language proficiency test or receiving conditional or exemptional enrolment. For the English proficiency test, University A sets a minimum score requirement for each accepted language test and these requirements may vary depending on the program and level of study. Although University A will accept six different English language tests for admission purposes, this study has data from the two most prominent tests: the IELTS and the TOEFL. Therefore, the minimum score requirements of these tests will be presented while the other tests, which this study did not gather data for, will be excluded.

For admission via IELTS Academic (the IELTS General is not accepted), University A normally requires an overall band score of 6.5. However, some faculties set higher thresholds—for instance, Education and Business both require an overall 7.0. For the TOEFL iBT, the

university's baseline is 88, but Education and Business raise the bar to 94. All test scores are valid for two years.

University applicants who cannot present a qualifying test score can gain entry by completing University A's English for Academic Purposes (EAP) program. The program is organised into nine levels:

- **Levels 1–5** build foundational reading, writing, listening and speaking skills.
- After Level 5, students may enter a bridging option that lets them take **one** for-credit course while finishing their language studies.
- **Level 7** is an eight-week intensive course designed to bring learners to roughly IELTS 6.0.
- **Levels 8 and 9**, also delivered as a single eight-week intensive course, complete the program. Successful completion is deemed equivalent to an IELTS overall score of 6.5, after which students may begin full-time degree study.

University B

University B is a large research-intensive institution in western Canada. As a member of the U-15 (i.e., the association of leading research universities in Canada), University B has an international reputation for both teaching and research. As such, University B boasts an international graduate student enrollment of 39%; but only a 14% international undergraduate enrollment. The percentage of first year international undergraduate students could not be located in public data.

According to The Academic Ranking of World Universities as presented by the Shanghai Ranking, University B is ranked in the top 100 institutions globally. When exploring by subject,

University B has numerous subject areas that rank in the top 100 as well, with some subject areas ranking as high as the top 30 and top 20 in the world.

For university applicants whose first language is not English, University B accepts five standardized tests, but only IELTS and TOEFL appear in this dataset. The minimums are an IELTS Academic overall score of 6.5 with no band below 6.0 and a TOEFL iBT overall score of 90 with at least 21 in reading, writing, listening, and speaking. Applicants who completed all prior schooling in English, or who graduated from certain approved overseas institutions, may be exempt, but otherwise one of these benchmarks must be met before admission

Students who do not or cannot meet these scores can qualify through the university's English-for-Academic-Purposes (EAP) curriculum. The program begins with three Academic Foundations levels, followed by ESL 135, ESL 140, and ESL 145.

- **Academic Foundations 1 → 3** – entry levels that build core academic-language skills.
- **ESL 135** – intermediate academic language.
- **ESL 140** – upper-intermediate; focuses on academic reading & writing.
- **ESL 145** – advanced; integrates research, presentation and disciplinary tasks.

Successful completion of both ESL 140 and ESL 145 is deemed equivalent to meeting the IELTS-6.5 / TOEFL-90 standard. All EAP courses are graded on a pass/fail basis, and a bridging option permits conditionally admitted students to start degree work while finishing ESL 140/145.

Overview of Dataset

Data for this study were received from University A's Research Office and from University B's Office of Registrar. Data were received in the form of large excel files. University A's data were accessible first due to the author's doctoral supervisor's work as part of a separate research project. This larger data set contained many variables, some of which were relevant to this study, while others were not. Based on the literature review, key variables were identified in University A's data set that were needed to answer the research questions of this study. As a result, University B's Office of the Registrar was contacted after ethics approval for numerous variables to match relevant variables found in the University A data set. University B provided a substantial amount of data; however, they were unable to provide all variables requested (e.g., students' first language, EAP course grades, subject of study). It is possible that some of these data were not collected (e.g., first language) but despite repeated inquiry, no concrete reasons were provided as to why some data were inaccessible. The total student counts in the datasets prior to any cleaning were as follows:

Table 3.1: Raw data counts

Institution	Sample count
University A	$n = 24,083$
University B	$n = 4,327$

As a result of data collection noted above, two large data sets were provided by the two institutions with numerous variables – each of which will be described in the following section but briefly outlined here. A summary table is also provided below (Table 3.2). In terms of the main dependent variable of this study (i.e., academic achievement operationalized by GPA),

University A provided student cumulative GPA for academic years from 2009 to 2019, while University B provided 1st year and 2nd year student cumulative GPA from 2017 to 2021. As requested, each institution also provided the number of credits attempted and the number of credits earned for each year of data. As the key independent variable for this dissertation is English proficiency pathway, both institutions were requested to provide overall test scores for both the IELTS and the TOEFL iBT. While sub-test scores (i.e., scores on sections such as reading/writing) are not necessary to answer the research questions of this study, these data were requested from both institutions as a precaution and for future research purposes. University A provided sub-test scores, University B did not. Also, despite the request, no variable was provided that noted if overall scores were sufficient to meet university admission requirements, thus one was created. This was done by comparing university admission cut scores to student overall scores on the IELTS or TOEFL and inferring that students met ELP admission requirements by obtaining sufficient overall scores. This created variable will be discussed shortly in the following section. Also, key for the pathway variable, both institutions were requested to provide indication of any course-based ELP route. University A provided a binary Y/N variable indicating that students were admitted through the EAP program, while University B provided two separate variables that indicated student enrollment in two out of three EAP courses noted in the context section above (i.e., Y/N). These two courses, dubbed the mid-level and high-level EAP courses, were used to create the EAP pathway variable discussed shortly. No explanation was provided as to why data for the lowest level EAP course were not provided. A description on how these EAP variables were interpreted and coded into a pathway variable is also provided shortly. Additionally, each university provided data about each student's age, gender, and country of citizenship— with slight differences in coding that will be described below.

Lastly, University A provided data about each student's study major and discipline of study, while University B was able to provide discipline of study but did not provide study major.

Table 3.2: Variable overview

Variables	University A	University B
Cumulative GPA for each academic year	✓	✓
IELTS test scores	✓	✓
IELTS sub-section scores	✓	-
TOEFL test scores	✓	✓
TOEFL sub-section scores	✓	-
EAP program enrollment	✓	-
EAP course enrollment (mid level)	-	✓
EAP course enrollment (high level)	-	✓
The number of credits attempted	✓	✓
The number of credits earned	✓	✓
Study major	✓	-
Discipline of study	✓	✓
Age	✓	✓
Gender	✓	✓
Country of Citizenship	✓	✓

Description of Variables

This section provides an overview of the variables obtained for this dissertation. In addition to serving as the predictor, control, and dependent variables of the study, these variables were used for data cleaning. Data cleaning removed a considerable number of cases from the

original datasets due to a variety of issues; therefore, this section will present the raw data before any cleaning took place. The final sample will be presented in a subsequent section shortly.

English Proficiency Pathway

Besides the dependent variable(s), the most important variable in this dissertation is the primary predictor variable – English proficiency pathway. However, this variable had to be constructed from data which indicated that students had either submitted a standardized ELP test score, or from variables which indicated that students were admitted through an EAP course or program. Through the data collection process, it was possible to collect data on IELTS and TOEFL test scores from both University A and B. University A's raw data contained 12,431 students with an IELTS reading, writing, listening, speaking, and overall score. The overall scores ranged from 0 to 9 with a mean of 6.33 (out of 9). These data were dated from 2009 to 2020. University B's raw data contained 2,822 students with an IELTS overall test score. The scores ranged from 3.5 to 8.5 with an overall score mean of 6.39. These data were dated from 2017 to 2020. It was not possible to collect sub-test scores. For the TOEFL iBT data, University A provided *data for* 4,840 students with reading, writing, listening, speaking, and overall TOEFL scores. The overall scores ranged from 23 to 120 with a mean score of 95 (out of 120). These data were also dated from 2009 to 2020. University B's TOEFL data contained 270 students with an overall TOEFL score ranging from 43 to 120. The mean score was also 95. These data were also dated from 2017 to 2020, and it was also not possible to collect sub-test scores.

For indication of EAP, University A provided a separate excel file with 1335 cases which were marked to indicate admission through the institutions EAP program from 2009 to 2020. Conversely, University B used a binary Y/N variable to indicate that 1098 cases were admitted through the mid-level EAP course and 32 were admitted through the high-level EAP course. Due

to the volume of available data, it was possible to construct a new variable that split students into categories not only based on their pathway such as having sufficient IELTS or TOEFL scores required for admission, but also to group students into distinct categories based on their ELP test scores. In other words, a new variable was created by grouping students who were admitted to university with ELP test scores below the admission standards (e.g. <6.5 on IELTS), those who met admission standards (e.g., 6.5 to 7.5 on IELTS), and those who were admitted with scores that significantly outperformed admission standards (e.g. >7.5 on IELTS) and to compare these groups to the EAP pathway students. Creating this variable was a key component of the data cleaning process as unusable cases were identified throughout its creation. Therefore, how this variable was created will be discussed in more detail in the data cleaning section below with final counts presented in the final sample section.

Cumulative grade point average

As discussed in the literature review, academic achievement for this dissertation is operationalized by cumulative grade point average (CGPA). This dissertation specifically used CGPA as the outcome measure to assess academic achievement with the rationale that CGPA provides a more comprehensive proxy of student success over time. While first term GPA may capture immediate adjustments of international students to a new academic environment, CGPA, conversely, encompasses a broader range of academic experiences and captures students' performance across multiple semesters and courses. Furthermore, it was theorized that as students gain more experience in an English study environment, language performance would become less of a hinderance as their skill set improved over time. Testing this assumption is a possible avenue for future research. Additionally, the operationalization of academic achievement as CGPA follows norms found in the literature in both predictive validity research

(see: Al-Malki, 2014; Gagen, 2019; Ihlenfeldt & Rios, 2022) and the limited amount of research that explores questions similar to this dissertation (see: Ruegg et al., 2020; Johnson & Tweedie, 2021; Tweedie & Chu, 2019). Lastly, data limitations in terms of access to information were also a factor when considering additional operationalizations of academic achievement.

As this study examines the relationship between English proficiency pathways and academic achievement at the end of the first year of study, CGPA was the key dependent variable and, therefore, obtained from both University A and B. University A data contained 14,381 cases of first year CGPA on a 0-9.0 GPA scale. The mean was found to be 6.9 with a SD = 2.36. The count of CGPA at the end of the first year was significantly lower than the total data count noted above ($n = 24,083$) – as there were significant blanks found. This will be discussed subsequently in the data cleaning section below. University B data contained 4,327 cases with CGPA data on a 0-4.0 scale. The mean was 2.0 with a SD = 1.40. This data also had significant data gaps – discussed below as well.

Credits earned

It was possible to collect credits completed/earned data from both University A and B. Both institutions calculated credits similarly with courses at both institutions usually equaling three credits per course per semester. Therefore, data for these variables tended to range in intervals of three from 3 credits (one course completed) to 30 credits (10 courses) per academic year. There were oddities found in these variables at both institutions with outlier cases such as 4.5, 6.1, 9.1, 17, etc. With no explanation for these oddities, these cases were removed from the final sample.

Discipline of study

As discussed in the literature review, discipline of study has been shown to be a meaningful variable to include in research that explores the relationship between ELP and academic achievement (e.g., Feast, 2002; Harsch et al., 2017; Light et al., 1987; Schoepp & Garinger, 2016). For this dissertation, each institution provided data that included discipline of study; however, the sheer number of categories made interpretation of variables unpractical in analyses. Therefore, a new collapsed variable was created to include in statistical analyses. This new variable collapsed disciplines into the broad areas of Social Science and Arts, Business, and STEM. This was deemed the most practical coding considering the conventions found in the literature and the disciplines presented below. As a result, *Table 3.3* presents a summary of the disciplines retrieved, the counts before data cleaning, an indicator if this information was provided by both or one institution, and which broad category each discipline was coded into or if the data were removed.

Table 3.3: Discipline data received

Discipline	University A	Count	University B	Count	Recode
Agri, Life & Enviro Sciences	-	-	✓	259	STEM
Arts	✓	12,050	✓	1347	SS and Arts
Art, Media, Design	✓	658	-	-	SS and Arts
Business	✓	1853	✓	120	Business
Education	✓	162	✓	8	SS and Arts
Engineering	✓	1422	✓	1030	STEM
Environmental Studies	✓	214	-	-	SS and Arts
Enviro and Urban change	✓	12	-	-	SS and Arts
Graduate Studies	✓	2870	-	-	Removed
Health	✓	1976			STEM
Kinesiology, Sport, & Rec	-	-	✓	45	STEM
Law	✓	47	✓	6	SS and Arts
Native Studies	-	-	✓	3	SS and Arts
Nursing	-	-	✓	16	STEM
Open Studies	-	-	✓	82	Removed
Other	✓	486	✓	183	Removed
Rehabilitation Medicine	-	-	✓	7	STEM
Science	✓	2333	✓	1210	STEM

As noted above, three broader subject areas were created which resulted in the creation of the subject variable and subsequent counts (*Table 3.4*). The graduate studies category from University A was removed due to lack of fit with the research questions and ‘Other’ was removed due to the lack of interpretive significance; however, this will be discussed in more detail in the data cleaning section below.

Table 3.4: Discipline data recoded

	University A	University B
Social Science and Arts	13,143	1,364
Business	1,853	129
STEM	5,731	2,567
Removed	3,356	267
Total	24,083	4,327

Age

University A age data ranged from 15 to 65 with a mean age of 21 while University B age data ranged from 16 to -37 also with a mean age of 21. As a continuous variable, no additional coding was required for age.

Gender

For University A, gender was presented with three categories: male, female, and unknown. Female students accounted for 51.4% of the total sample or 11661 cases, where males accounted for 48% of the sample or 10960. Unknown accounted for 0.6% of the total sample or 147. This variable was coded to have men as the reference category and due to the small sample size and to match University B, the unknown category was removed. Gender at University B was

presented as a binary variable composing of female and male. Female students accounted for 45% of the total sample or 1927 cases, whereas males accounted for 55% of the sample or 2,400. This variable was coded to have men as the reference category to match University A.

Country of Citizenship

Both University A and B provided country of citizenship data which created numerous coding options. University A provided student data represented by 162 countries, while University B provided 35 countries worth of data. The country most represented at University A was China at 47% of the total sample, followed by India at 6.5% and Iran at 6.2%. At University B, China was also the most represented country at 66% of the total sample, followed by India at 12% and Bangladesh at 4%. Due to the small sample sizes of most of the countries, and for modelling and interpretive purposes, each institution's country of origin data was recoded into two groups where China was the reference category (0), and all other countries were coded together as 1.

The above section describes the predictors included in this dissertation; however, significant data cleaning was necessary as various issues were found in the data. Therefore, the next section presents how data cleaning was completed to create the final sample for analyses.

Data Cleaning

Data cleaning began for both institutions with analyses of the predictor variables described above to note any discrepancies. The first variables analyzed were indicators of English proficiency pathway as this was the key predictor for this dissertation and required construction from the raw data.

Exclusion based on pathway

Both University A and B had a significant number of cases which had no clear indication of their admission pathway. For example, some students had no indication that they were ever enrolled in an EAP program, but they also had no recorded ELP test score. This made pathway identification for these cases impossible. At University A, 6,660 students were found to not have any indication of pathway, while at University B 659 students had no such indication. These cases were removed as it was logically impossible to code them into the final data set. The remaining cases were examined based on indicators for the IELTS, the TOEFL, or the EAP pathways. Additionally, due to the availability of ELP test score data at both institutions, it was possible to account for performance within pathways. In other words, as discussed above, data was provided for overall test scores, not simply a yes/no indicator that students met ELP requirements through completing one of the standardized ELP tests. Therefore, it was possible to group students into ELP ranges based on their performance on either IELTS or TOEFL instead of simply grouping all students together into a singular category based on pathway. Scholars with robust data such as that obtained for this dissertation have also categorized IELTS and TOEFL students based on their performance on ELP tests and compared the academic achievement of the different groups (e.g., Cho & Bridgeman, 2012; Schoepp & Garinger, 2016; Woodrow, 2006). However, the novel approach of this dissertation is the inclusion of students who were admitted to both universities with ELP test scores below the admission standards cut scores and the inclusion and comparison of students with different ELP test score ranges to EAP students.

To create the pathway variable the following logic was used.

- If students had two scores on two ELP tests, and one score met the cut score and the other did not, the score that met admission requirements was assumed to be the student's

pathway (e.g., a student with a 5.5 on IELTS and a 90 on TOEFL was coded into the TOEFL pathway).

- If students had an EAP indicator and an ELP test score but the ELP test score was below the cut score, they were coded as admitted through the EAP pathway.
- If students had two scores, one on IELTS and one on TOEFL and both scores were above the admission cut score, it was impossible to discern which test was used to meet admission requirements. Thus, these cases were removed ($n = 105$ at University A and $n = 7$ at University B).
- If students' ELP test scores were significantly below the cut score (e.g., <5.5 on IELTS or <46 on TOEFL) and they had no indication of EAP, they were also removed ($n = 898$ at University A and $n = 33$ at University B).

The lower cutoff of 5.5 on IELTS and the equivalent score on TOEFL was applied based on a range of concordance with the highest degree of confidence for comparison between the IELTS and TOEFL overall scores ($r = .73$; ETS, 2010). This range ends at 5.5 on IELTS and the corresponding TOEFL score of 46.

As a result of the processes outlined above, the remaining students were coded into one of seven categories – six which corresponded to a specific ELP test score range and one for EAP if students were admitted through this pathway. The IELTS admission requirement at both institutions was the same (6.5), thus the three categories and their corresponding ranges applied to both universities. The following categories were created.

- Those students with an IELTS score ranging from 5.5 to 6.0, indicating students admitted with IELTS scores below the admission requirements.
- Those students with an IELTS score ranging from 6.5 to 7.5, indicating a middle range of IELTS scores and the meeting of admission requirements.
- And those students with an IELTS score ranging from 7.5 to 9.0 indicating students admitted with a high IELTS score and the meeting of admission requirements.

For TOEFL, University A had a TOEFL admission requirement of 88, while University B had a requirement of 90. Therefore, the following ranges slightly differed for each university. The following categories were created.

- Those students with a TOEFL score ranging from 46 to 87 (University A) and 46 to 89 (University B), indicating students admitted with TOEFL scores below the admission requirements.
- Those students with a TOEFL score ranging from 88 to 109 (University A) and 90 to 109 (University B), indicating a middle range of TOEFL scores and the meeting of admission requirements.
- and those with a TOEFL score ranging from 110 to 120 representing a high TOEFL score.

As a result of this process, 7 total categories were created including the reference category of those in the EAP pathway. This key predictor variable with the final counts of each category will be displayed in the final sample section shortly. However, additional cleaning was required based on other predictor variables, as well as on the dependent variable (CGPA).

Exclusion based on other predictor variables

Additional cases were removed if they could not be coded and interpreted on the other predictor variables included in this dissertation. For discipline of study, most students had a declared subject area, however, at University A, 807 cases were found to have an unknown discipline while at University B 140 such cases were located. These cases were removed. Regarding gender, University A had a small sample (<1%) in an unknown/undeclared gender category while University B data did not contain this category. Therefore, these cases were removed from the University A data to match samples. As noted above, country of origin was coded into a binary category where Chinese students formed the reference category, and all other students were coded as the comparison group. However, through the cleaning process some students ($n = 1,279$) at University A were identified as born in Canada. As this is a study of international students, these cases were removed. There was no such issue for University B data. As mentioned above, discipline of study was coded into three categories (social sciences and humanities, business, and STEM). Cases that could not be coded as such (due to missing data or lack of fit) were removed. Thus, 807 and 140 cases were removed from University A and University B, respectively.

As described above, University A data came from a larger data set provided by the author's doctoral supervisor. As a result, additional data cleaning was required for this sample as much more data was provided than was necessary for this dissertation. For example, both master's and PhD graduate students were included in the data set, which were not the focus of this dissertation ($n = 2,185$ and $n = 385$, respectively); therefore, these cases were removed. Furthermore, the data cleaning process located students who submitted ELP test scores after admission at University A. Since university policy is that scores must be submitted for admission

purposes, these cases were removed ($n = 57$). Also, cases were found of students who had submitted their ELP test scores more than two years prior to admission. University A policy dictates that English proficiency tests are valid for a period of two years; therefore, 715 cases were removed from the IELTS sample, and 837 cases were removed from the TOEFL sample, as these cases fell outside the two-year window. Furthermore, a variable for first language spoken was provided in the University A data set and 650 cases were identified with English listed as the first language spoken. As this is a study of students who speak English as an additional language, these cases were also removed from the final University A sample. This concluded data cleaning based on the predictors of this dissertation; however, additional data cleaning was required based on the dependent variable (CGPA).

Exclusion based on dependent variable

At University A, an additional 821 cases were located and removed due to blank CGPA data. Additionally, it was observed that there were 446 cases in the data set that had a recorded CGPA of zero. Additionally, all these cases had a credits earned value of zero, but a ranging credits attempted value of 3 to 21. No demographic or pathway similarities were found for these cases. With no explanation for these cases, they were removed.

At University B 874 cases were located that had a recorded CGPA of zero but also had earned university credits. None of these cases were identified as EAP. It is possible that these students were enrolled in courses that were pass/fail only and that offered credits for completion. This was theorized by the data office when this anomaly was brought to their attention. However, they could not provide a concrete explanation for this group. A small number of cases ($n = 7$) were also found at University A. Therefore, as to not skew the data with a significant number of extremely low GPA data, these cases were removed. Additionally, as with University A, 168

cases were located that had a recorded CGPA of zero with a credits earned value of zero. Also, no demographic or pathway similarities were found for these cases. Therefore, these cases were also removed.

As a result of data cleaning, the final sample for University A was 9,673 students and 2,480 students for University B.

Final Sample

This section provides the final counts of all relevant variables noted above, post data cleaning. Counts at both universities are presented together in one table by predictor for ease of comparison. Dependent variable tables will be presented in the subsequent findings chapter.

English proficiency pathways and ELP levels

As described above, for both institutions, pathway was coded into seven categories with EAP as the reference category. The final count of each pathway at both institutions is presented below (*Table 3.5*). As stated above, various score ranges for each of the ELP tests were used to group students in terms of their ELP level. Students with *low* ELP were those admitted to the universities with test scores below admission requirements (IELTS 5.5 to 6.0 at both institutions, and TOEFL 46 to 87 at University A and TOEFL 46-89 at University B). Students with mid ELP were those admitted to the universities with test scores that met ELP admission requirements or were slightly above the admission cut score (IELTS 6.5 to 7.5 at both institutions and TOEFL 88-109 at University A and 90-109 at University B). Lastly, Students with high ELP were those admitted to the universities with test scores ranging from 8.0-9.0 on the IELTS or 110-120 on the TOEFL.

Table 3.5: Final pathway count by ELP level, pathway, and university

Pathway Group	University A:	University A:	University B:	University B:
	Frequency	Percent	Frequency	Percent
EAP	908	9.39%	795	32.06%
IELTS (low)	2,814	29.09%	197	7.94%
TOEFL (low)	582	6.02%	16	0.65%
IELTS (mid)	3,627	37.50%	1,217	49.07%
TOEFL (mid)	1,323	13.68%	120	4.84%
IELTS (high)	301	3.11%	112	4.52%
TOEFL (high)	118	1.22%	23	0.93%
Total	9,673	100.00%	2,480	100.00%

This table highlights several key patterns in the distribution of students admitted through different English proficiency pathways across the two universities. At University A, the majority of students (37.50%) were admitted with mid-range English proficiency on the IELTS test. A substantial proportion of students (29.09%) were also admitted with low IELTS scores, indicating that this university has a relatively high number of students admitted below the IELTS cut score. TOEFL mid-range scores accounted for 13.68% of the sample, with lower TOEFL scores making up 6.02% of admissions. High IELTS and TOEFL scores represented the smallest groups, with 3.11% and 1.22% of students, respectively. EAP admittance only accounted for

9.39% of students at University A, included in the final sample. Overall, University A shows a particularly large proportion of students with low to mid ELP levels, especially on the IELTS.

University B has a significantly different profile. The majority of students (49.07%) were admitted with mid-level IELTS scores, similar to University A, but University B stands out for its high proportion of students (32.06%) admitted through the EAP pathway, which is much higher compared to University A (9.39%). Additionally, University B has a much smaller share of students with low IELTS scores (7.94%) compared to University A (29.09%), and similarly, it has very few students with low TOEFL scores (0.65%). Given the small sample sizes in certain categories for University B, particularly for students with low and high TOEFL scores, it is crucial to approach any findings of these groups with caution.

Like University A, high-level ELP admissions are rare, but high TOEFL admissions at University B are slightly lower than at University A (0.93% vs. 1.22%). This observation changes to the inverse when the samples were matched by enrollment year (i.e., including data for both samples from 2017-2020, only). Doing so reduces the University A sample to 4,658 students from 9,673, and the number of high TOEFL admissions to .24% ($n = 11$) of the new sample. Matching by enrollment year produces somewhat different results in terms of statistical significance and size of the coefficients of some pathways at University A; therefore, these findings will be presented as their own regression analysis in the findings chapter to provide a comparison by enrollment period.

Overall, while both universities admitted the majority of students with mid-range ELP above their admission cut scores University A has a larger proportion of students admitted with low proficiency, particularly on the IELTS, while University B admitted a significantly larger portion of students through EAP pathways. These differences suggest a potential divergence in

admission strategies or student demographics across the two institutions. Below, the final counts for the other variables included in the analyses at both institutions are presented and discussed.

Discipline of study

Table 3.6 highlights key patterns in the distribution of students across different subject groups at University A and University B. At University A, students are more evenly distributed across the three subjects. The largest proportion of students is found in Business, accounting for 36.28% of the total, followed closely by Social Science and Humanities at 34.72%. The STEM group represents 29.01% of the student sample.

University B has a significantly different distribution, with an overwhelming majority of students (73.10%) enrolled in STEM subjects. Social Science and Humanities represent only 24.52% of the student body, while Business is almost non-existent, comprising just 2.38% of the sample. This suggests that University B has a strong focus on STEM disciplines, while University A shows a more diversified academic profile. These patterns indicate that the academic focus and student distribution are markedly different between the two institutions.

Table 3.6: Subject group by university

Subject Group	University A:	University A:	University B:	University B:
	Frequency	Percent	Frequency	Percent
Social Science and Humanities	3,358	34.72%	608	24.52%
Business	3,509	36.28%	59	2.38%
STEM	2,806	29.01%	1,813	73.10%
Total	9,673	100.00%	2,480	100.00%

Country of origin

Table 3.7 provides insights into the distribution of students by country of citizenship at University A and University B, specifically comparing the proportion of students from China with those from all other countries.

At University A, a majority of students come from China, accounting for 54.34% of the total student body, while the remaining 45.10% are from other countries. University B has an even stronger concentration of students from China, with 65.71% of the student population. Students from other countries represent only 34.29%, meaning that University B is more heavily dominated by students from China compared to University A. These patterns suggest that University B has a higher dependency on Chinese students, whereas University A shows a more balanced distribution between students from China and other countries, though Chinese students still make up the majority. This difference highlights how the two universities vary in their levels

of international diversity, with University A having a slightly more globally diverse student body compared to the more concentrated Chinese population at University B.

Table 3.7: Country of citizenship by university

Origin Group	University A:	University A:	University B:	University B:
	Frequency	Percent	Frequency	Percent
China	5,256	54.34%	1,622	65.71%
All other countries	4,517	45.10%	858	34.29%
Total	9,673	100.00%	2,480	100.00%

Gender

Table 3.8 presents the distribution of students by gender at University A and University B. At University A, the student population is nearly evenly split between men (48.02%) and women (51.98%). In contrast, University B shows more of a gender imbalance where men represent a majority, comprising 58.47% of the student body, while women make up only 41.53%. The gender gap at University B could be influenced by factors such as the university's focus on certain academic disciplines (e.g., STEM) or its appeal to specific demographics (e.g., country of origin).

Table 3.8: Gender by university

Gender Group	University A:	University A:	University B:	University B:
	Frequency	Percent	Frequency	Percent
Men	4,645	48.02%	1,450	58.47%
Women	5,028	51.98%	1,030	41.53%
Total	9,673	100.00%	2,480	100.00%

Age

Age was treated as a continuous variable. Therefore, table 3.9 below summarizes the relevant information. At University A, the minimum age of students is 16, and the maximum age is 55, with a mean age of 19.9 years (Mdn = 21, SD= 3.1). In comparison, University B has a narrower age range, with a minimum age of 16 and a maximum age of 33 (Mdn = 19, SD= 1.5). The mean age of students at University B is slightly higher at 20.7 years compared to 19.9 at University A. Overall, University A appears to accommodate a broader range of student ages, while University B has a more concentrated age group with a slightly older average age.

Table 3.9: Age by university

Age	Minimum	Maximum	Mean	Total
University A	16	55	19.9	9,673
University B	16	33	20.7	2,480

Statistical Analyses

The goal of this study is to investigate the relationship between a key predictor variable—English proficiency pathway—and academic achievement, operationalized by cumulative GPA, while controlling for various moderator variables (e.g., academic discipline and demographic characteristics). To achieve this, multiple regression analyses were first conducted.

Regression analyses estimate the relationship between the main independent variable (pathway) and the dependent variable (CGPA) while including multiple control variables (such as subject, origin, and gender) (Agresti, 2018). Because GPA is a continuous variable (ranging, for example, from 0–9 at University A and 0–4 at University B) and the predictors include both categorical (e.g., pathway, subject) and continuous (e.g., age) variables, data met the basic requirements for multiple regression. In addition, key assumptions of regression—linearity, homoscedasticity, normality of residuals, and the absence of multicollinearity—were rigorously tested to ensure the validity of findings.

Building on the regression findings, Analysis of Covariance (ANCOVA) was employed to provide a more nuanced comparison of the different pathways. ANCOVA allows for the comparison of average outcomes (in this case, CGPA and Credits earned) across groups while controlling for other influential factors such as age and academic discipline (Agresti, 2018). This approach facilitated post hoc pairwise comparisons—for example, comparisons between IELTS High and TOEFL High, or between TOEFL Low and TOEFL Mid—that not only confirmed the differences observed in the regression models (e.g., differences between the EAP group and other pathways) but also extended the analysis to include additional comparisons not represented in the initial regression models.

To further isolate the effects of each predictor and to explore potential interactions, margin and interaction analyses were conducted. These analyses generated predicted outcomes (margins) for each combination of factors—such as pathway by gender or pathway by origin—while holding other variables constant. Post hoc pairwise comparisons of the margins provided a detailed analysis of where statistically significant differences were found. Margin analyses allowed the comparison of the predicted cumulative credits or CGPA for each additional predictor within each pathway, revealing, in some cases, exceptions to the overall trend. In this way, the margin and interaction analyses not only served as a valuable check on the regression and ANCOVA findings but also offered a broader perspective on how different English proficiency pathways interact with other predictors to moderate academic achievement.

Before presenting assumption testing, it is essential to outline the regression equations that were used to address the research questions of this study and to summarize the coding structure of the variables included in the equations. The regression equations reflect the research questions, capturing the relationship between English language proficiency pathways and academic achievement. Each model was structured to examine how these pathways, along with relevant control variables, predict academic achievement.

Multiple Regression Analysis

This dissertation asks:

How do international undergraduate students who are required to provide evidence of ELP and are admitted to university through different pathways compare in terms of their academic achievement at the end of their first year of study?

To answer this question, two different regression equations were created to explore the following:

1. the relationship between English language proficiency pathway and academic achievement operationalized by first year cumulative grade point average (CGPA):

$$\text{Equation: } CGPA = \beta_0 + \beta_1 Pathway_i + \beta_2 Discipline_i + \beta_3 Origin_i + \beta_4 Gender_i + \beta_5 Age_i + \epsilon_i$$

2. the relationship between English proficiency pathway and academic achievement operationalized by first year cumulative credits earned

$$\text{Equation: } CreditsEarned = \beta_0 + \beta_1 Pathway_i + \beta_2 Discipline_i + \beta_3 Origin_i + \beta_4 Gender_i + \beta_5 Age_i + \epsilon_i$$

All variables included in this dissertation are presented in table 3.12 below.

Table 3.10: Overview of modelled variables

Variable	Type	Range/Levels
CGPA (Dependent Variable 1)	Continuous	University A: 0–9, University B: 0–4
Credits Earned (Dependent Variable 2)	Continuous	University A: 0–33, University B: 0–21
Pathway (English Language Proficiency Pathway)	Categorical (7 levels)	Reference category: EAP pathway; Other levels: IELTS (low), TOEFL (low), IELTS (mid), TOEFL (mid), IELTS (high), TOEFL (high)
Discipline of Study	Categorical (3 levels)	Reference category: Social Science and Arts; Other levels: Business, STEM
Country of Origin	Categorical (2 levels)	Reference category: China; Other level: All other countries
Gender	Categorical (2 levels)	Reference category: men; Other level: women
Age	Continuous	University A: 16–55, University B: 16–33

The above section presents the regression equations used to answer the research questions of this dissertation and provides a summary of all relevant variables, post data cleaning. Once data cleaning was complete, methods were explored to analyze the final sample. Statistical assumption testing was conducted to discern if cleaned data met the requirements for various statistical analyses.

Method of Analysis and Assumption Testing

Linearity

The linearity assumption in multiple regression refers to the relationship between the predictors included in the model and the dependent variable being linear. However, it is important to note that categorical variables, such as the English proficiency pathways (e.g., IELTS and TOEFL scores), do not require a linear relationship with the dependent variable. Categorical variables represent distinct groups and are interpreted based on differences between group means rather than a continuous trend. As such, the linearity assumption is not applicable to these categorical predictors. In regression analysis software such as STATA, categorical variables are converted into dummy variables, which allows for clear differentiation between the categories without imposing the requirement of linearity. Given that this study uses categorical predictors such as English proficiency pathways and discipline of study, the linearity assumption does not pose any issues for these variables.

Age is the only continuous predictor in the models, and while its relationship with CGPA is not strictly linear, the clustering of CGPA around the most common age (i.e., 21) creates a skew that affects the linearity between the two variables. However, this pattern was expected and does not substantially violate the linearity assumption or meaningfully impact the results. To

further explore this assumption, age was transformed into a categorical variable ($<25 = 0$, $>25 = 1$) and the regression was re-run. The results remained consistent, indicating that the non-linearity of age does not significantly alter the model's outcomes or interpretation.

Homoscedasticity

Homoscedasticity refers to the state where the variance of residuals (or errors) is constant. This means that the spread or distribution of the residuals fit a particular pattern or shape without significant outliers or asymmetry. Homoscedasticity is important for multiple regression as it ensures the reliability and validity of results by ensuring that the standard errors of the coefficients are correct and thereby limiting incorrect p-values. Such incorrect p-values can lead to Type 1 (false positives) and Type 2 (false negatives) errors. If homoscedasticity is not present then the data is typically heteroscedastic, and estimates can be imprecise. There are different methods to test for the presence of homoscedasticity such as the Breusch-Pagan/Cook-Weisberg test or White's test. However, these tests are sensitive to sample sizes, as even small deviations from homoscedasticity can be detected in larger samples which can lead to a misdiagnosis of heteroscedasticity that has no practical significance (Long & Ervin, 2000). Both tests were run on both data sets and both produced results that would indicate heteroscedasticity. One solution, if heteroscedasticity is present, is to use robust standard errors (i.e., a measure of the variability of the coefficients in a regression model) which accounts for heteroscedasticity. This adjustment helps ensure that the standard errors are reliable despite the presence of heteroscedasticity. This method will be shown in the findings chapter in comparison with modelling that does not use robust standard errors.

Additionally, it is possible to interpret a scatterplot of the predicted values of a model (i.e., estimates of the dependent variable) versus residuals (i.e., the differences between the

observed values of the dependent variable and the predicted values). Both variables were created in STATA as outputs of the regression model. There are specific patterns of scatterplots that indicate heteroscedasticity (e.g., funnel or cone shaped), whereas homoscedasticity is typically indicated by a cloud of data points or a random distribution (Neter et al., 1996). Such scatterplots were created for both data sets (Figure 3.1 and 3.2). University A's scatterplot took the appearance of a cloud of points, whereas University B's was more randomly distributed. The appearance of lines found particularly in University A's plot indicates clustering around CGPA values ranging from 0-9. This is ideal, as the residuals closely resemble the distribution of CGPA data indicating that the models' predictions are close to the actual values of CGPA; therefore, they are a good fit. CGPA at university B is less clustered at specific CGPA values, hence the presence of strict lines is lessened. As both scatterplots reflect the distribution of CGPA well, heteroscedasticity is not a concern. However, as stated above, robust standard errors will be shown in a model below for comparative purposes.

Figure 3.1: University A Scatterplot (fitted values and residuals)

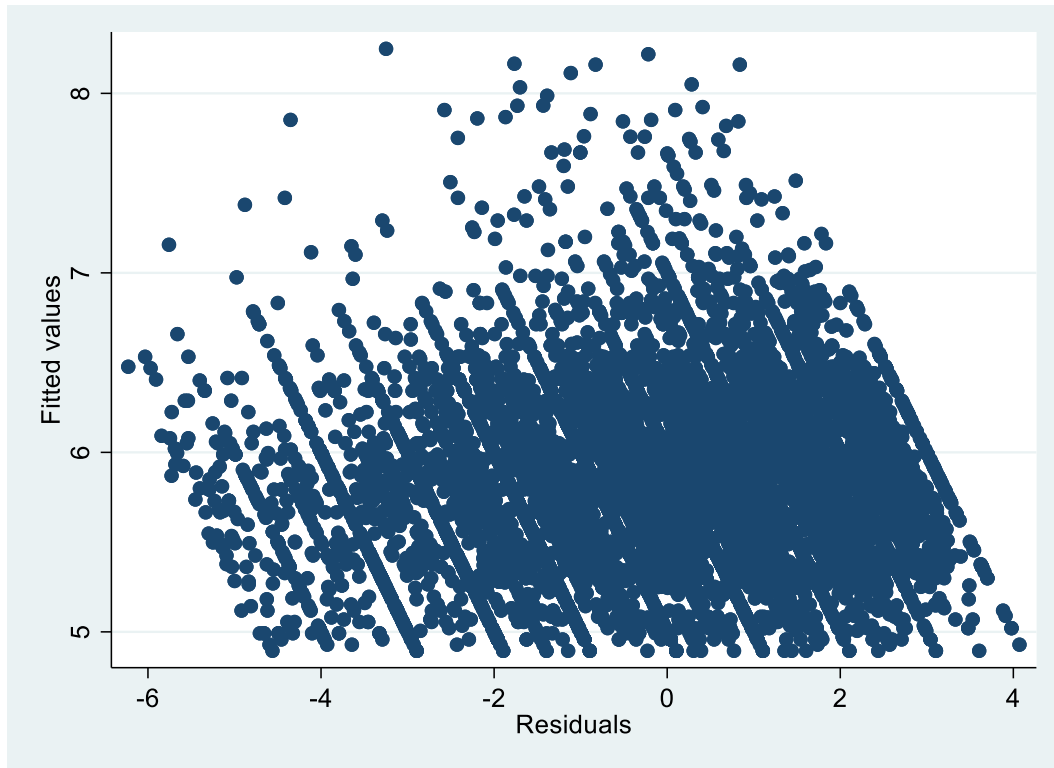
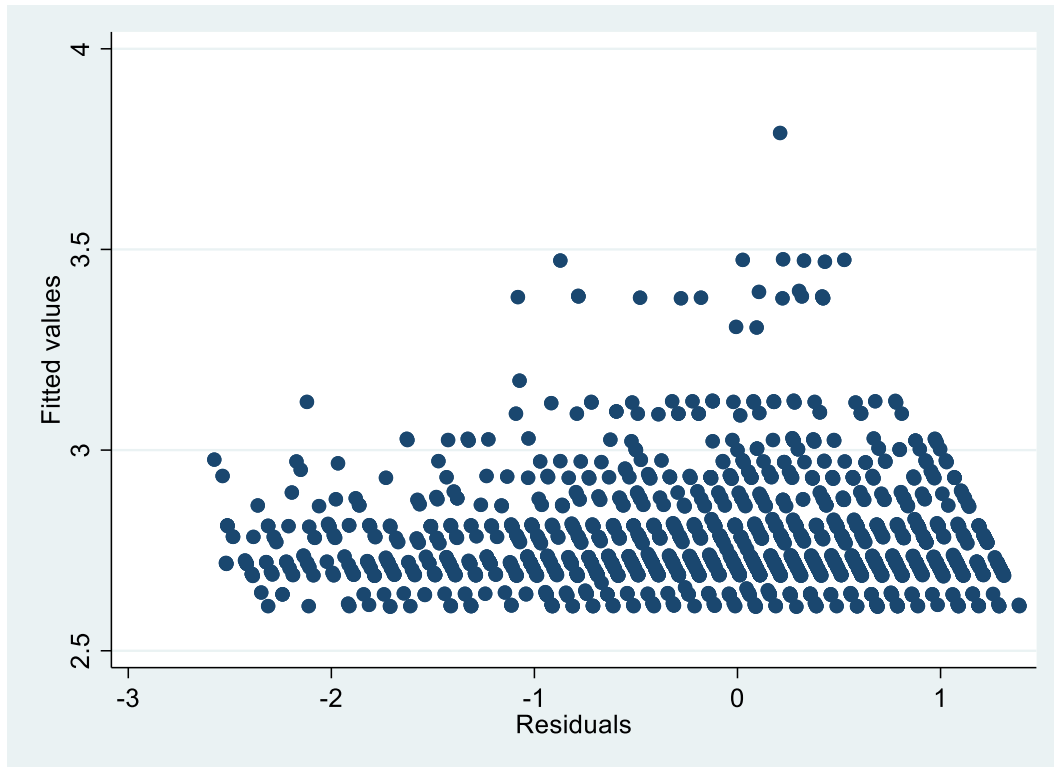


Figure 3.2: University B Scatterplot (fitted values and residuals)



Normality of Residuals

The normality assumption found in multiple regression assumes that the errors or residuals of the regression are normally distributed. Furthermore, there is no assumption or requirement that the predictor variable (i.e., CGPA) be normally distributed (Agresti, 2018). Such normality can simply be shown in a histogram of residuals and a Quintile- Quintile plot (QQ-plot). A QQ-plot demonstrates if a dataset follows a theoretical distribution (typically normal). Both visual tools were generated for each data set (see Figures 3.3 to 3.6).

The university A histogram (Figures 3.3) resembles a bell-shaped curve suggesting a normal distribution. The university B histogram (Figures 3.4) does have some residuals that

reach above the curve, and the residuals distribution appears to be slightly left-skewed, as there is a heavier tail on the left side. The skewness may suggest the presence of outliers.

Figure 3.3: University A Histogram (normality of residuals)

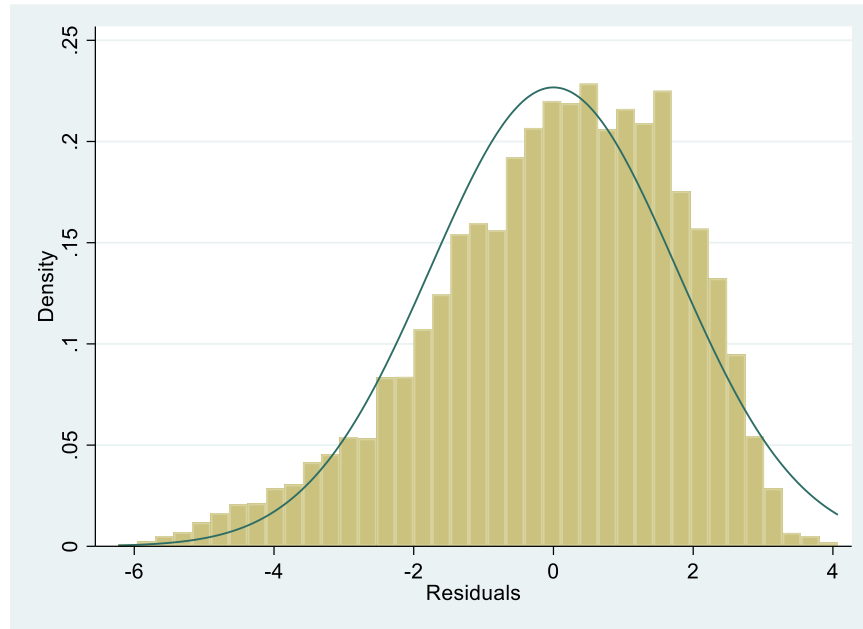
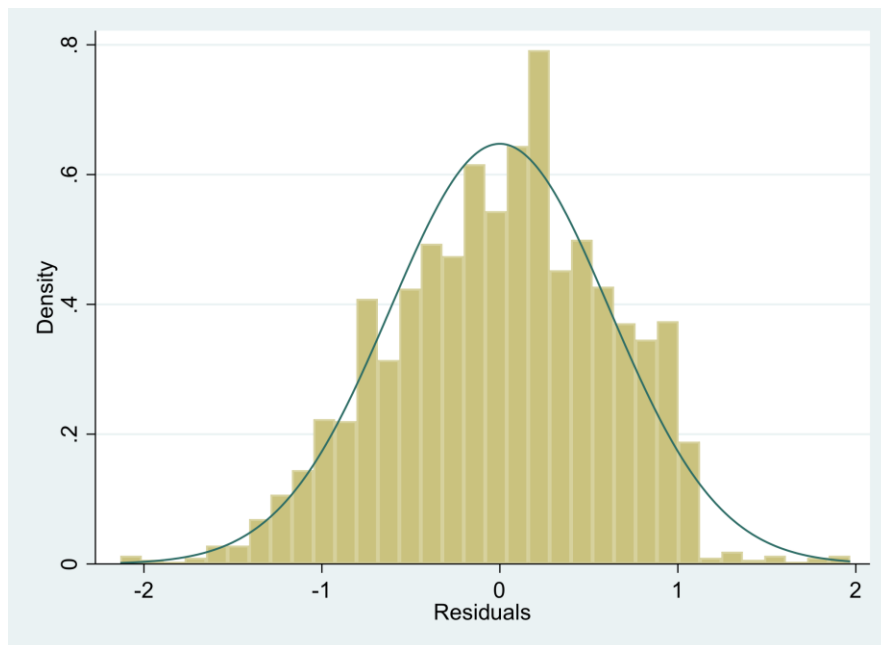


Figure 3.4: University B Histogram (normality of residuals)



To explore outlier influence and normality further, QQ-plots were generated for each data set. Both QQ-plots (Figure 3.5 and 3.6) deviate from the straight reference line indicating that the residuals have heavier tails than a normal distribution. In other words, the data are more spread out at extremes than what would be expected under normality. Considering the representation of very high CGPA values in both data sets (i.e., 4.0/9.0), this is not surprising. To account for the presence of outliers, robust regression was used (*see Appendix A*). This method down-weights the influence of outliers. Minimal differences between the OLS and robust regression results was found at the coefficient level. Key variables retain their significance across both methods, and coefficient estimates are closely aligned. This suggests that the OLS estimates are relatively robust to the non-normality and the presence of outliers in the dataset. In other words, this finding demonstrates that the differences between the OLS and robust regression results are minimal despite the presence of non-normal residuals due to heavy tails. As a result, despite the skewness noted in figures 3.3-3.6, negligible differences in coefficients and significance levels affirm that OLS regression provides reliable estimates for both datasets.

Figure 3.5: University A QQ-plot (normality of residuals)

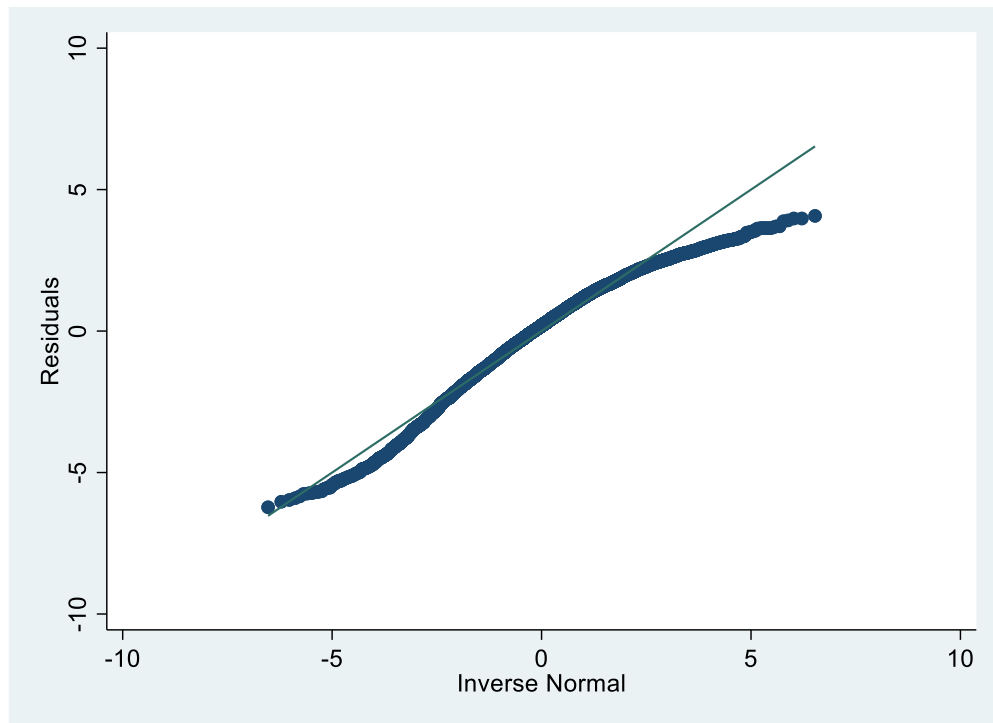
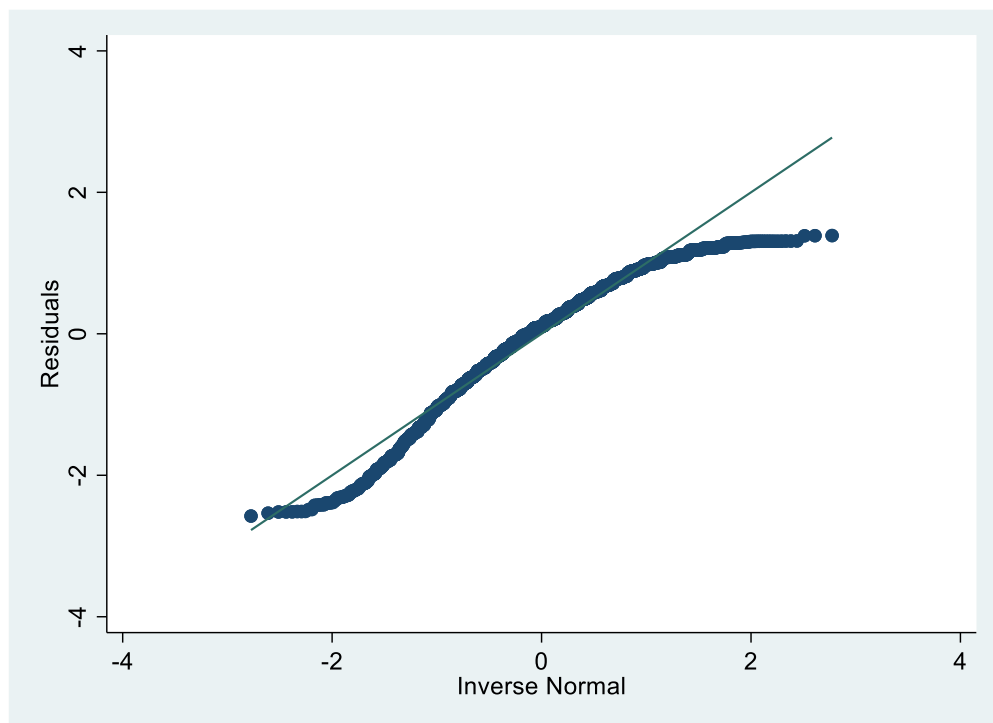


Figure 3.6: University B QQ-plot (normality of residuals)



Lack of Multicollinearity.

Multicollinearity can cause issues with unreliable coefficients within regression outputs. The concern is that as multicollinearity increases, the model coefficients become unstable. Furthermore, if two predictor variables have a high degree of multicollinearity, then one or more of the variables may be redundant to the model (Dormann et al., 2013). To examine multicollinearity, variance inflation factor (VIF) tests (Figure 3.7-3.8) were run on the full regression model which included all predictor variables for both data sets. Conventionally, a variable which has a VIF above 10 usually must be investigated further (Agresti, 2019). However, as the VIF tests demonstrate, there are no concerns for multicollinearity with either data set. Therefore, the assumption is met, as no evidence for multicollinearity was found.

Figure 3.7: University A variance inflation factor test

	VIF	1/VIF
1.IELTS (L)	2.93	.342
2.TOEFL (L)	1.60	.625
3. IELTS (M)	3.46	.29
4. TOEFL (M)	2.21	.439
5. IELTS (H)	1.39	.725
6. TOEFL (H)	1.13	.882
1.Business	1.40	.714
2.STEM	1.38	.725
Origin	1.50	.668
Gender	1.03	.963
Age	1.03	.973
Mean VIF	1.73	.

Figure 3.8: University B variance inflation factor test

	VIF	1/VIF
1.IELTS (L)	1.42	.68
2.TOEFL (L)	1.05	.941
3. IELTS (M)	1.82	.543
4. TOEFL (M)	1.18	.846
5. IELTS (H)	1.23	.813
6. TOEFL (H)	1.04	.958
1.Business	1.15	.873
2.STEM	1.21	.827
Origin	1.27	.785
Gender	1.08	.93
Age	1.51	.656
Mean VIF	1.27	.

Conclusion

This chapter has outlined the methods and data sources central to this dissertation, providing a detailed account of how the study was designed and the steps taken to ensure the rigor and reliability of the analysis. The institutional contexts of University A and University B were examined which allows for an understanding of the approaches and policies these universities have in terms of ELP admission requirements.

The chapter has also detailed the extensive data collection process, highlighting the challenges and considerations involved in gathering and cleaning data from two large institutions. The creation of the English proficiency pathway variable, the key predictor variable in this study, represents a novel approach to categorizing students based on their performance on ELP tests or their participation in EAP programs. This variable, along with other selected predictors, forms the foundation of the subsequent statistical analysis.

Through assumption testing, this chapter has established that the application of multiple regression techniques is apt for exploring the relationship between English proficiency pathways and academic achievement. The analysis that follows will build on this, offering insights into this relationship while controlling relevant moderator variables.

In conclusion, the methodological choices made in this chapter were aimed at ensuring that the findings of this dissertation are both valid and meaningful, providing valuable contributions to the literature on English language proficiency and academic achievement among international students in higher education. The next chapter will present the findings of the study followed by a discussion which offers a detailed examination of the implications of the results for both policy and practice.

Chapter 4: Findings

This chapter presents the findings of this study, addressing the two central research questions on the relationships between English Language Proficiency (ELP) pathways and academic achievement operationalized by cumulative grade point average (CGPA) and cumulative credits earned (CCE) among international undergraduates at two Canadian universities, referred to as University A and University B. The chapter is organized into two main sections; each section includes three main sub-sections:

1. **Descriptive Comparisons** of CGPA and credits earned by ELP pathway, discipline, country of origin, gender, and age at both universities.
2. **Regression Analyses** which present the results of models that examine the relationship between pathway and academic achievement whilst controlling for the effects of discipline of study, country of origin, gender, and age.
3. **Pairwise Comparisons and Interaction Effects** that offer deeper explorations into how results regarding ELP pathways and academic achievement vary dependent upon student characteristics.

Relationship between ELP pathways and CGPA

The following sections present the descriptive statistics for Research Question 1, which examines the relationship between English Language Proficiency (ELP) pathways and academic achievement as measured by first-year cumulative GPA (CGPA) while controlling for various moderators. Each section includes CGPA comparisons by pathway, discipline of study, country of citizenship, gender, and age across University A and University B.

Descriptive Statistics

CGPA by ELP Pathway

The mean CGPA for each ELP pathway group for both universities is presented in Table 4.1. At University A, CGPA is calculated on a 9-point scale. Among the 9,673 international undergraduate students, the mean CGPA was 5.83 (SD = 1.82), with a minimum of 0.14 and a maximum of 9.00. At University B, CGPA is calculated on a 4-point scale. Among the 2,480 international undergraduate students, the mean CGPA was 2.76 (SD = 0.84), with a minimum of 0.2 and a maximum of 4.00.

Table 4.1

Mean CGPA by English language proficiency pathway by university

Pathway Group	University A			University B		
	M	SD	N	M	SD	N
EAP	5.31	1.79	908	2.72	0.88	795
IELTS Low	5.67	1.68	2,814	2.78	0.90	197
IELTS Mid	6.11	1.80	3,627	2.75	0.80	1,217
IELTS High	6.72	1.78	301	2.96	0.76	112
TOEFL Low	5.72	1.75	582	2.76	0.60	16
TOEFL Mid	5.52	2.00	1,323	2.90	0.84	120
TOEFL High	6.29	1.87	118	3.42	0.51	23

At both University A and University B, students admitted through the EAP pathway had the lowest mean CGPA within their respective institutions (University A: $M = 5.31$ on a 9-point scale; University B: $M = 2.72$ on a 4-point scale), indicating that these students performed below the overall average at their universities. At each university, there was a trend of increasing mean CGPA corresponding with higher ELP test scores. For instance, at University A, the mean CGPA rose from 5.31 for the EAP group to 6.72 for the IELTS High group. Similarly, at University B, the mean CGPA increased from 2.72 for the EAP group to 3.42 for the TOEFL High group.

Additionally, students in the higher IELTS and TOEFL score groups tended to achieve higher mean GPAs compared to those in the lower score groups within the same university, with some exceptions. For example, at University A, students in the IELTS Mid group had a higher mean CGPA ($M = 6.11$) than those in the IELTS Low group ($M = 5.67$), while at University B, students in the TOEFL Mid group had a higher mean CGPA ($M = 2.90$) than those in the TOEFL Low group ($M = 2.76$). An exception to this trend was at University B, where students in the IELTS Mid group had a slightly lower mean CGPA ($M = 2.75$) than those in the IELTS Low group ($M = 2.78$). However, generally, students in the higher IELTS and TOEFL score groups tended to achieve higher mean GPAs compared to those in the lower score groups within the same university. Overall, this pattern suggests a positive association between English language proficiency and academic achievement at each institution.

CGPA by Discipline of Study

The mean CGPA by discipline of study for both universities is presented in Table 4.2. At University A, students in STEM disciplines had the highest mean CGPA ($M = 6.06$, $SD = 1.82$, $n = 2,806$), followed by those in Social Science and Arts ($M = 5.87$, $SD = 1.75$, $n = 3,358$). Business students had the lowest mean CGPA ($M = 5.60$, $SD = 1.85$, $n = 3,509$). At University B, a different finding was observed. Business students had the highest mean CGPA within their university ($M = 3.09$, $SD = 0.62$, $n = 59$), followed by students in STEM disciplines ($M = 2.77$, $SD = 0.84$, $n = 1,813$) and then those in Social Science and Arts ($M = 2.71$, $SD = 0.83$, $n = 608$). However, the small sample size for Business students at University B ($n = 59$) should be considered when interpreting these results, as observations might not be fully representative and would benefit from a larger sample size to enhance the reliability and generalizability of this finding. Overall, these patterns indicate that the relationship between discipline of study and academic achievement varies between institutions.

Table 4.2

Mean CGPA by discipline of study by university

Discipline	University A			University B		
	M	SD	N	M	SD	N
Social Science & Arts	5.87	1.75	3,358	2.71	0.83	608
Business	5.60	1.85	3,509	3.09	0.62	59
STEM	6.06	1.82	2,806	2.77	0.84	1,813
Total	5.83	1.82	9,673	2.76	0.84	2,480

CGPA by Country of Citizenship

The mean CGPA by country of citizenship is shown in Table 4.3 for both universities. At University A, students from countries other than China had a higher mean CGPA ($M = 6.12$, $SD = 1.80$, $n = 4,389$) compared to students from China ($M = 5.58$, $SD = 1.78$, $n = 5,284$)—a difference of approximately 0.54 CGPA points. Similarly, at University B, students from countries other than China had a slightly higher mean CGPA ($M = 2.79$, $SD = 0.84$, $n = 858$) compared to students from China ($M = 2.75$, $SD = 0.83$, $n = 1,622$), although the difference was smaller—approximately 0.04 CGPA points on a 4-point scale.

Table 4.3

Mean CGPA by country of citizenship by university

Country of Origin	University A			University B		
	M	SD	N	M	SD	N
China	5.58	1.78	5,284	2.75	0.83	1,622
All Other Countries	6.12	1.80	4,389	2.79	0.84	858
Total	5.83	1.82	9,673	2.76	0.84	2,480

CGPA by Gender

The mean CGPA by gender is presented in Table 4.4 for both universities. At University A, women had a higher mean CGPA ($M = 6.01$, $SD = 1.76$, $n = 5,006$) compared to men ($M = 5.62$, $SD = 1.86$, $n = 4,667$), amounting to a difference of approximately 0.39 GPA points on a 9-point scale. Similarly, at University B, women achieved a higher mean CGPA ($M = 2.81$, $SD =$

0.81, $n = 1,030$) compared to men ($M = 2.73$, $SD = 0.85$, $n = 1,450$)—a difference of approximately 0.08 GPA points on a 4-point scale.

Table 4.4

Mean CGPA by gender by university

Gender	University A			University B		
	M	SD	N	M	SD	N
Men	5.62	1.86	4,667	2.73	0.85	1,450
Women	6.01	1.76	5,006	2.81	0.81	1,030
Total	5.83	1.82	9,673	2.76	0.84	2,480

CGPA by Age

Age and cumulative Grade Point Average (CGPA) among international undergraduate students at both University A and University B could not easily be presented in a table due to the number of values. Therefore, a summarized analysis of the interaction between age and CGPA is provided (Table 4.5) with age categorized into two groups: students aged 21 and under and students aged over 21. For the presentation of this data, this age range was chosen because, at both universities, most students were 18-21. This is likely because age 21 often represents the upper limit of traditional college-aged students. As a reminder, age was treated as a continuous variable in this dissertation, thus this grouping into two categories is for ease of data presentation only.

At University A, most international undergraduate students were between the ages of 18 and 21, with the most populous age groups being 18 (2,684 students) and 19 years old (2,683 students). Students aged 18 had a mean CGPA of 5.74 ($SD = 1.80$, $n = 2,684$), while those aged 19 had a mean CGPA of 5.71 ($SD = 1.79$, $n = 2,683$). The mean CGPA remained relatively stable within this age range as students aged 20 achieved an average CGPA of 5.75 ($SD = 1.85$, $n = 1,639$), and those aged 21 had a mean CGPA of 5.71 ($SD = 1.85$, $n = 869$). Starting from age 22, a noticeable upward trend in mean CGPA emerged. Students aged 23 attained a higher mean CGPA of 6.21 ($SD = 1.79$, $n = 290$), and this increase continued with students aged 25 reaching a mean CGPA of 6.60 ($SD = 1.66$, $n = 111$). The highest mean CGPAs were observed among students aged 31 and above, with mean CGPAs exceeding 7.00. However, due to smaller sample sizes in these older age groups, the reliability of these observations was limited. Overall, students aged over 21 had a higher mean CGPA ($M = 6.31$, $SD = 1.77$, $n = 1,476$) compared to those aged 21 and under ($M = 5.74$, $SD = 1.81$, $n = 8,197$).

Similarly, at University B, most international undergraduate students were also between 18 and 21 years old. Within this age group, the mean CGPA was relatively consistent. Students aged 18 had a mean CGPA of 2.67 ($SD = 0.84$, $n = 141$), and those aged 19 had a mean CGPA of 2.73 ($SD = 0.86$, $n = 417$). Students aged 20 and 21 had slightly higher mean CGPAs of 2.79 ($SD = 0.83$, $n = 535$) and 2.79 ($SD = 0.82$, $n = 624$), respectively. Beyond age 21, a slight increase in mean CGPA was observed; students aged 22 had a mean CGPA of 2.77 ($SD = 0.84$, $n = 481$), and those aged 24 achieved a mean CGPA of 2.94 ($SD = 0.76$, $n = 59$). However, sample sizes in older age groups were smaller, which might affect the reliability of these findings. Overall, students aged over 21 had a marginally higher mean CGPA ($M = 2.77$, $SD =$

0.83, $n = 750$) compared to those aged 21 and under ($M = 2.76$, $SD = 0.84$, $n = 1,730$), indicating a minimal difference of approximately 0.01 CGPA points, on average.

Table 4.5

Mean CGPA by age (range) by university

Age	University A			University B		
	M	SD	N	M	SD	N
21 years or less	5.74	1.81	8,197	2.76	0.84	1,730
More than 21 years	6.31	1.77	1,476	2.77	0.83	750
Total	5.83	1.82	9,673	2.76	0.84	2,480

Regression Analysis Results

This section presents the findings from multiple regression analyses conducted on the final samples of each university. The analyses aimed to examine the relationship between ELP pathway (represented by the EAP pathway and different levels of language proficiency through IELTS and TOEFL scores) and CGPA at the end of the first year of study. Furthermore, analyses sought to progressively control for the moderator variables of academic discipline, student origin, gender, and age.

Five sequential models were developed to assess the incremental contribution of these variables, allowing for a comprehensive understanding of their impact on academic achievement

operationalized by CGPA. Model 1 focused solely on ELP pathway, while subsequent models progressively included discipline (Model 2), origin (Model 3), gender (Model 4), and age (Model 5). As a result, the following section details the regression coefficients, statistical significance, and the explanatory power of each model for both universities.

University A

Table 4.6 presents the results of regression analyses for University A. Each model is described with emphasis placed on statistically significant results and notable changes between models.

The initial model (Model 1) explored the relationship between CGPA and ELP pathway with no moderator variables included in the analysis ($R^2 = 0.03$, $F(6, 9666) = 51.87$). The adjusted R^2 was 0.031, indicating that approximately 3.1% of the variance in CGPA was explained by the pathway categories alone. Significant positive associations were found between all pathway categories and CGPA compared to the EAP reference group. Specifically, students in the IELTS Low group had a CGPA that was, on average, 0.362 points higher than those in the EAP group ($p < 0.001$). Students in the TOEFL Low group exhibited a significant difference in CGPA by 0.412 points ($p < 0.001$). Students in the IELTS Mid group had a CGPA difference of 0.800 points ($p < 0.001$). The TOEFL Mid group showed a significant difference of 0.206 points ($p < 0.01$). The IELTS High group demonstrated the most substantial positive effect, with a CGPA that was 1.406 points higher ($p < 0.001$). Lastly, students in the TOEFL High group had a significantly higher CGPA, with an average difference of 0.978 points compared to those in the EAP group ($p < 0.001$).

In Model 2, discipline of study was introduced as the first moderator variable, resulting in a slight increase in the adjusted R^2 to 0.035 ($R^2 = 0.04$, $F(8, 9664) = 44.86$). The significant positive associations for the pathway categories observed in Model 1 remained relatively stable, with slight decreases in coefficients. The introduction of discipline of study revealed significant effects for both Business and STEM disciplines in comparison to Social Sciences and Humanities. Students in the Business discipline had a significantly lower CGPA by 0.180 points compared to those in Social Sciences and Humanities ($p < 0.001$). Conversely, students in the STEM discipline exhibited a 0.134 point increase in CGPA ($p < 0.01$) when compared to those in Social Sciences and Humanities.

In Model 3, country of origin was incorporated as an additional predictor alongside pathway and discipline ($R^2 = 0.04$, $F(8, 9663) = 47.24$, adjusted $R^2 = 0.041$). Pathway group coefficients and their significance levels remained largely unchanged from Model 2, with slight decreases in most categories. The exception was the TOEFL Mid group, which saw its coefficient decrease to 0.0512 and become statistically insignificant ($p = 0.658$). The negative association for the Business discipline diminished to -0.122 points ($p < 0.01$), while the positive effect of the STEM discipline decreased to 0.0915 points ($p < 0.05$). Origin emerged as a significant positive predictor, with students from all other countries seeing an increase of 0.178 points in CGPA compared to those from China ($p < 0.001$).

Gender was introduced as an additional predictor in Model 4, increasing the adjusted R^2 to 0.051 ($R^2 = 0.05$, $F(10, 9662) = 53.29$). Pathway coefficients remained similar, with no changes in statistical significance. The negative association for the Business discipline further diminished to -0.0741 points and became statistically insignificant ($p = 0.091$). Conversely, the positive effect of the STEM discipline increased to 0.153 points ($p < 0.001$). Gender emerged as a significant

positive predictor, with women achieving, on average, 0.371 points higher in CGPA than men ($p < 0.001$).

In the final model, Model 5, age was introduced as an additional predictor alongside pathway, discipline, origin, and gender ($R^2 = 0.06$, $F(11, 9661) = 59.94$, adjusted $R^2 = 0.063$). Most pathway groups continued to be statistically significant, with slight changes noted. However, the difference between the TOEFL Low group and the EAP group became statistically insignificant ($p = 0.186$), similar to the TOEFL Mid group noted in Model 3. The Business discipline remained statistically insignificant in comparison to its reference category, while the positive effect of the STEM discipline slightly decreased to 0.118 points ($p < 0.05$). Origin remained a robust positive predictor, with students from all other countries seeing a 0.161 point increase in CGPA ($p < 0.001$). Gender continued to significantly predict CGPA, with women achieving an average increase of 0.355 points ($p < 0.001$). Lastly, Age emerged as a significant positive predictor, with each additional year of age associated with a 0.063 point increase in CGPA ($p < 0.001$).

Table 4.6*Regression Analysis Estimates (and SE) for CGPA for University A*

	Model 1	Model 2	Model 3	Model 4	Model 5
EAP	0	0	0	0	0
1.IELTS LOW	0.362*** (0.0683)	0.347*** (0.0682)	0.385*** (0.0682)	0.373*** (0.0678)	0.404*** (0.0675)
2. TOEFL LOW	0.412*** (0.0951)	0.362*** (0.0951)	0.236* (0.0961)	0.209* (0.0957)	0.186 (0.0951)
3. IELTS MID	0.800*** (0.0664)	0.718*** (0.0674)	0.575*** (0.0695)	0.567*** (0.0692)	0.576*** (0.0687)
4. TOEFL MID	0.206** (0.0772)	0.161* (0.0774)	0.0512 (0.0784)	0.0600 (0.0780)	0.0343 (0.0775)
5. IELTS HIGH	1.406*** (0.119)	1.317*** (0.120)	1.086*** (0.123)	1.069*** (0.122)	1.066*** (0.121)
6. TOEFL HIGH	0.978*** (0.175)	0.920*** (0.175)	0.786*** (0.175)	0.794*** (0.175)	0.774*** (0.173)
Social Sci. + Hum.		0	0	0	0
1.Business		-0.180*** (0.0438)	-0.122** (0.0443)	-0.0741 (0.0443)	-0.0769 (0.0440)
2.STEM		0.134** (0.0461)	0.0915* (0.0463)	0.153*** (0.0464)	0.118* (0.0463)
Origin			0.178*** (0.0222)	0.171*** (0.0221)	0.161*** (0.0220)
Gender				0.371*** (0.0365)	0.355*** (0.0363)
Age					0.0632*** (0.00578)
Constant	5.312*** (0.0594)	5.386*** (0.0658)	5.290*** (0.0667)	5.075*** (0.0696)	3.833*** (0.133)
Adjusted R-squared	0.031	0.035	0.041	0.051	0.063
N	9673	9673	9673	9673	9673

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

University B

As with University A above, Table 4.7 similarly presents the results of regression analyses for University B. Each model is described with emphasis placed on statistically significant results and notable changes between models.

The initial model (Model 1) explored the relationship between CGPA and ELP pathway with no moderator variables included in the analysis ($R^2 = 0.01$, $F(6, 2473) = 4.40$). The adjusted R^2 was 0.008, indicating that approximately 0.8% of the variance in CGPA is explained by the pathway categories alone. Significant positive associations were found between certain pathway categories and CGPA compared to the EAP group. Specifically, students in the TOEFL Mid group had a CGPA that was, on average, 0.184 points higher than those in the EAP group ($p < 0.05$). Students in the IELTS High group had a CGPA increase of 0.237 points ($p < 0.01$), and those in the TOEFL High group exhibited the most substantial positive effect with a CGPA that was 0.700 points higher ($p < 0.001$). Other pathway categories (IELTS Low, TOEFL Low, IELTS Mid) did not show statistically significant differences compared to the EAP group.

In Model 2, discipline of study was introduced as the first moderator variable, resulting in a slight increase in the adjusted R^2 to 0.012 ($R^2 = 0.02$, $F(8, 2471) = 4.88$). The significant positive associations for the pathway categories observed in Model 1 remained relatively stable. The introduction of discipline of study revealed significant effects for the Business discipline in comparison to Social Sciences and Humanities. Students in the Business discipline had a significantly higher CGPA by 0.400 points compared to those in Social Sciences and Humanities ($p < 0.001$). The STEM discipline did not show a significant difference in CGPA compared to the reference group.

In Model 3, Country of origin was incorporated as an additional predictor alongside pathway and discipline ($R^2 = 0.02$, $F(9, 2470) = 4.88$, adjusted $R^2 = 0.012$). Pathway group coefficients and their significance levels remained largely unchanged from Model 2. TOEFL Mid, IELTS High, and TOEFL High continued to be significant positive predictors of CGPA. Regarding discipline, the positive association for Business remained significant with a coefficient of 0.400 ($p < 0.001$). Origin did not emerge as a significant predictor, with a coefficient close to zero, indicating that country of origin did not significantly impact CGPA in this model.

In Model 4, Gender was introduced as another predictor in this model, slightly increasing the adjusted R^2 to 0.014 ($R^2 = 0.02$, $F(10, 2469) = 4.62$). Pathway coefficients remained similar, with TOEFL Mid, IELTS High, and TOEFL High continuing to be significant positive predictors of CGPA. The effect of the Business discipline remained significant, with a coefficient of 0.387 ($p < 0.001$). The STEM discipline remained non-significant. Gender emerged as a significant positive predictor, with women achieving, on average, 0.093 points higher in CGPA than men ($p < 0.01$).

In the final model (Model 5), age was introduced as an additional predictor alongside pathway, discipline, origin, and gender ($R^2 = 0.02$, $F(11, 2468) = 4.20$, adjusted $R^2 = 0.014$). The significant pathway categories from previous models remained consistent, with TOEFL Mid, IELTS High, and TOEFL High continuing to significantly predict higher CGPA compared to the EAP pathway. The difference between the Business discipline and the reference category remained significant (coefficient = 0.384, $p < 0.01$), while the STEM discipline continued to show no significant effect. Gender remained a significant positive predictor, with women achieving an average increase of 0.093 points in CGPA ($p < 0.01$) compared to men. Age did not

emerge as a significant predictor in this model, with a negligible coefficient and no statistical significance, indicating that age did not significantly impact CGPA.

Table 4.7
Regression Analysis Estimates (and SE) for CGPA for University B

	Model 1	Model 2	Model 3	Model 4	Model 5
EAP	0	0	0	0	0
1.IELTS LOW	0.0664 (0.0662)	0.0898 (0.0669)	0.0857 (0.0686)	0.0808 (0.0685)	0.0777 (0.0734)
2.TOEFL LOW	0.0446 (0.210)	0.0534 (0.210)	0.0504 (0.210)	0.0466 (0.210)	0.0421 (0.213)
3.IELTS MID	0.0304 (0.0380)	0.0328 (0.0380)	0.0306 (0.0389)	0.0289 (0.0388)	0.0262 (0.0450)
4.TOEFL MID	0.184* (0.0815)	0.194* (0.0815)	0.195* (0.0815)	0.189* (0.0814)	0.186* (0.0845)
5.IELTS HIGH	0.237** (0.0840)	0.241** (0.0842)	0.246** (0.0862)	0.241** (0.0861)	0.238** (0.0890)
6.TOEFL HIGH	0.700*** (0.176)	0.693*** (0.176)	0.697*** (0.176)	0.690*** (0.176)	0.687*** (0.178)
Social Sci. + Hum		0	0	0	0
1. Business		0.400*** (0.114)	0.400*** (0.114)	0.387*** (0.114)	0.384** (0.117)
2. STEM		0.0550 (0.0398)	0.0571 (0.0405)	0.0771 (0.0412)	0.0774 (0.0413)
Origin			-0.0106 (0.0391)	0.00215 (0.0393)	0.00245 (0.0394)
Gender				0.0926** (0.0351)	0.0927** (0.0352)
Age					0.00152 (0.0128)
Constant	2.718*** (0.0295)	2.664*** (0.0414)	2.668*** (0.0432)	2.612*** (0.0480)	2.582*** (0.257)
Adjusted R-squared	0.008	0.012	0.012	0.014	0.014
N	2480	2480	2480	2480	2480

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In summary, the regression analyses for both University A and University B revealed significant associations between English language proficiency (ELP) pathways and cumulative grade point average (CGPA). At University A, all pathway categories had significantly higher CGPA compared to the EAP reference group, with the IELTS High and IELTS Mid groups exhibiting the highest CGPA. Variables such as discipline of study, origin, gender, and age also played significant roles in predicting CGPA.

At University B, significant positive effects were observed primarily in the higher proficiency groups (TOEFL Mid, IELTS High, and TOEFL High). The Business discipline emerged as a significant positive predictor of CGPA when compared to Social Science and Humanities, and gender remained a significant factor, with women outperforming men. Unlike University A, origin and age did not significantly impact CGPA at University B.

These findings highlighted the nuanced relationships between ELP pathways and academic achievement across both institutions. To further investigate these relationships, the next section presents pairwise comparisons through analyses of covariance, offering a detailed examination of the differences between specific groups within the study.

Analysis of Covariance

This section examines the academic achievement of international students from various language proficiency pathways by comparing their adjusted mean scores. The primary focus is to present how students in the EAP pathway perform academically compared to students in different language testing pathways (TOEFL, IELTS) and proficiency levels (Low, Mid, High). To conduct this comparison, an Analysis of Covariance (ANCOVA) was used as a follow up to the analyses above. ANCOVA allows comparison of the average outcomes (in this case,

CGPA) across multiple groups while controlling other influential factors, such as age and academic discipline (Agresti, 2018). Post hoc pairwise comparisons were conducted to identify statistically significant differences between tests pathways (e.g., IELTS High vs TOEFL High) and within test pathways (e.g., TOEFL Low vs TOEFL Mid). By isolating each pair, post hoc pairwise comparisons provide a clearer picture of which pathways stand out in terms of academic achievement. While ANCOVA does not significantly differ from regression analyses and comparisons of the EAP group to the other pathways produced identical results, this method does allow for additional comparisons between other pathways not represented in the regression models above.

Comparative Analysis between IELTS and TOEFL Pathways

To evaluate the differences in academic achievement between students who entered through IELTS and TOEFL ELP pathways, pairwise comparisons were conducted for each university. The analysis aimed to determine whether there were significant differences in CGPA between students assumed to have similar proficiency levels on IELTS and TOEFL. This comparison was made with all variables discussed in this dissertation included in the ANCOVA.

Students in the IELTS Low group had a significantly higher CGPA than those in the TOEFL Low group (contrast = 0.219, SE = 0.083, 95% CI = [0.056, 0.382], $p < 0.01$) as well as those in the TOEFL Mid group (contrast = 0.370, SE = 0.062, 95% CI = [0.249, 0.492], $p < 0.001$). Conversely, students in the IELTS Low group had a significantly lower CGPA compared to those in the TOEFL High group (contrast = -0.369, SE = 0.167, 95% CI = [-0.697, -0.041], $p < 0.05$).

Students in the IELTS Mid group achieved a significantly higher CGPA compared to those in the TOEFL Low group (contrast = 0.391, SE = 0.079, 95% CI = [0.236, 0.545], $p < 0.001$) and the TOEFL Mid group (contrast = 0.542, SE = 0.057, 95% CI = [0.430, 0.654], $p < 0.001$). No significant difference in CGPA was found between students in the IELTS Mid and TOEFL High groups.

Students in the IELTS High group significantly outperformed those in the TOEFL Low group (contrast = 0.880, SE = 0.126, 95% CI = [0.633, 1.127], $p < 0.001$) and the TOEFL Mid group (contrast = 1.031, SE = 0.114, 95% CI = [0.809, 1.254], $p < 0.001$). Conversely, there was no significant difference in CGPA between students in the IELTS High and TOEFL High groups.

Table 4.8*Pairwise comparisons between pathways at University A*

Comparison	Contrast	Std. Error	95% Confidence Interval	Significance
TOEFL Low vs. IELTS Low	-0.219	0.083	[-0.382, -0.056]	$p < 0.01$
TOEFL Mid vs. IELTS Low	-0.370	0.062	[-0.492, -0.249]	$p < 0.001$
TOEFL High vs. IELTS Low	0.369	0.167	[0.041, 0.697]	$p < 0.05$
IELTS Mid vs. TOEFL Low	0.391	0.079	[0.236, 0.545]	$p < 0.001$
IELTS High vs. TOEFL Low	0.880	0.126	[0.633, 1.127]	$p < 0.001$
TOEFL Mid vs. IELTS Mid	-0.542	0.057	[-0.654, -0.430]	$p < 0.001$
TOEFL High vs. IELTS Mid	0.197	0.165	[-0.126, 0.520]	Not Significant
IELTS High vs. TOEFL Mid	1.031	0.114	[0.809, 1.254]	$p < 0.001$
TOEFL High vs. IELTS High	-0.292	0.192	[-0.668, 0.084]	Not Significant

At University B, students in the IELTS Low group had a significantly lower CGPA compared to those in the TOEFL High group (contrast = -0.5889, SE = 0.1872, 95% CI = [-0.9559, -0.2218], $p < 0.01$). Students in the IELTS Low group had a CGPA that was not significantly different from those in the TOEFL Low group and similarly, there was no significant difference between the IELTS Low and TOEFL Mid groups.

Students in the IELTS Mid group had a significantly lower CGPA compared to those in the TOEFL Mid group (contrast = -0.1601, SE = 0.0802, 95% CI = [-0.3173, -0.0028],

$p < 0.05$). Additionally, the IELTS Mid group had a significantly lower CGPA compared to the TOEFL High group (contrast = -0.6477, SE = 0.1773, 95% CI = [-0.9955, -0.2999], $p < 0.001$).

There was no significant difference between the IELTS Mid and TOEFL Low groups.

Students in the IELTS High group had a significantly lower CGPA compared to those in the TOEFL High group (contrast = -0.4481, SE = 0.1911, 95% CI = [-0.8227, -0.0735], $p < 0.05$). There was no significant difference between the IELTS High group and the TOEFL Mid group nor between the IELTS High and TOEFL Low groups.

The pairwise comparisons at University B reveal that students entering through the TOEFL pathway generally achieved higher CGPAs than those entering through the IELTS pathway, particularly at mid and high proficiency levels. The findings differ from those at University A, where the IELTS pathway was more strongly associated with academic achievement.

Table 4.9*Pairwise comparisons between pathways at University B*

Comparison	Contrast	Std. Error	95% Confidence Interval	Significance
IELTS Low vs. TOEFL Low	0.0130	0.2195	[-0.4175, 0.4434]	Not Significant
IELTS Low vs. TOEFL Mid	-0.1012	0.0984	[-0.2943, 0.0918]	Not Significant
IELTS Low vs. TOEFL High	-0.5889	0.1872	[-0.9559, -0.2218]	$p < 0.01$
IELTS Mid vs. TOEFL Low	0.0459	0.2133	[-0.3724, 0.4642]	Not Significant
IELTS Mid vs. TOEFL Mid	-0.1601	0.0802	[-0.3173, -0.0028]	$p < 0.05$
IELTS Mid vs. TOEFL High	-0.6477	0.1773	[-0.9955, -0.2999]	$p < 0.001$
IELTS High vs. TOEFL Low	0.1537	0.2281	[-0.2936, 0.6011]	Not Significant
IELTS High vs. TOEFL Mid	-0.0395	0.1107	[-0.2567, 0.1776]	Not Significant
IELTS High vs. TOEFL High	-0.4481	0.1911	[-0.8227, -0.0735]	$p < 0.05$

Comparative Analysis within IELTS and TOEFL Pathways

At University A, for students admitted through an IELTS pathway, significant differences were observed across all comparisons within the IELTS pathway. Students in the IELTS Mid group had a significantly higher CGPA than those in the IELTS Low group (contrast = 0.1719, SE = 0.0506, 95% CI = [0.0728, 0.2710], $p < 0.001$). Similarly, the IELTS High group outperformed both the IELTS Low group (contrast = 0.6612, SE = 0.1126, 95% CI = [0.4406, 0.8819], $p < 0.001$) and the IELTS Mid group (contrast = 0.4894, SE = 0.1064, 95% CI = [0.2808, 0.6979], $p < 0.001$). These results indicate a strong positive relationship between higher IELTS proficiency levels and CGPA at University A.

For students admitted through a TOEFL pathway, significant differences were found between the TOEFL High group and both the TOEFL Low and TOEFL Mid groups. Students in the TOEFL High group had a significantly higher CGPA compared to those in the TOEFL Low group (contrast = 0.5879, SE = 0.1779, 95% CI = [0.2391, 0.9367], $p < 0.001$) and the TOEFL Mid group (contrast = 0.7392, SE = 0.1692, 95% CI = [0.4075, 1.0710], $p < 0.001$). However, there was no significant difference between the TOEFL Mid and TOEFL Low groups, indicating that the most substantial CGPA increases are associated with moving from mid to high TOEFL proficiency levels.

Table 4.10*Pairwise Comparisons within IELTS and TOEFL Pathways at University A*

Comparison	Contrast	Std. Error	95% Confidence Interval	Significance
IELTS Mid vs. IELTS Low	0.1719	0.0506	[0.0728, 0.2710]	$p < 0.001$
IELTS High vs. IELTS Low	0.6612	0.1126	[0.4406, 0.8819]	$p < 0.001$
IELTS High vs. IELTS Mid	0.4894	0.1064	[0.2808, 0.6979]	$p < 0.001$
TOEFL Mid vs. TOEFL Low	-0.1514	0.0877	[-0.3233, 0.0206]	Not Significant
TOEFL High vs. TOEFL Low	0.5879	0.1779	[0.2391, 0.9367]	$p < 0.001$
TOEFL High vs. TOEFL Mid	0.7392	0.1692	[0.4075, 1.0710]	$p < 0.001$

At University B, within test comparisons results for the IELTS pathways were as follows (Table 4.11). No significant differences were found between the IELTS Mid and IELTS Low groups or between the IELTS High and IELTS Low groups. However, a significant difference was observed between the IELTS High and IELTS Mid groups (contrast = 0.1996, SE = 0.0863, 95% CI = [0.0304, 0.3688], $p < 0.05$), suggesting that only at the highest proficiency levels does IELTS proficiency impact CGPA at University B.

For students admitted through a TOEFL pathway, the TOEFL High group outperformed the TOEFL Low group (contrast = 0.6018, SE = 0.2759, 95% CI = [0.0609, 1.1428], $p < 0.05$) and the TOEFL Mid group (contrast = 0.4876, SE = 0.1905, 95% CI = [0.1140, 0.8613], $p < 0.05$). No significant difference was found between the TOEFL Mid and TOEFL Low groups.

Table 4.11*Pairwise Comparisons within IELTS and TOEFL Pathways at University B*

Comparison	Contrast	Std. Error	95% Confidence Interval	Significance
IELTS Mid vs. IELTS Low	-0.0588	0.0656	[-0.1875, 0.0698]	Not Significant
IELTS High vs. IELTS Low	0.1408	0.1056	[-0.0663, 0.3478]	Not Significant
IELTS High vs. IELTS Mid	0.1996	0.0863	[0.0304, 0.3688]	$p < 0.05$
TOEFL Mid vs. TOEFL Low	0.1142	0.2254	[-0.3278, 0.5561]	Not Significant
TOEFL High vs. TOEFL Low	0.6018	0.2759	[0.0609, 1.1428]	$p < 0.05$
TOEFL High vs. TOEFL Mid	0.4876	0.1905	[0.1140, 0.8613]	$p < 0.05$

Findings for both universities demonstrated that higher IELTS proficiency levels are associated with better academic achievement, though the strength of this relationship varied between the institutions. At University A, significant differences were observed across all proficiency levels within the IELTS pathway. The IELTS High group significantly outperformed both the IELTS Mid and IELTS Low groups, and the IELTS Mid group also achieved higher CGPAs than the IELTS Low group. At University B, however, the pattern differed. No significant differences were found between the IELTS Low and IELTS Mid groups or between the IELTS Low and IELTS High groups, while the only significant difference observed was

between the IELTS Mid and IELTS High groups, indicating that students made the most significant academic gains once they reached higher proficiency levels.

Interaction Effects

As discussed in Chapter 2, the relationship between language proficiency and academic achievement does not operate in isolation; it is often moderated by other variables such as discipline of study, gender, and country of origin.¹

By incorporating interaction terms into the statistical models, the aim was to uncover whether the impact of English language proficiency on academic achievement varied across different student subgroups and institutional contexts. For example, the effect of a high IELTS score on CGPA might differ between students in the sciences and those in the humanities, or between men and women. Understanding these interaction effects provides deeper insights into the multifaceted nature of academic achievement among international students. All findings presented below are statistically significant at $p < 0.001$ and are visually represented.² Figures show the marginal prediction for variables of interest while holding all other variables constant at their averages. They also include an ‘as-observed’ line which shows the predicted outcome when each observation’s real values for the other variables are used. In other words, instead of holding all other variables at their average, STATA “observes” each record’s actual values and uses those in the prediction. Representing this shows how the predicted outcomes behave in

¹ Age was not included in the interaction effects because incorporating a continuous variable like age in interactions makes the effects conditional on specific age values (which there are many), and doing so can increase multicollinearity, thereby reducing the stability and clarity of the model estimates. Age was therefore treated as a covariate in the overall regression models.

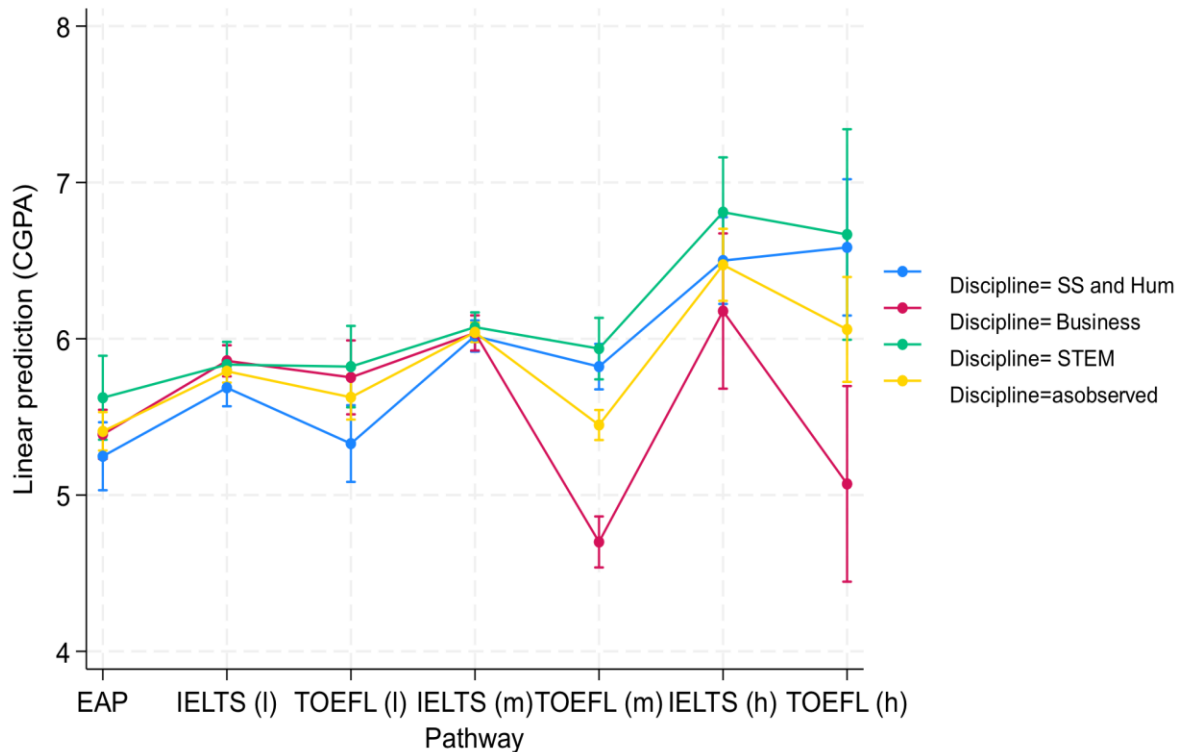
² All figures of this type show the marginal prediction for variables of interest while holding all other variables constant at their averages and include an ‘as-observed’ line which shows the predicted outcome when each observation’s real values for the other variables are used. In other words, instead of holding all other variables at their average, STATA “observes” each record’s actual values and uses those in the prediction. Representing this shows how the predicted outcomes behave in comparison to the actual data.

comparison to the actual data. The results for all interaction effects are presented in Appendices A to F.

Interaction Effects between English Language Proficiency Pathways and Academic Discipline

The full results of the interaction effects between ELP pathways and discipline for both universities are reported in Appendices A and B. As Appendix A shows, at University A, students entering via IELTS or TOEFL pathways often outperformed EAP, with the strongest differences observed at High proficiency levels. For instance, students' CGPA in IELTS High and in STEM reached 6.81, exceeding EAP's 5.62 by +1.19. In SS & Humanities, students' CGPA in TOEFL High averaged 6.58, +1.34 above EAP's 5.25. Conversely, in Business, TOEFL Mid fell to 4.70, -0.69 below EAP's 5.39, while TOEFL High also was found to be slightly below EAP (5.07 vs. 5.39; -0.32). By contrast, students in IELTS High and in Business outperformed at 6.18, +0.79 higher than EAP. Overall, these results demonstrate the advantage of high-proficiency IELTS and TOEFL entrants relative to EAP, with TOEFL Mid and TOEFL High in Business emerging as notable exceptions.

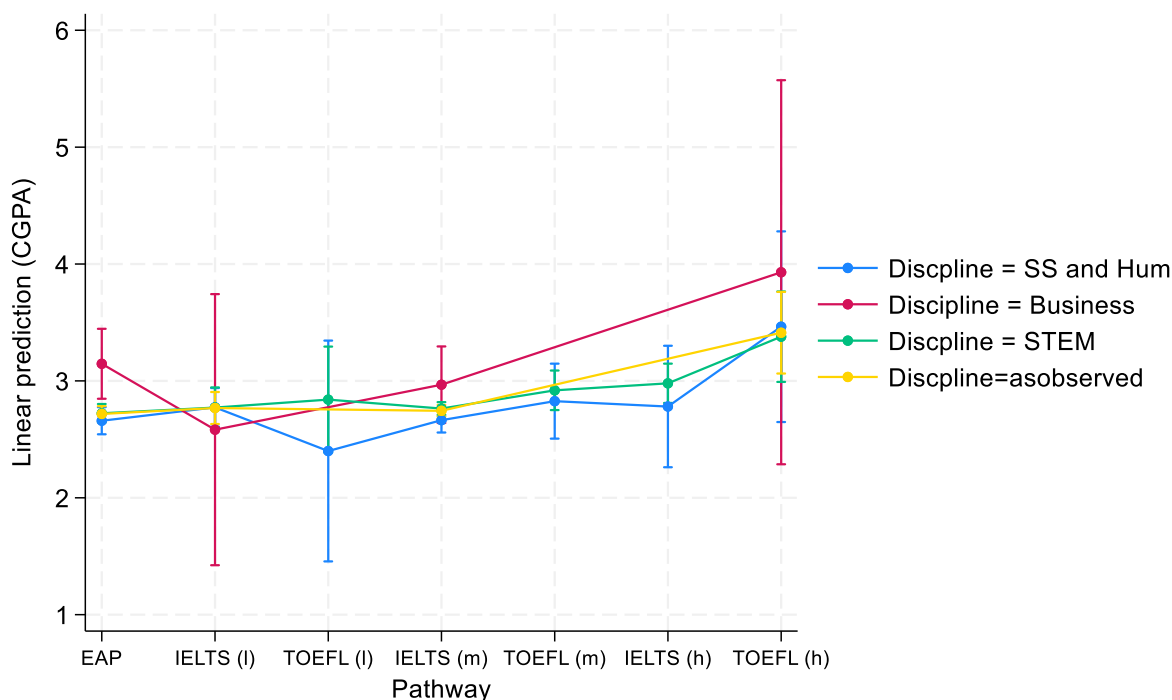
Figure 4.1: Predictive margins with 95% CIs - interaction effects between pathway and discipline (University A) ³



At University B (see Appendix B), testing revealed that none of the interaction effects between pathway and discipline were statistically significant. This indicates that the differences in predicted CGPA across these groups do not vary in a statistically reliable way; however, the predictive margins remain useful for showing the estimated average outcomes for each combination of pathway and discipline group. The general trend of the margins indicated that higher ELP is associated with higher predicted CGPA. Due to small counts within the business discipline specifically, some pathways contained no cases and therefore are not visually represented in Figure 4.2.

³ The as-observed' line represents how the predicted outcomes behave in comparison to the actual data.

Figure 4.2: *Predictive margins with 95% CIs - interaction effects between pathway and discipline (University B)*

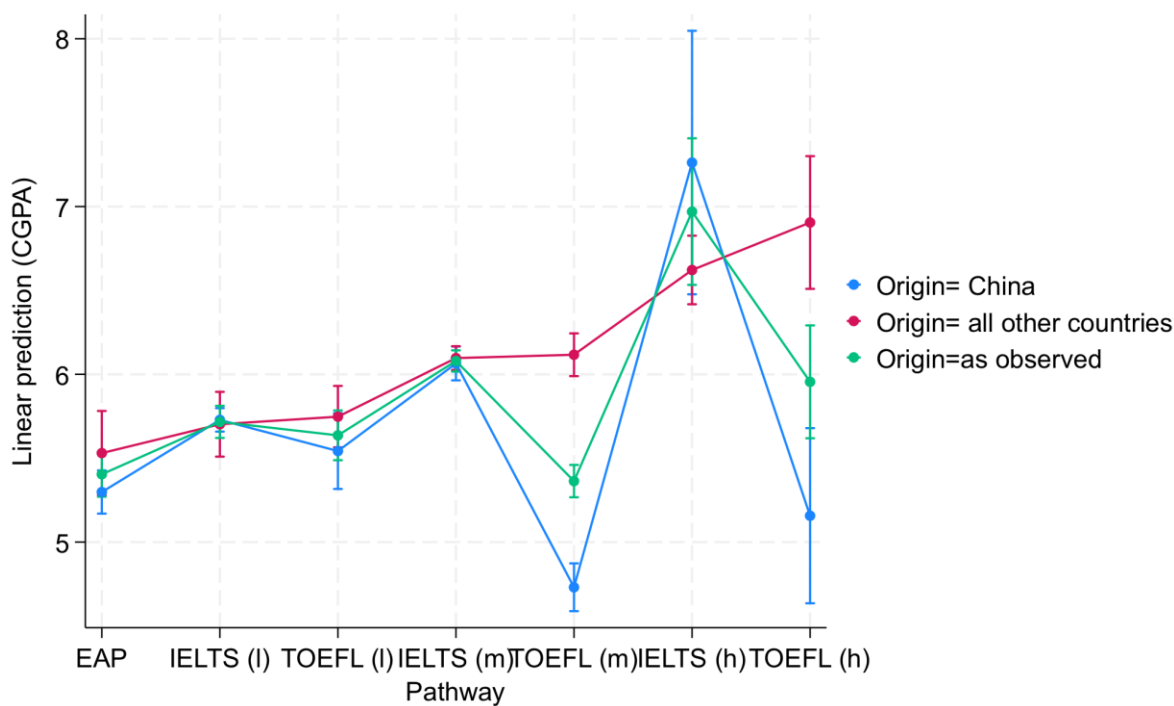


Interaction Effects between English Language Proficiency Pathways and Country of Origin

The full results of the interaction effects between ELP pathways and country of origin for both universities are reported in Appendices C and D. Overall at University A (see Appendix C), students from all other countries achieved a slightly higher predicted CGPA (5.94) than their Chinese counterparts (5.71). Notably, Chinese IELTS High students outperformed EAP Chinese students 1.96 CGPA. Other noteworthy comparisons are that among TOEFL Mid students, non-Chinese students attained a predicted CGPA of 6.12, exceeding the 4.73 observed for Chinese students by approximately +1.39, while TOEFL high students from all other countries reached 6.90—+1.75 above the 5.16 for Chinese students. Conversely, in the IELTS high group, Chinese students outperformed their peers from other countries (7.26 vs. 6.62; +0.64), but by a smaller

margin. Overall, these results underscore the advantage of high-proficiency IELTS and TOEFL pathways relative to EAP, with students' country of origin further moderating academic achievement.

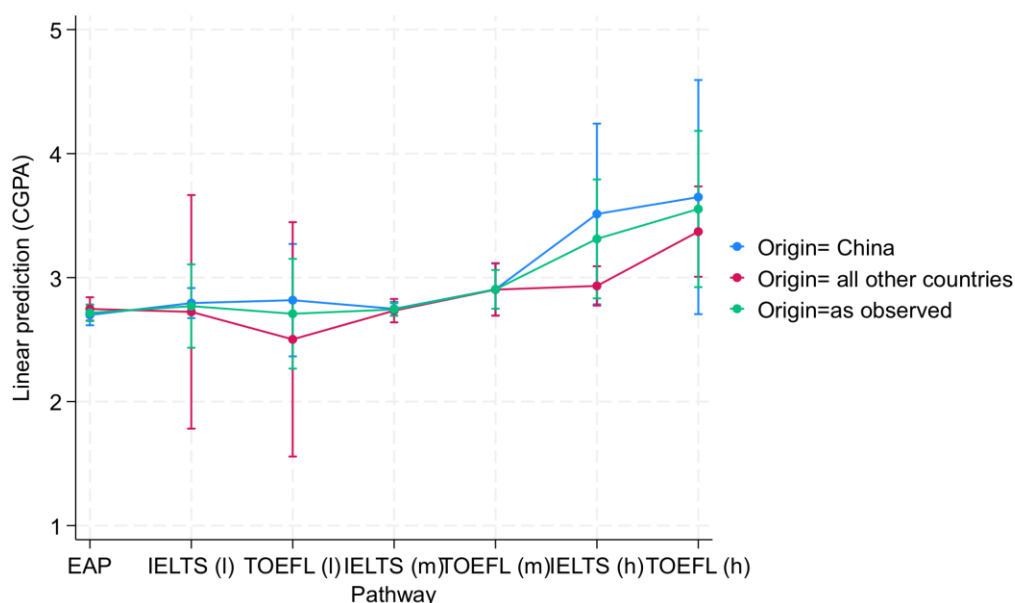
Figure 4.3: *Predictive margins with 95% CIs - interaction effects between pathway and origin (University A)*



At University B (see Appendix D), most pairwise comparisons among pathway–origin groups did not reach statistical significance. However, the contrast between TOEFL high students from other countries and EAP entrants from China was statistically significant, with a difference of +0.67. Despite lack of statistical significance during interaction comparisons, predicted margins revealed a clear increase in CGPA with higher proficiency pathways. Predicted scores rise from approximately 2.69 for baseline Chinese EAP students and 2.74 for students from other countries to Chinese students at IELTS high achieving a CGPA of 3.51 and

TOEFL high at 3.65. In comparison, non-Chinese students achieved a CGPA of 2.93 for IELTS high and 3.37 for TOEFL high. This trend confirms that students in the highest proficiency categories consistently outperform those in lower categories, with slight moderating effects based on country of origin.

Figure 4.4: *Predictive margins with 95% CIs - interaction effects between pathway and origin (University B)*

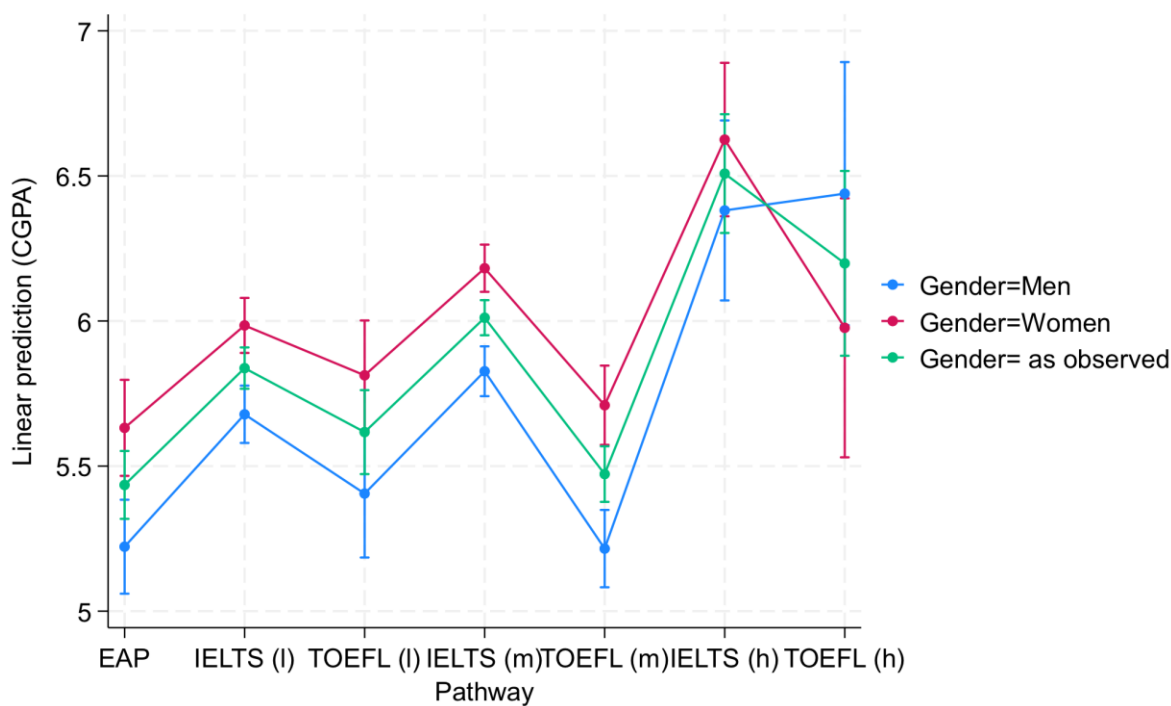


Interaction Effects between English Language Proficiency Pathways and Gender

The full results of the interaction effects between ELP pathways and gender for both universities are reported in Appendices E and F. At University A (see Appendix E), gender differences in predicted CGPA consistently favored women across most proficiency levels and pathways. For example, among EAP entrants women scored an average of 5.63 CGPA compared to 5.22 for men. Similarly, in the IELTS low and IELTS mid groups, women outperformed men by 0.31 and 0.36 CGPA, respectively, and in the TOEFL mid group by 0.49 CGPA. No

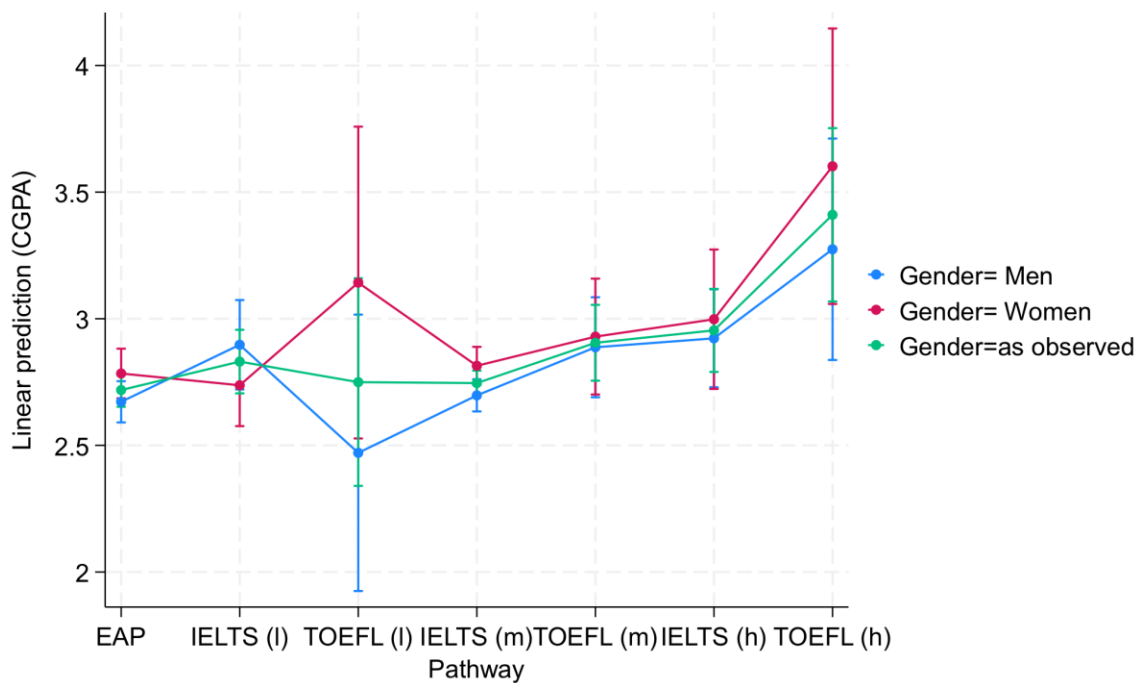
statistically significant gender differences were observed in the IELTS high and TOEFL high categories. These findings suggest that while the overall trend favors women, the magnitude of the gender gap varies by pathway group.

Figure 4.5: *Predictive margins with 95% CIs - interaction effects between pathway and gender (University A)*



At University B (see Appendix F), analyses demonstrated that women generally achieved higher predicted CGPAs than men across most English language proficiency pathways at University B. However, testing revealed that none of the interaction effects between pathway and gender were statistically significant. However, the predictive margins demonstrated that women's CGPA was consistently higher than those of men, reinforcing the trend that women tend to perform better on average—even though these differences did not reach statistical significance.

Figure 4.6: Predictive margins with 95% CIs - interaction effects between pathway and gender
(University B)



Having established a clear relationship between English language proficiency pathways and academic achievement as measured by CGPA, the subsequent section will explore how these pathways relate to another critical aspect of academic achievement: credits earned during the first year of study.

Relationship between ELP pathways and Credits Earned

The following sections present the descriptive statistics for Research Question 2, which examines the relationship between English Language Proficiency (ELP) pathways and academic achievement as measured by cumulative credits earned during the first year of study. Similar to the previous section, pathway and moderator variables by credits earned by University A and University B are presented side by side to facilitate comparison.

Descriptive Statistics

Credits Earned by ELP Pathway

Table 4.12 presents the mean number of credits earned by students in each ELP pathway group at both universities. At University A, the total sample consisted of 9,673 international undergraduate students, with an overall mean of 10.35 cumulative first-year credits earned (SD = 4.40). At University B, the total sample included 2,480 international undergraduate students, with an overall mean of 13.24 units earned (SD = 3.70).

Table 4.12

Mean credits earned by English language proficiency pathway by university

Pathway Group	University A			University B		
	M	SD	N	M	SD	N
EAP	8.10	4.71	908	12.81	3.85	795
IELTS Low	9.80	3.79	2,814	15.63	5.50	197
TOEFL Low	10.37	3.98	582	14.00	4.13	16
IELTS Mid	11.27	4.47	3,627	12.90	3.15	1,217
TOEFL Mid	10.01	4.58	1,323	13.53	3.72	120
IELTS High	12.24	4.17	301	14.99	1.89	112
TOEFL High	10.98	4.27	118	14.75	1.54	23
Total	10.35	4.40	9,673	13.24	3.70	2,480

At University A, students admitted through the EAP pathway had the lowest mean credits earned ($M = 8.10$, $SD = 4.71$, $n = 908$), indicating that these students, on average, completed fewer credits in their first year compared to other groups. A progressive increase in mean credits earned was observed with higher ELP. For instance, students in the IELTS Low group had a mean of 9.80 credits earned ($SD = 3.79$, $n = 2,814$), while those in the IELTS Mid and IELTS High groups had higher means of 11.27 ($SD = 4.47$, $n = 3,627$) and 12.24 ($SD = 4.17$, $n = 301$) credits earned, respectively. Similarly, students in the TOEFL High group also showed an increase in credits earned compared to those in the TOEFL Low group.

At University B, a somewhat similar pattern emerged, albeit with some notable deviations. Students in the EAP group had the lowest mean credits earned ($M = 12.81$, $SD = 3.85$, $n = 795$), and those in the IELTS Low group had a higher mean of 15.63 credits earned ($SD = 5.50$, $n = 197$), as did those in the TOEFL Low group ($M = 14.00$, $SD = 4.13$, $n = 16$). However, the IELTS Mid group had a lower mean of credits earned ($M = 12.90$, $SD = 3.15$, $n = 1,217$) compared to the IELTS Low group, which deviated from the trend. Despite this deviation, both mid groups had higher means than the EAP group. Students in the IELTS High group had mean credits earned of 14.99 ($SD = 1.89$, $n = 112$), which was lower than the IELTS Low group. This anomaly was not replicated for the same TOEFL groups, which followed more of the predicted trend.

Generally, both universities demonstrated a positive association between higher English language proficiency and the number of credits earned in the first year. Students with higher IELTS and TOEFL scores tended to accumulate more credits, suggesting that greater English proficiency might facilitate academic progress.

Credits Earned by Discipline of Study

Table 4.13 presents the mean number of credits earned by students in each discipline of study at both universities. At University A, students in the Social Science & Arts disciplines had the highest mean credits earned ($M = 11.08$, $SD = 4.57$, $n = 3,358$), indicating that these students, on average, completed more credits in their first year compared to other disciplines. Students in the STEM disciplines had a mean credits earned of 10.27 ($SD = 4.24$, $n = 2,806$), while Business students had the lowest mean credits earned ($M = 9.70$, $SD = 4.24$, $n = 3,509$).

At University B, students in the STEM disciplines had the highest mean units passed ($M = 13.48$, $SD = 3.46$, $n = 1,813$), followed by those in Social Science & Arts ($M = 12.63$, $SD = 4.36$, $n = 608$). Business students had the lowest mean units passed ($M = 12.20$, $SD = 2.22$, $n = 59$). However, similar to previous observations, the small sample size for Business students at University B should be considered when interpreting these results.

Overall, the data indicated that at University A, students in Social Science & Arts tended to earn more credits in their first year compared to those in Business and STEM disciplines. Conversely, at University B, students in STEM disciplines earned the most credits, followed by Social Science & Arts, with Business students earning the least. These observations highlighted further differences between the two universities.

Table 4.13*Mean credits earned by discipline of study by university*

Discipline	University A			University B		
	M	SD	N	M	SD	N
Social Science & Arts	11.08	4.57	3,358	12.63	4.36	608
Business	9.70	4.24	3,509	12.20	2.22	59
STEM	10.27	4.24	2,806	13.48	3.46	1,813
Total	10.35	4.40	9,673	13.24	3.70	2,480

Credits Earned by Country of Citizenship

Table 4.14 presents the mean number of credits earned by students grouped by country of origin at both universities. At University A, students from All Other Countries had higher mean credits earned ($M = 11.32$, $SD = 4.48$, $n = 4,417$) compared to students from China ($M = 9.53$, $SD = 4.16$, $n = 5,256$). Similarly, at University B, students from All Other Countries also had higher mean credits earned ($M = 13.84$, $SD = 3.03$, $n = 858$) compared to students from China ($M = 12.92$, $SD = 3.97$, $n = 1,622$). Overall, the mean credits earned at University B was higher than that at University A. This was reflected in the total mean credits earned at University B being 13.24 ($SD = 3.70$) compared to the total mean credits earned at University A of 10.35 ($SD = 4.40$)¹

Table 4.14*Mean credits earned by country of citizenship by university*

Country of Origin	University A			University B		
	M	SD	N	M	SD	N
China	9.53	4.16	5,256	12.92	3.97	1,622
All Other Countries	11.32	4.48	4,417	13.84	3.03	858
Total	10.35	4.40	9,673	13.24	3.70	2,480

Credits Earned by Gender

Table 4.15 presents the mean number of credits earned by students based on their gender at both universities. At University A, women had higher mean credits earned ($M = 10.78$, $SD = 4.35$, $n = 5,028$) compared to men ($M = 9.88$, $SD = 4.41$, $n = 4,645$). This indicated that, on average, women completed approximately one more credit than men in their first year. In contrast, at University B, men had a higher mean units passed ($M = 13.42$, $SD = 3.60$, $n = 1,450$) compared to women ($M = 12.98$, $SD = 3.83$, $n = 1,030$). This suggested that, on average, men completed approximately 0.44 more credits than women.

Table 4.15*Mean credits earned by gender by university*

Gender	University A			University B		
	M	SD	N	M	SD	N
Men	9.88	4.41	4,645	13.42	3.60	1,450
Women	10.78	4.35	5,028	12.98	3.83	1,030
Total	10.35	4.40	9,673	13.24	3.70	2,480

Credits Earned by Age

Table 4.16 presents the mean number of credits earned by students based on their age at both universities. Similarly, this information is categorized into two age groups (<21 and >21) with CGPA data above for ease of presentation.

At University A, students aged 18 had a mean credits earned of 10.81 (SD = 4.41, $n = 2,684$), while those aged 19 had a slightly lower mean of 10.28 credits (SD = 4.48, $n = 2,683$). There was a subtle decline in mean credits earned as age increased within this range; students aged 20 earned an average of 10.09 credits (SD = 4.35, $n = 1,639$), and those aged 21 earned 9.85 credits (SD = 4.41, $n = 869$). For students over 21, the mean credits earned fluctuated without a consistent pattern. Overall, students aged 21 and under tended to earn slightly more credits ($M = 10.42$, SD = 4.43, $n = 8,197$) than those over 21 ($M = 9.90$, SD = 4.26, $n = 1,476$).

Similarly, at University B, students aged 18 had a mean credit passed of 13.13 (SD = 3.29, $n = 141$), and those aged 19 had a mean of 13.08 units (SD = 3.73, $n = 417$). Students aged 20 and 21 had slightly higher means, with averages of 13.62 units (SD = 3.47, $n = 535$) and 13.46 units (SD = 3.56, $n = 624$). Beyond age 21, there was a gradual decrease in mean units passed; students aged 22 earned an average of 13.28 units (SD = 3.90, $n = 481$), and those aged 24 earned 11.75 units (SD = 3.98, $n = 59$). Overall, students aged 21 and under at University B had a higher mean units passed ($M = 13.37$, SD = 3.55, $n = 1,730$) compared to students over 21 ($M = 12.94$, SD = 3.95, $n = 750$).

Table 4.16*Mean CGPA by age (range) by university*

Age	University A			University B		
	M	SD	N	M	SD	N
< 21	10.42	4.43	8,197	13.37	3.55	1,730
> 21	9.90	4.26	1,476	12.94	3.95	750
Total	10.35	4.40	9,673	13.24	3.70	2,480

Regression Analysis Results

This section presents the findings from multiple regression analyses conducted for University A and University B, focusing on the second dependent variable: credits earned during the first year of study. The analyses aimed to examine the relationship between English Language Proficiency (ELP) pathways—represented by the EAP pathway and varying levels of language proficiency through IELTS and TOEFL scores—and the number of cumulative credits earned by international undergraduate students at the end of the first year of study. Additionally, the analyses sought to progressively control for moderator variables including academic discipline, country of origin, gender, and age. As with the previous section, five sequential models were developed to assess the incremental contribution of these variables, allowing for a comprehensive understanding of their impact on academic achievement operationalized by cumulative credits earned (CCE).

University A

Table 4.17 presents the results of regression analyses for University A. Each model is described with emphasis placed on statistically significant results and notable changes between models.

The initial model examined the relationship between cumulative credits earned and ELP pathways without any moderator variables (Adjusted $R^2 = 0.052$, $F(6, 9666) = 88.62$, $p < 0.001$).

The results indicated that all ELP pathway groups had a significant positive association with credits earned compared to the reference group (EAP pathway). Students in the IELTS Low group earned, on average, 1.697 more credits than EAP students ($p < 0.001$). Those in the TOEFL Low group earned 2.269 additional credits ($p < 0.001$). The IELTS Mid group showed the largest increase, with students earning 3.165 more credits than the EAP group ($p < 0.001$). Students in the TOEFL Mid group earned 1.903 more credits ($p < 0.001$). The IELTS High group showed a significant increase of 4.135 credits ($p < 0.001$). Finally, students in the TOEFL High group earned 2.875 more credits than those in the EAP group ($p < 0.001$).

In Model 2, discipline of study was introduced as a moderator variable, resulting in a slight increase in the adjusted R^2 to 0.064 ($F(8, 9664) = 83.40$, $p < 0.001$). The inclusion of discipline slightly decreased the coefficients of the ELP pathways, but all remained significant. For instance, the coefficient for IELTS Mid decreased from 3.165 to 3.096 credits ($p < 0.001$). The introduction of academic discipline revealed significant differences for both Business and STEM disciplines in comparison to Social Sciences and Humanities. Students in the Business discipline earned 1.039 fewer credits than those in Social Sciences and Humanities ($p < 0.001$). Similarly, students in the STEM discipline earned 1.025 fewer credits ($p < 0.001$). The negative coefficients

for Business and STEM indicate that students in these disciplines earned fewer credits during their first year compared to their peers in Social Sciences and Humanities.

Country of origin was added to model 3 as an additional predictor, resulting in an adjusted R-squared of 0.080 ($F(9, 9663) = 94.56, p < 0.001$). The pathway group coefficients showed slight increases and remained significant. The negative association for the Business discipline weakened slightly to -0.819 credits ($p < 0.001$), while the negative effect of the STEM discipline became more pronounced, with a coefficient of -1.188 credits ($p < 0.001$). Country of origin emerged as a significant predictor, with students from countries other than China earning 0.678 more credits than students from China ($p < 0.001$).

Gender was introduced as an additional predictor in Model 4, increasing the adjusted R-squared to 0.085 ($F(10, 9662) = 90.40, p < 0.001$). The pathway coefficients remained relatively stable and significant. The negative association for the Business discipline further weakened to -0.730 credits ($p < 0.001$), and the negative effect of the STEM discipline slightly reduced to -1.074 credits ($t = -9.74, p < 0.001$). Country of origin remained significant, with a coefficient of 0.666 credits ($p < 0.001$). Gender emerged as a significant predictor, with female students earning 0.690 more credits than male students ($p < 0.001$).

In the final model, age was included alongside the other variables, resulting in a final adjusted R-squared of 0.089 ($F(11, 9661) = 87.20, p < 0.001$). The pathway coefficients remained significant with minimal changes, as did the coefficients for discipline of study, while country of origin's coefficient increased marginally to 0.680 credits ($p < 0.001$). Gender's effect slightly increased to 0.711 credits ($p < 0.001$). Age emerged as a significant predictor, with each additional year of age associated with earning 0.084 fewer credits ($p < 0.001$).

Table 4.17
Regression Analysis Estimates (and SE) for CCE for University A

	Model 1	Model 2	Model 3	Model 4	Model 5
EAP	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)
1.IELTS LOW	1.697***	1.652***	1.797***	1.773***	1.731***
	(0.163)	(0.163)	(0.162)	(0.161)	(0.161)
2. TOEFL LOW	2.269***	2.203***	1.720***	1.669***	1.699***
	(0.227)	(0.227)	(0.228)	(0.227)	(0.227)
3. IELTS MID	3.165***	3.096***	2.548***	2.534***	2.522***
	(0.159)	(0.161)	(0.165)	(0.164)	(0.164)
4. TOEFL MID	1.903***	1.748***	1.328***	1.345***	1.379***
	(0.185)	(0.184)	(0.186)	(0.185)	(0.185)
5. IELTS HIGH	4.135***	3.882***	2.997***	2.966***	2.971***
	(0.285)	(0.285)	(0.291)	(0.290)	(0.290)
6. TOEFL HIGH	2.875***	2.615***	2.101***	2.117***	2.145***
	(0.419)	(0.417)	(0.416)	(0.415)	(0.414)
Social Sci. + Hum.		0	0	0	0
		(.)	(.)	(.)	(.)
1.Business		-1.039***	-0.819***	-0.730***	-0.726***
		(0.104)	(0.105)	(0.105)	(0.105)
2.STEM		-1.025***	-1.188***	-1.074***	-1.028***
		(0.110)	(0.110)	(0.110)	(0.110)
Origin			0.678***	0.666***	0.680***
			(0.0527)	(0.0525)	(0.0525)
Gender				0.690***	0.711***
				(0.0866)	(0.0865)
Age					-0.0836***
					(0.0138)
Constant	8.104***	8.853***	8.485***	8.086***	9.729***
	(0.142)	(0.157)	(0.158)	(0.165)	(0.317)
Adjusted R-squared	0.052	0.064	0.080	0.085	0.089
N	9673	9673	9673	9673	9673

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

University B

As with University A above, Table 4.18 similarly presents the results of regression analyses for University B. Each model is described with emphasis placed on statistically significant results and notable changes between models.

The initial model examined the relationship between cumulative credits earned and ELP pathways without any moderator variables. The adjusted R-squared was 0.052 ($F(6, 2473) = 22.81, p < 0.001$). The results indicated that certain ELP pathway groups had a significant positive association with credits earned compared to the reference group (EAP pathway). Students in the IELTS Low group earned, on average, 2.822 more credits than EAP students ($p < 0.001$). Those in the IELTS High group earned 2.179 additional credits ($p < 0.001$) compared to the EAP group ($p < 0.001$). The TOEFL High group showed an increase of 1.936 credits ($p < 0.05$). Students in the TOEFL Mid group earned 0.717 more credits than EAP students ($p < 0.05$). The coefficients for TOEFL Low (1.188 credits) and IELTS Mid (0.087 credits) were not statistically significant in this model.

In Model 2, discipline of study was introduced as a moderator variable, resulting in an increase in the adjusted R-squared to 0.068 ($F(8, 2471) = 22.44, p < 0.001$). The inclusion of discipline slightly affected the coefficients of the ELP pathways. For instance, the coefficient for IELTS Low increased slightly to 3.074 credits ($p < 0.001$) from 2.822, and the coefficient for IELTS High decreased to 1.941 credits ($p < 0.001$) from 2.179. The coefficient for TOEFL Mid decreased to 0.622 credits and became not statistically significant. Discipline revealed significant differences for the STEM group in comparison to Social Sciences and Humanities. Students in the STEM discipline earned 1.116 more credits than those in Social Sciences and Humanities

($p < 0.001$). The coefficient for Business was positive (0.033 credits) but not statistically significant.

Country of origin was added to model 3 as an additional predictor, resulting in an adjusted R-squared of 0.080 ($F(9, 2470) = 22.90, p < 0.001$). The pathway group coefficients showed slight changes. The coefficient for IELTS Low increased to 3.439 credits ($p < 0.001$), while the coefficient for IELTS High decreased to 1.501 credits from 1.94 and 2.17 in the previous models but remained significant ($p < 0.001$). The coefficients for TOEFL Low, TOEFL Mid, and TOEFL High were not statistically significant in this model. Country of origin emerged as a significant predictor. Students from countries other than China earned 0.944 more credits than students from China ($p < 0.001$). The coefficient for the STEM discipline remained significant, with students earning 0.929 more credits than those in Social Sciences and Humanities ($p < 0.001$).

In Model 4, gender was introduced as an additional predictor, with the adjusted R-squared remaining at 0.080 ($F(10, 2469) = 20.00, p < 0.001$). The coefficients in this model remained relatively stable and unaffected from the introduction of the gender variable. Gender was also found to not be statistically significant at University B.

In the final model, age was included alongside the other variables, resulting in a final adjusted R-squared of 0.087 ($F(11, 2468) = 20.85, p < 0.001$). The pathway coefficients showed notable increases. The coefficient for IELTS Low increased to 3.944 credits ($p < 0.001$) from 3.45 in the previous model, while the coefficient for TOEFL low increased from 1.3 to 2.05 and became statistically significant ($p < 0.05$) for the first and only time across all models. Similarly, the coefficient for IELTS Mid increased from .20 to .60 and became statistically significant ($p < 0.001$) also for the first and only time across all models. The coefficient for IELTS High

increased to 1.931 credits ($p < 0.001$) from 1.51. The coefficient for TOEFL High increased to 1.894 credits and became statistically significant ($p < 0.05$), which was not the case since model 2. Age emerged as a significant predictor, with each additional year of age associated with earning 0.241 fewer credits ($p < 0.001$). This indicates that older students tended to earn fewer credits during their first year. The coefficient for gender remained not statistically significant. The effect of country of origin remained significant, with students from other countries earning 0.865 more credits than students from China ($p < 0.001$). The coefficient for the STEM discipline decreased slightly to 0.834 credits but remained significant ($p < 0.001$). The coefficient for Business increased to 0.557 credits but was still not statistically significant.

Table 4.18
Regression Analysis Estimates (and SE) for CCE for University B

	Model 1	Model 2	Model 3	Model 4	Model 5
EAP	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)
1.IELTS LOW	2.822***	3.074***	3.439***	3.451***	3.944***
	(0.287)	(0.288)	(0.293)	(0.293)	(0.313)
2. TOEFL LOW	1.188	1.060	1.326	1.335	2.051*
	(0.910)	(0.902)	(0.898)	(0.898)	(0.909)
3. IELTS MID	0.0868	0.00399	0.200	0.204	0.631***
	(0.164)	(0.163)	(0.166)	(0.166)	(0.192)
4. TOEFL MID	0.717*	0.622	0.587	0.602	1.025**
	(0.353)	(0.351)	(0.348)	(0.348)	(0.360)
5. IELTS HIGH	2.179***	1.941***	1.501***	1.512***	1.931***
	(0.364)	(0.362)	(0.369)	(0.369)	(0.379)
6. TOEFL HIGH	1.936*	1.840*	1.458	1.475	1.894*
	(0.762)	(0.756)	(0.754)	(0.754)	(0.757)
Social Sci. + Hum.		0	0	0	0
		(.)	(.)	(.)	(.)
1.Business		0.0328	-0.00852	0.0223	0.557
		(0.489)	(0.486)	(0.487)	(0.500)
2.STEM		1.116***	0.929***	0.881***	0.834***
		(0.171)	(0.173)	(0.176)	(0.176)
Origin			0.944***	0.913***	0.865***
			(0.167)	(0.168)	(0.168)
Gender				-0.223	-0.236
				(0.150)	(0.150)
Age					-0.241***
					(0.0546)
Constant	12.81***	12.03***	11.74***	11.88***	16.63***
	(0.128)	(0.178)	(0.184)	(0.205)	(1.095)
Adjusted R-squared	0.052	0.068	0.080	0.080	0.087
N	2480	2480	2480	2480	2480

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In summary, the regression analyses for both University A and University B revealed significant associations between pathways and cumulative credits earned during the first year of study. At University A, all ELP pathway categories had significantly higher credits earned compared to the EAP group, with the IELTS High and IELTS Mid groups exhibiting the largest differences. Variables such as academic discipline, country of origin, gender, and age also played significant roles in predicting credits earned. Specifically, students in the Business and STEM disciplines earned fewer credits than those in Social Sciences and Humanities. Female and younger students generally earned more credits, and students from countries other than China also earned more.

At University B, significant positive effects were observed primarily in the IELTS Low and IELTS High proficiency groups, with the IELTS Low group showing the most substantial gains in credits earned. The STEM discipline emerged as a significant positive predictor of credits earned, indicating that students in STEM fields earned more credits than those in Social Sciences and Humanities. Age was a significant predictor, with younger students earning more credits. Unlike University A, gender did not significantly impact credits earned at University B, suggesting that men and women accumulated credits at similar rates. Additionally, students from countries other than China earned more credits, mirroring the findings at University A.

Similarly to the previous section on CGPA, to further investigate these relationships and gain a deeper understanding of how different ELP pathways influence the number of credits earned, the next section presents pairwise comparisons through analyses of covariance (ANCOVA).

Analysis of Covariance

This section examines the academic achievement, operationalized by cumulative credits earned, of international students from various language proficiency pathways by comparing their adjusted mean scores. The primary focus is to present how students in different language testing pathways (TOEFL, IELTS) and proficiency levels (Low, Mid, High) perform compared to students in the EAP pathway.

Comparative analysis between IELTS and TOEFL pathways

At University A, most pairwise comparisons were statistically significant and showed that IELTS students earned more credits than TOEFL students, particularly at mid and high proficiency levels. For example, IELTS Low students surpassed TOEFL Mid (contrast = -0.330, 95% CI=[-0.621, -0.038]), while IELTS Mid exceeded both TOEFL Low (contrast=0.796 [0.426, 1.166]) and TOEFL Mid (contrast=-1.119 [-1.388, -0.851]). IELTS High likewise outperformed TOEFL Low (contrast=1.233 [0.642, 1.824]) and TOEFL Mid (1.556 [1.023, 2.090]), though no significant difference emerged between IELTS High and TOEFL High. Overall, these findings suggest that IELTS proficiency may be a stronger indicator of academic progression at University A, especially for mid- and high-level English speakers.

Table 4.19

Pairwise Comparisons between IELTS and TOEFL Pathways at University A for Cumulative Credits Earned

Comparison	Contrast	Std. Error	95% Confidence Interval	Significance
TOEFL Low vs. IELTS Low	0.007	0.199	[-0.396, 0.383]	Not Significant
TOEFL Mid vs. IELTS Low	0.330	0.149	[-0.621, -0.038]	$p < 0.05$
IELTS Mid vs. TOEFL Low	0.796	0.189	[0.426, 1.166]	$p < 0.001$
IELTS Mid vs. TOEFL Mid	1.119	0.137	[-1.388, -0.851]	$p < 0.001$
IELTS High vs. TOEFL Low	1.233	0.301	[0.642, 1.824]	$p < 0.001$
IELTS High vs. TOEFL Mid	1.556	0.272	[1.023, 2.090]	$p < 0.001$
TOEFL High vs. IELTS High	0.805	0.458	[-1.703, 0.093]	Not Significant

At University B, most IELTS and TOEFL pathway comparisons showed no significant differences. An exception was TOEFL Mid students earning significantly fewer credits than IELTS Low students (contrast = -3.006 [SE = 0.418, 95% CI = [-3.825, -2.186], $p < 0.001$]). This finding was contrary to the typical expectation that higher proficiency levels correlate with better academic outcomes.

Table 4.20

Pairwise Comparisons between IELTS and TOEFL Pathways at University B for Cumulative Credits Earned

Comparison	Contrast	Std. Error	95% Confidence Interval	Significance
TOEFL Low vs. IELTS Low	-1.624	0.932	[-3.452, 0.204]	Not Significant
TOEFL Mid vs. IELTS Low	-3.006	0.418	[-3.825, -2.186]	$p < 0.001$
IELTS Mid vs. TOEFL Low	-1.765	0.906	[-3.542, 0.011]	Not Significant
IELTS Mid vs. TOEFL Mid	0.384	0.340	[-0.284, 1.051]	Not Significant
IELTS High vs. TOEFL Low	-0.475	0.969	[-2.375, 1.424]	Not Significant
IELTS High vs. TOEFL Mid	0.906	0.470	[-0.016, 1.828]	Not Significant
TOEFL High vs. IELTS High	-0.108	0.811	[-1.698, 1.483]	Not Significant

Comparative analysis within IELTS and TOEFL pathways

At University A, for students admitted through the IELTS pathway, the Mid group earned significantly more credits than the Low group (contrast=0.789 [SE=0.121, 95% CI=0.552–1.026], $p < 0.001$), and the High group significantly outperformed the Low group (contrast=1.226 [SE=0.269, 95% CI=0.699–1.754], $p < 0.001$). The remaining comparisons were found to not be statistically significant. For students who were admitted through the TOEFL pathway, all comparisons were found to not be statistically significant. The following section presents results for the same analyses conducted for University B.

Table 4.21

Pairwise Comparisons within IELTS and TOEFL Pathways at University A for Cumulative Credits Earned

Comparison	Contrast	Std. Error	95% Confidence Interval	Significance
IELTS Mid vs. IELTS Low	0.789	0.121	[0.552, 1.026]	$p < 0.001$
IELTS High vs. IELTS Low	1.226	0.269	[0.699, 1.754]	$p < 0.001$
IELTS High vs. IELTS Mid	0.437	0.254	[-0.062, 0.935]	Not Significant
TOEFL Mid vs. TOEFL Low	-0.323	0.210	[-0.734, 0.088]	Not Significant
TOEFL High vs. TOEFL Low	0.428	0.426	[-0.406, 1.262]	Not Significant
TOEFL High vs. TOEFL Mid	0.751	0.404	[-0.042, 1.544]	Not Significant

At University B, IELTS Mid students earned about 3.389 fewer credits than IELTS Low students (contrast=-3.389 [SE=0.279, 95% CI=-3.936 to -2.843], $p < 0.001$), and IELTS High students also trailed Low students by 2.099 credits (contrast=-2.099 [SE=0.448, 95% CI=-2.978 to -1.220], $p < 0.001$). These results are unexpected, as one would anticipate higher proficiency students earning more credits. However, IELTS High students outperformed Mid students by 1.290 credits (contrast=1.290 [SE=0.366, 95% CI=0.572 to 2.009], $p < 0.001$). Among TOEFL students, no proficiency-level comparisons were statistically significant.

Table 4.22

Pairwise Comparisons within IELTS and TOEFL Pathways at University B for Cumulative Credits Earned

Comparison	Contrast	Std. Error	95% Confidence Interval	Significance
IELTS Mid vs. IELTS Low	-3.389	0.279	[-3.936, -2.843]	p < 0.001
IELTS High vs. IELTS Low	-2.099	0.448	[-2.978, -1.220]	p < 0.001
IELTS High vs. IELTS Mid	1.290	0.366	[0.572, 2.009]	p < 0.001
TOEFL Mid vs. TOEFL Low	-1.382	0.957	[-3.258, 0.495]	Not Significant
TOEFL High vs. TOEFL Low	-0.583	1.171	[-2.880, 1.714]	Not Significant
TOEFL High vs. TOEFL Mid	0.799	0.809	[-0.788, 2.385]	Not Significant

At University A, higher IELTS proficiency levels were associated with increased cumulative credits earned, highlighting a positive relationship between higher IELTS performance and credits earned. Conversely, proficiency levels within the TOEFL pathway did not have a significant impact on credits earned. At University B, an unexpected pattern emerged where IELTS Low proficiency students earned more credits than their higher proficiency counterparts, and no significant differences were observed within the TOEFL pathway.

Interaction Effects

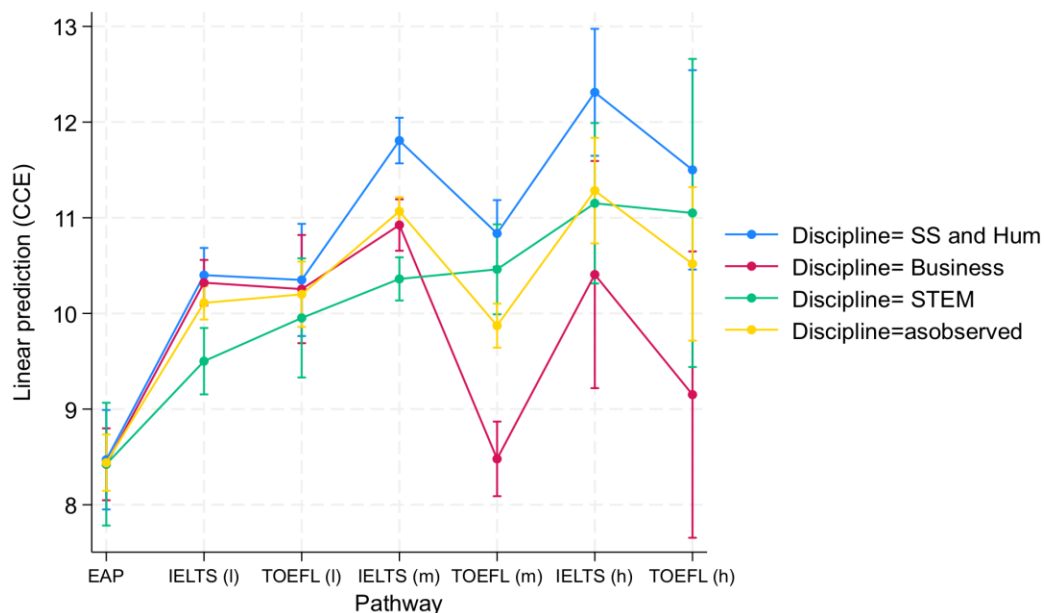
As with the previous section on CGPA, this section incorporated interaction terms into the statistical models to explore whether the impact of English language proficiency on academic achievement, operationalized by cumulative credits earned, varied across different

student subgroups and institutional contexts. All findings presented below are statistically significant at $p < 0.001$ and are visually represented. All interaction effects are presented in Appendix G to L.

Interaction Effects between English Language Proficiency Pathways and Academic Discipline

The full results of the interaction effects between ELP pathways and discipline for both universities are reported in Appendices G and H. At University A, cumulative credits earned varied by English language proficiency pathway and discipline. Compared to EAP entrants (who earned about 8.44 credits on average), students admitted via higher proficiency pathways earned significantly more credits across discipline. For example, in Social Sciences and Humanities, IELTS Low students earned approximately 10.40 credits—an increase of about +1.93 credits over EAP - while TOEFL Low students in the same discipline earned about 10.35 credits, an improvement of +1.88 credits. Similarly, in Business, TOEFL Low students earned 10.25 credits, which was found to be +1.78 credits higher than EAP. Among the higher proficiency groups, IELTS Mid entrants, in Social Sciences and Humanities, earned 11.81 credits, or +3.34 credits relative to EAP and IELTS High students earned 12.31 credits - an increase of +3.84 credits. In addition, TOEFL Mid students demonstrated significant advantages in Social Sciences and Humanities, where students earned 9.87 credits, and in STEM 10.46 credits which represented +2.37 and +1.99 credits compared to EAP. Finally, TOEFL High entrants in Social Sciences and Humanities earned 11.50 credits, a significant improvement of +3.03 credits compared to EAP. Overall, these results underscore that higher proficiency, whether via IELTS or TOEFL, is associated with an increase in cumulative credits earned relative to EAP across multiple disciplines.

Figure 4.7. Predictive margins with 95% CIs - interaction effects between pathway and discipline (University A)⁴

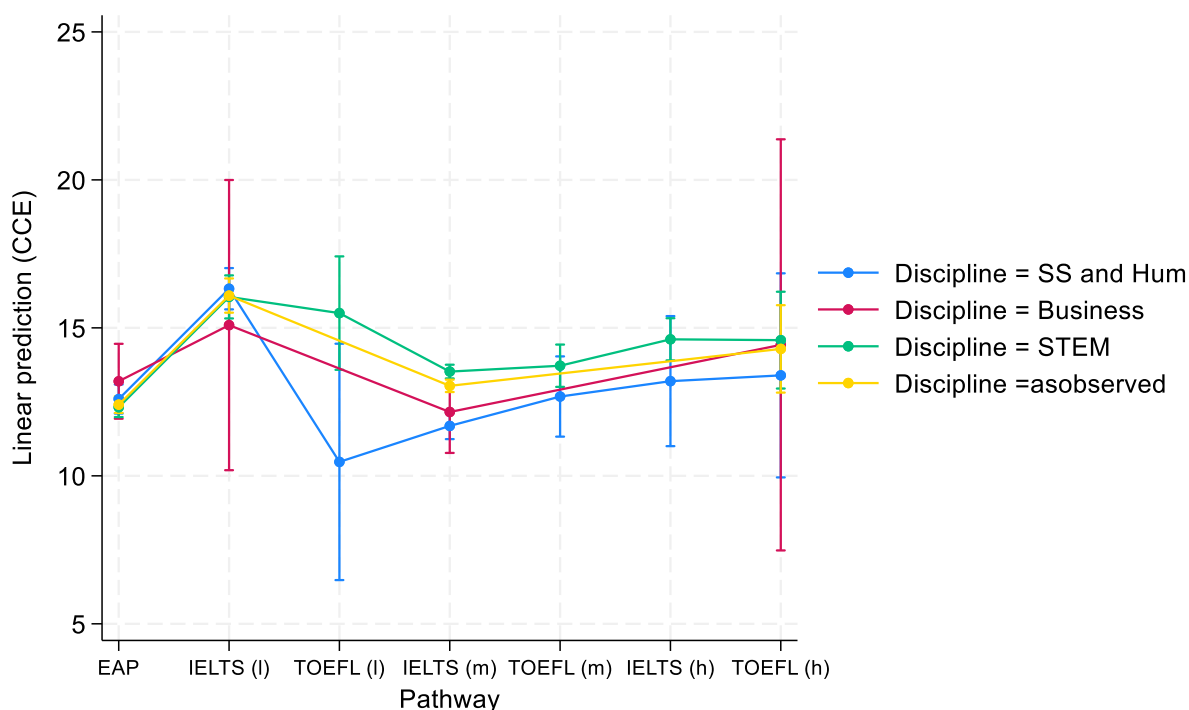


At University B, cumulative credits earned varied substantially by English language proficiency pathway and discipline. For instance, EAP students in Social Sciences and Humanities earned 12.59 credits on average, whereas IELTS Low students in the same discipline earned 16.32 credits—an increase of +3.73 credits. Similarly, in STEM, IELTS Low students earned 16.05 credits compared to 12.32 credits for EAP students, reflecting a significant gain of around +3.45 credits. In addition, significant differences were observed for IELTS High students. For example, those in STEM earned 14.61 credits, which was significantly higher than the EAP benchmark. In contrast, although TOEFL High students achieved an overall predicted margin of 14.29 credits—approximately +1.88 credits higher than EAP—the pairwise comparisons across individual disciplines (e.g., in STEM, a difference of +1.99 credits) did not reach statistical

⁴ The 'as-observed' line represents how the predicted outcomes behave in comparison to the actual data.

significance. Overall, these results indicate that while higher proficiency via IELTS pathways is robustly associated with increased cumulative credits relative to EAP, the advantage for TOEFL entrants is less pronounced and less statistically reliable across disciplines.

Figure 4.8. *Predictive margins with 95% CIs - interaction effects between pathway and discipline (University B)*

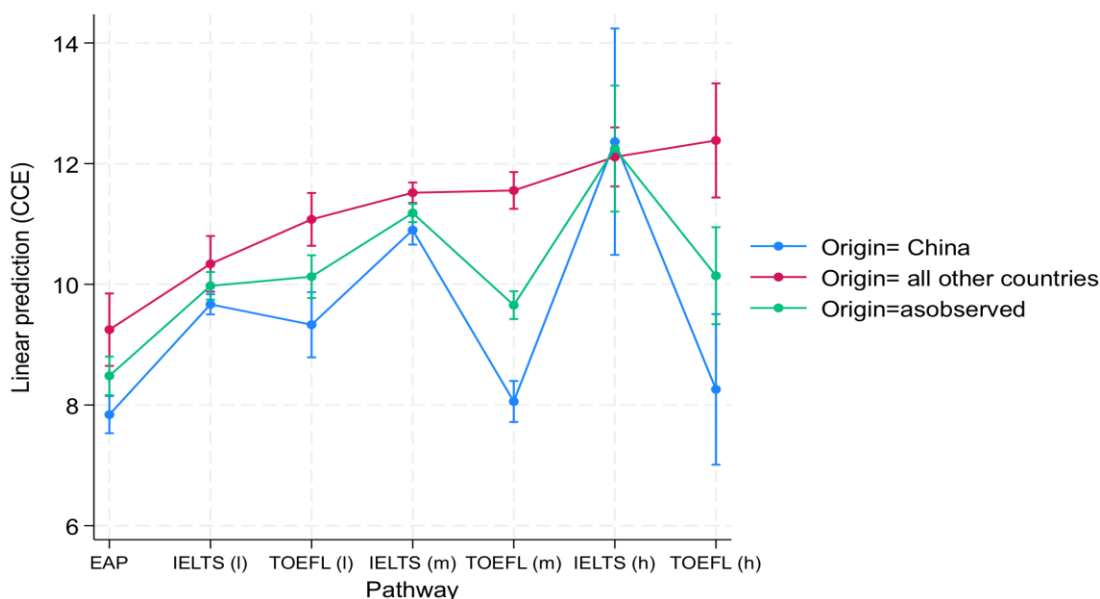


Interaction Effects between English Language Proficiency Pathways and Country of Origin

The full results of the interaction effects between ELP pathways and discipline for both universities are reported in Appendices I and J. At University A, across all pathways, students from other countries generally earned more credits than Chinese students across all pathways. For instance, among EAP students, non-Chinese students earned an average of 9.25 credits compared to 7.84 credits for Chinese students—a difference of +1.41 credits. The largest disparities were observed in TOEFL mid where students from other countries earned 11.56

credits versus 8.06 for Chinese students (a difference of +3.72 credits), and in TOEFL high, the gap was even wider at approximately +4.13 credits (12.38 vs. 8.26 credits). Notably, in the IELTS High group, the difference between non-Chinese and Chinese students was negligible and not statistically significant. Overall, these results suggest that—with the exception of the highest IELTS group—students from other countries consistently accumulate more cumulative credits than Chinese students across ELP pathways.

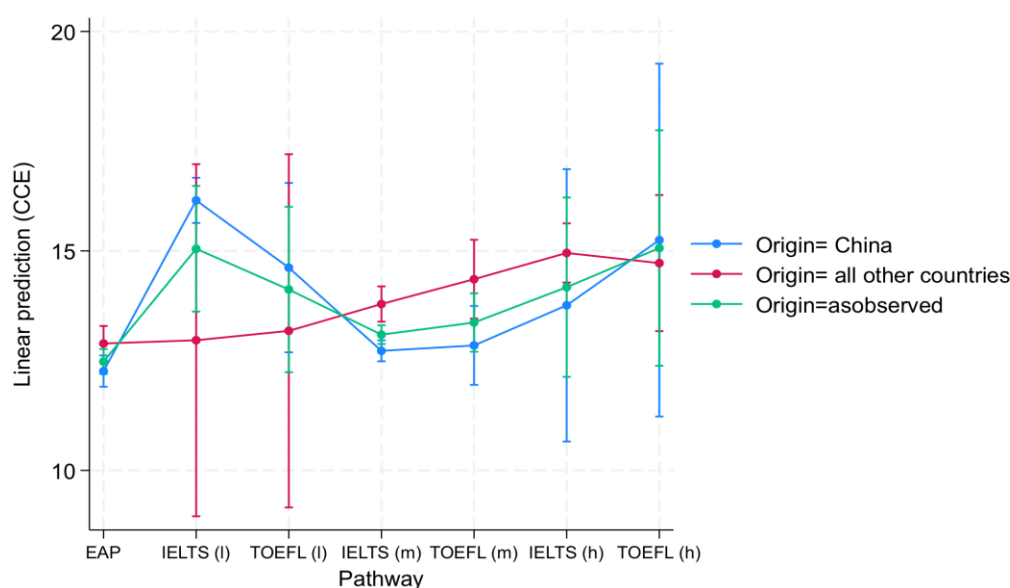
Figure 4.10. *Predictive margins with 95% CIs - interaction effects between pathway and origin (University A)*



At University B, the overall effect of origin shows that non-Chinese students (13.52 credits) tend to earn slightly more than Chinese students (12.94 credits). However, when examining the interaction between ELP pathway and origin, most differences between Chinese and non-Chinese students were not statistically significant. Some notable exceptions were Chinese students in the IELTS Low pathway compared to Chinese EAP students where a significant difference of approximately +3.89 credits was found. Similarly, Chinese students in IELTS High earned +2.09 credits more than Chinese EAP students.

In addition, when directly comparing Chinese students across pathways, those in the IELTS Mid group earned significantly fewer credits than those in the IELTS Low group with a difference of -3.43 credits. Also, within the IELTS Mid pathway, non-Chinese students earned approximately 1.07 credits more than their Chinese peers.

Figure 4.11. *Predictive margins with 95% CIs - interaction effects between pathway and origin (University B)*

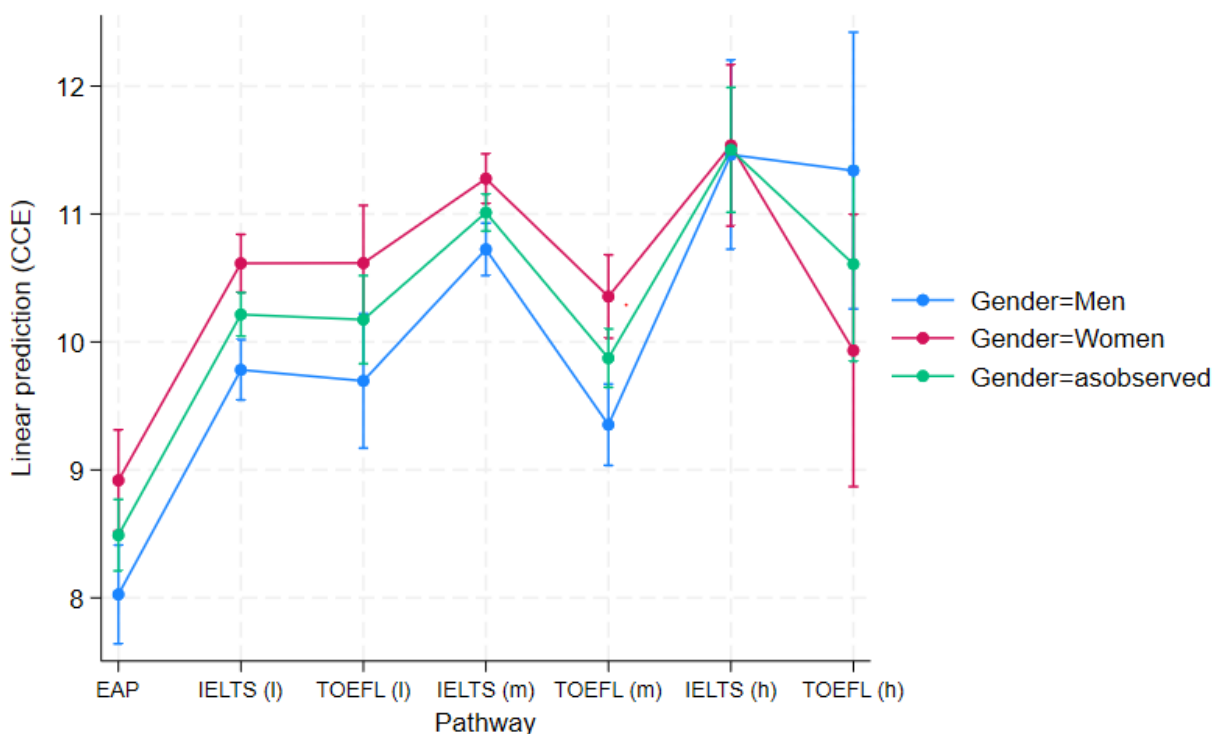


Interaction Effects between English Language Proficiency Pathways and Gender

The full results of the interaction effects between ELP pathways and discipline for both universities are reported in Appendices K and L. At University A, the interaction between ELP pathway and gender in cumulative credits earned shows that female students consistently outperform male students within the same proficiency pathway. For example, in the IELTS Low group, women earned an average of 10.614 credits compared to 9.782 credits for men, a significant difference of 0.83 credits. In the IELTS Mid group, women earned 11.28 credits versus 10.72 credits for men—a difference of 0.552 credits. Similarly, in the IELTS High group,

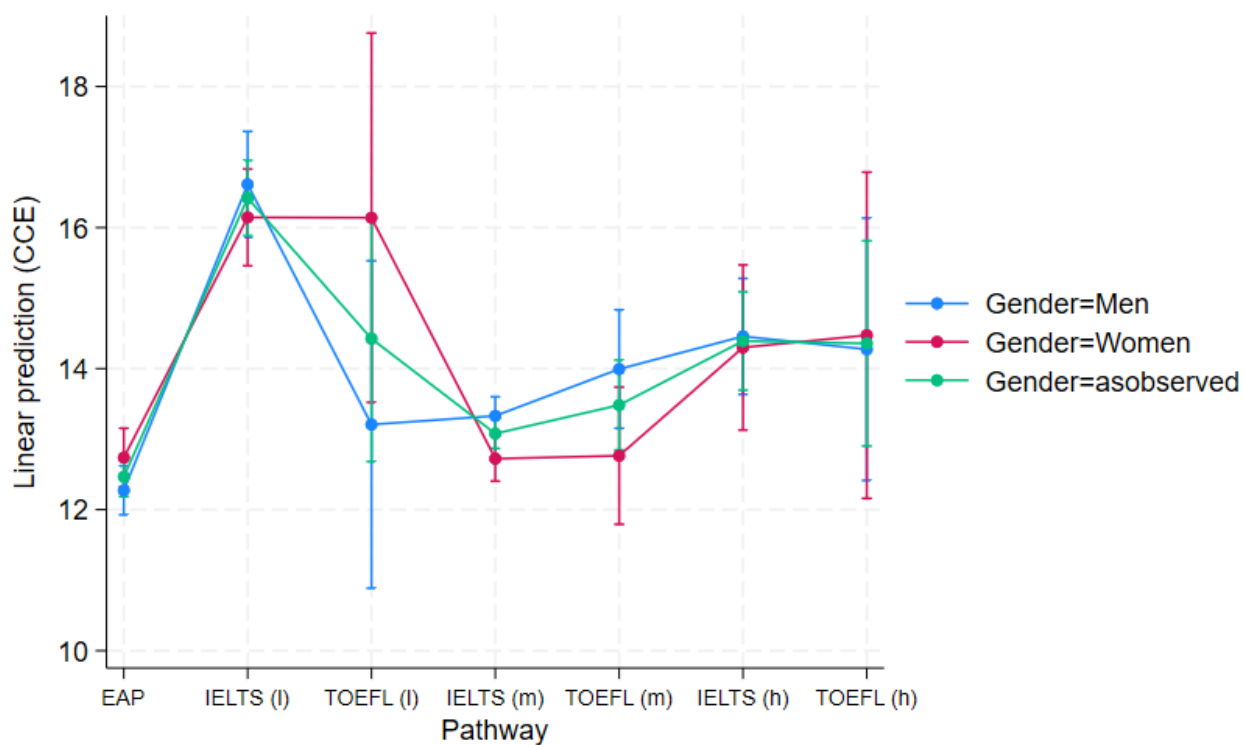
the contrast between women and men revealed that women earned 1.002 credits more than men. These findings indicate that, across multiple ELP pathways, women accumulate more credits than men students.

Figure 4.12. *Predictive margins with 95% CIs - interaction effects between pathway and gender (University A)*



At University B, gender advantage shifted across pathways and proficiency levels – although many comparisons were not statistically significant. Notable findings were that women led at EAP (12.74 vs. 12.27 for men) and in the IELTS Low pathway where men earned 9.78 credits while women earned 10.61 credits—a difference of 0.83 credits. Similarly, in the IELTS Mid group, women outperformed men by 0.55 credits, and in the IELTS High pathway, women earned roughly 1.00 credit more than men. Significantly, men in the IELTS Low pathway earned 4.34 credits more than men in EAP.

Figure 4.13. Predictive margins with 95% CIs - interaction effects between pathway and gender
(University B)



In the next chapter, these results will be interpreted in relation to the research questions and the literature.

Chapter 5: Discussion

This chapter synthesizes and interprets the findings from the previous chapter, placing them in the context of the study's research questions and the broader literature. The discussion is organized to

1. synthesize the main findings, highlighting patterns and trends observed across the two universities and the two measures of academic achievement (cumulative GPA and cumulative credits earned),
2. discuss implications of these findings for university policy
3. note limitations of this study; and
4. provide recommendations for future research

In doing so, this chapter addresses the overarching research question of this dissertation:

How do international undergraduate students who are required to provide evidence of ELP and are admitted to university through different pathways compare in terms of their academic achievement at the end of their first year of study?

Synthesis of Main Findings

The findings of this study provide a comprehensive view of how English language proficiency (ELP) pathways—namely, the English for Academic Purposes (EAP) pathway and the standardized test pathways (e.g., IELTS and TOEFL)—relate to academic achievement as operationalized by both cumulative grade point average (CGPA) and cumulative credits earned (CCE) at the end of the first year of study. Drawing from descriptive statistics, multiple regression models, analysis of covariance (ANCOVA), and interaction effects, this section synthesizes the most salient outcomes across both University A and University B.

Standardized Tests vs. EAP Pathways: Key Findings

In the literature on international student success in English-medium universities, there is a significant lack of research that directly compares academic outcomes of standardized ELP tests (e.g., IELTS and TOEFL) to alternative or course-based measures of proficiency. The limited amount of research in this field has called into question the equivalency of such different measures of proficiency through findings that have favoured the test-based route (e.g. Clark et al., 2021; Pearson, 2020; Schoepp, 2018; Tweedie and Chu, 2019). Conversely, an even more limited amount of research has presented conflicting findings and suggested that course-based options (typically EAP) may have a better association with academic achievement (e.g., Johnson & Tweedie, 2021). Given the lack of definitive conclusions and conflicting evidence in the literature, the expectation for this study was that data would indicate either a clear advantage for standardized test pathways over EAP or a more nuanced, context-dependent relationship.

While analyses confirmed that international students admitted via either IELTS or TOEFL tended to outperform those admitted through English for Academic Purposes pathways at both universities, modelling revealed significant nuance to this key finding. At both universities, students who scored in the highest proficiency bands on standardized tests (IELTS High and TOEFL High) achieved significantly higher cumulative GPAs and earned more cumulative credits than students admitted through the EAP pathway. However, it is important to recognize that comparing EAP students to high-scoring test takers (for example, those scoring IELTS 8.5) has theoretical and practical flaws, as the purpose of the EAP pathway is not to match the performance of students who already significantly exceed minimum proficiency requirements but rather to bring students up to that baseline level. Therefore, while findings indicated that test-based routes generally yielded better academic outcomes, the evaluation of the EAP pathway

should focus on whether these students achieved outcomes comparable to those at or close to the cut-score level, not the high-proficiency end of the spectrum.

This being the case, in terms of CGPA outcomes, only the IELTS low and mid proficiency groups outperformed EAP significantly (.40 and .58, respectively, $p < 0.001$) at University A in the final model. While both TOEFL low and mid proficiency pathways initially outperformed the EAP pathway in models 1-4, the final model did not retain statistical significance. This finding was not replicated at University B whereas IELTS and TOEFL low and mid proficiency levels significantly outperformed EAP in most models at University A, only the TOEFL mid proficiency was found to have a significant difference at University B in comparison to EAP, and with a much smaller p value ($p < 0.05$). Furthermore, while R^2 values were small at both institutions, results did vary significantly. The initial models which simply examined pathway as a predictor of CGPA found an adjusted R^2 of 3% for University A and only .8% at University B. In the final model, which introduced all the moderator variables of this dissertation (gender, age, etc.), R^2 values improved to 6% for University A and only to 1% for University B.

These findings can be explained by several factors adding a significant amount of complexity to the limited literature on this topic. First, while previous research has shown that standardized ELP test scores (e.g., IELTS and TOEFL) generally predict stronger academic outcomes compared to alternative, course-based measures (Clark et al., 2021; Pearson, 2020; Schoepp, 2018; Tweedie & Chu, 2019), findings from this dissertation regarding the relationship between academic achievement and CGPA indicate that the advantage of test-based pathways is nuanced. For example, at University A, only the IELTS low and mid groups maintained significant positive differences in CGPA relative to EAP, whereas the TOEFL low and mid groups lost significance in the final model after accounting for moderator variables. This

suggests that while standardized test scores do capture aspects of academic readiness, the predictive power may vary by test type and be sensitive to the inclusion of factors such as discipline, country of origin, gender, and age. Additionally, at University B, the relationship between test-based pathways and academic achievement was less robust overall, as evidenced by much lower adjusted R^2 values (rising only from 0.8% to 1% in the final model), which could be due to differences in institutional ranking, and the dominance of STEM disciplines at this institution. These institutions likely attract different cohorts with varying levels of preparedness and support structures, which in turn influence how effectively ELP scores predict performance. Moreover, the literature on international student success (e.g., Johnson & Tweedie, 2021) emphasizes that the equivalency of different ELP measures is not absolute; comparing EAP students to high-scoring test-takers is inherently problematic since the goal of EAP is to raise students to the minimum proficiency required for academic success rather than to match the performance of those who already exceed that threshold. Thus, these findings reinforce that while high standardized test scores are generally associated with better academic outcomes, the strength of this association is moderated by local contextual factors and student demographics. Understanding these nuances is essential, as it suggests that institutions should consider a more tailored approach to admissions and support, rather than assuming that all ELP pathways are equivalent in predicting long-term academic achievement.

Similarly to the CGPA findings, students admitted through the EAP pathway at both institutions consistently earned fewer credits in their first year compared to those who met language proficiency requirements via standardized tests. Notably, while CGPA differences varied by proficiency level, the differences in cumulative credits earned remained statistically significant across all proficiency bands—from low to high—in the final models. This pattern

indicates that, irrespective of the specific level of proficiency achieved on standardized tests, EAP students typically underperformed in terms of total credits earned when compared to their peers meeting baseline or below baseline admission requirements. While University A followed the expected trend of higher proficiency levels earning more credits than lower proficiency levels in comparison to EAP, results at University B demonstrated that the lowest proficiency levels earned more credits compared to EAP than the highest proficiency levels compared to EAP. In other words, at University B, students with lower standardized test scores appear to benefit more—in terms of credits earned—from the test-based admission route relative to the EAP pathway, while the relative advantage of high proficiency becomes less pronounced when accounting for moderator variables. Pathway alone was found to predict more of the variance in cumulative credits earned than CGPA, as both institutions reported an R^2 of ~5% in the base model and with an increase to 9% found at both institutions.

The findings regarding cumulative credits earned may be explained by several factors inherent in both the nature of EAP programs and the institutional contexts of the universities studied. The EAP pathway is designed primarily to raise students to a minimum proficiency level, rather than to produce high academic performers. As such, students admitted through EAP may receive sufficient language training to meet entry requirements but may not acquire the full range of academic language and study skills that standardized tests capture. This could contribute to their consistently lower accumulation of credits compared to students admitted through test-based routes, regardless of whether those test scores fall in the low, mid, or high proficiency bands. Furthermore, the differential patterns observed between University A and University B likely reflect variations in institutional characteristics and student populations. At University A, where there is a balanced distribution of disciplines, the expected trend is evident: higher

proficiency scores correspond to greater credit accumulation, indicating that the test-based route effectively signals readiness across academic fields. In contrast, at University B—a highly ranked institution with a stronger focus on STEM—the finding that students with lower standardized test scores earn more credits in comparison to higher standardized test scores relative to EAP students suggests that the student body may possess additional academic strengths or that local practices in credit allocation and support services are different. These nuances could mean that at University B, even students with lower test scores benefit from the structure of test-based admissions, potentially due to stronger academic backgrounds. Finally, the fact that pathway alone predicted a higher proportion of the variance in cumulative credits earned relative to CGPA (with R^2 values rising from around 5% in the base model to 9% in the final model) underscores that the type of admission pathway is a more significant determinant of academic progression than it is compared to CGPA. This suggests that standardized tests might be capturing elements of academic readiness more comprehensively than EAP measures.

The variation across universities in the magnitude of the difference in academic achievement between the EAP students and those admitted through proficiency tests could reflect variations in the delivery and effectiveness of EAP programs across the two universities. For example, the level structures of the EAP programs at the two institutions are significantly different. Whereas University A's EAP program consists of nine courses ranging from beginner to advanced (levels 1-9), and students' level entry is based on internal placement tests or ELP test scores (e.g., IELTS or TOEFL), University B's program consists of 3 levels with the second level consisting of a full term of language instruction (260 hours). Furthermore, University A offers a blended (on campus and online) delivery format, while University B follows a more traditional in person classroom only instruction model. A thorough curricula analysis was not the

focus of this dissertation; therefore, it cannot be ascertained how these differences in programming affect outcomes. However, it is possible that these programmatic distinctions played a role in the smaller effect sizes observed at University B. In other words, variations in level structure (e.g., nine levels versus three), instructional hours, and delivery format (blended versus face-to-face) may create unique learning conditions that influence students' development and learning outcomes. A more detailed curricular analysis—beyond the scope of this dissertation—would be necessary to determine precisely how these differing program features shape learner progress and achievement. Such nuances echo prior observations (e.g., Schoepp, 2018) that institutional and program-specific factors can moderate the strength of the relationship between ELP and academic achievement.

Overall, these results are meaningful because they deepen our understanding of how different English language proficiency pathways—particularly standardized tests versus EAP-based coursework—relate to academic achievement for international students. Given that the existing literature presents conflicting evidence and has struggled to clearly establish whether EAP can match or surpass the predictive power of test-based pathways (e.g. Clark et al., 2021; Johnson & Tweedie, 2021; Tweedie & Chu, 2019), this dissertation finds that while higher proficiency levels consistently outperform EAP, comparisons at the baseline admission level (and below) vary in significance. This finding underscores how crucial it is for universities and policymakers to examine the practical outcomes of each pathway rather than relying solely on theoretical or anecdotal equivalencies across the sector.

From an institutional perspective, these findings highlight a potential need to revisit the design and rigor of EAP programs. If students who meet admissions requirements via EAP earn lower GPAs and fewer credits than those with TOEFL or IELTS scores assumed to be at the

same proficiency level, it suggests that EAP curricula may need further integration with discipline-specific language demands or additional scaffolding for academic writing and research skills. Moreover, the differences in effect sizes between University A and University B could be partially explained by the type of student population each institution attracts, given their distinct rankings and disciplinary profiles. University A, with a top 401–500 global rank, features a relatively balanced distribution of students across Business, Social Sciences, and STEM, whereas University B is ranked in the top 100 and exhibits a stronger focus on STEM subjects. This difference in rank and discipline composition may create varying levels of academic rigor or test-score competitiveness among admitted cohorts, thus amplifying or reducing the observed achievement gap between EAP and test-scoring students. In contexts like University A, where standardized test scores more strongly predict academic outcomes, refinements to EAP programs could help close the achievement gap and better equip EAP-admitted students for the rigors of first-year study. An important strength and contribution of this research lies in its multi-site comparative design. Previous studies, limited predominantly to single-site investigations, struggled with external validity. By contrast, this dissertation directly compares outcomes across two distinct university settings, thereby significantly enhancing both the validity and reliability of the findings. Such a methodological approach has demonstrated more comprehensive insights into the variability of student outcomes across institutional contexts.

Finally, these findings inform the broader literature and discourse on international student success in a globalized higher education landscape. By confirming that EAP-based pathways produce lower first-year academic outcomes than standardized test routes, this dissertation provides empirical grounding for arguments that EAP and test-based measures should not be treated as perfectly interchangeable or equivalent. Such evidence is crucial for students deciding

between pathways and for institutions deciding how to balance English admission options with ensuring academic preparedness. Ultimately, recognizing the gap between EAP and students admitted through test scores can prompt targeted interventions for students admitted through the EAP pathway—and in the long term, create a more supportive environment for international students' academic achievement.

Reevaluating IELTS–TOEFL Equivalencies

Consistent with previous research (e.g., Cho & Bridgeman, 2012; Dang & Dang, 2023; Feast; 2002; Hill et al., 1999; Ruegg et al., 2024; Schoepp, 2018; Wongtrirat, 2010; Woodrow, 2006), many post-hoc comparisons in this dissertation revealed that students in the high-proficiency groups (i.e., IELTS High and TOEFL High) outperformed their peers in the lower proficiency categories. For instance, at University A, both IELTS high and TOEFL high outperformed their IELTS and TOEFL mid and low peers in terms of CGPA. This trend also continued for IELTS for credits earned at University A, but did not result in statistically significant differences for TOEFL. At University B, an unexpected result emerged where IELTS Low students earned more credits than their higher proficiency counterparts, suggesting that factors beyond language proficiency—such as institutional course load expectations, disciplinary differences in credit distribution, or variations in academic support services—may be influencing credit accumulation. Despite this anomaly, within test comparisons suggested that these differences, although often modest in magnitude, are generally consistent with previous studies comparing IELTS and TOEFL across a wide range of study designs, disciplines, and institutional settings (Abunawas, 2014; Gagen, 2019; Wongtrirat, 2010).

Between test comparisons further highlighted the importance of understanding local contexts. At University A, IELTS students tended to perform better than TOEFL students at

comparable proficiency levels, while at University B, the trend was reversed or statistically insignificant. Such variability underscores that treating IELTS and TOEFL as universally equivalent, as often happens in admissions (Davies et al., 1999; Schoepp, 2018), is likely not pertinent. As some evidence has suggested, there are subtle differences in how each test correlates with course outcomes (e.g., Ihlenfeldt & Rios, 2023; Johnson & Tweedie, 2021); therefore, equivalency is questionable. These results are important precisely because they demonstrate that simple equivalencies in English language proficiency test scores do not always translate into uniform academic outcomes across different institutional and demographic contexts. Rather than relying on a universal assumption that IELTS and TOEFL can be treated interchangeably, the analysis of this dissertation shows that various demographic and contextual factors, including institutional context can significantly alter the extent to which higher language proficiency confers an academic advantage.

Moreover, by revealing that high-versus low-level proficiency is not a simple predictor of academic success in every scenario, the findings encourage more nuanced admissions and placement practices. For instance, at University A, IELTS mid-level test takers surpass TOEFL mid-level students in terms of CGPA and CCE, but at University B this comparison is inversed. This suggests that language proficiency alone does not capture the entirety of what determines a student's progression toward their degree and factors such as student demographics (e.g., discipline of study, student profile) and institutional context (e.g., ranking, teaching vs research institution) may shape how mid-level test scores relate to academic achievement. Consequently, rather than relying on universal cut-scores or simple test-score equivalencies, universities should conduct their own standard setting studies to establish cut scores for admissions. This approach ensures that language proficiency pathways are interpreted within the broader institutional

context, ultimately leading to more effective outcomes for diverse cohorts of international students. Ultimately, this more refined perspective enables higher education institutions to develop policies and interventions that truly align with students' varied needs, thereby enhancing both short-term academic performance and long-term educational success.

The Moderating Roles of Discipline, Origin, Gender, and Age

In addition to exploring the relationship between pathway and academic achievement, this dissertation examined how the impact of English language proficiency pathways on CGPA and cumulative credits earned are moderated by key student variables—discipline of study, country of origin, gender, and age. In the literature, researchers have consistently demonstrated that discipline and demographic factors are important predictors to include in predictive models of academic success, as they help explain variability in outcomes (e.g., Chen et al., 2020; Hellas et al., 2018; Lehmann & Meldrum, 2021; Gipson, 2018; Strayhorn, 2006; Van Herpen et al., 2017).

The results of this dissertation, demonstrated both in regression outputs and interaction effect analyses, have reinforced that the relationship between English language proficiency pathways and academic achievement is intricately moderated by discipline, country of origin, gender, and age. For example, analyses indicated that while higher ELP scores are generally associated with improved CGPA and increased credits earned, the degree of improvement varies significantly across academic disciplines—with students in STEM and Social Sciences & Humanities often outperforming those in Business. This finding is consistent with related research that documents how the predictive validity of standardized tests can differ by various moderator variables such as discipline of study (Cho & Bridgeman, 2012; Hu & Trenkic, 2019; Light et al., 1987).

Similarly, results regarding country of origin align with prior studies (e.g., Chen et al., 2020; Lehmann & Meldrum, 2021) that emphasize the moderating influence of cultural and educational background. In this dissertation, non-Chinese students frequently earned more credits and achieved higher CGPAs than their Chinese counterparts in several proficiency bands, although some oscillations were found and some comparisons were less pronounced at the highest proficiency levels (e.g., CCE at University B). In addition, the interaction effects with gender—where women consistently outperformed male students in key pathways—corroborate findings from Conger and Long (2010) and Chee et al. (2005), which have long documented women’s advantage in academic achievement in post-secondary education. Although age did not emerge as a dominant moderator, its inclusion is supported by previous work (Cassidy, 2012; Sheard, 2009) that identifies age as an important, albeit sometimes subtler, factor influencing academic success.

Together, these findings underscore the complexity of the relationship between ELP and academic achievement, confirming that while high standardized test scores generally predict stronger academic achievement, the context in which these scores are interpreted matters considerably as significant variation was found when moderator variables were included in analyses. The literature has repeatedly argued for the inclusion of discipline and demographic factors to explain variability in academic outcomes (e.g., Hellas et al., 2018; Gipson, 2018; Strayhorn, 2006; Van Herpen et al., 2017), and the results of this dissertation extend these insights by demonstrating that these moderators not only influence overall academic achievement but are relevant in terms of different ELP pathways.

These findings are significant because they demonstrate that the effectiveness of English language proficiency pathways in predicting academic achievement extends beyond a simple,

linear relationship with language scores or admission pathways. Instead, results confirm that the degree to which ELP pathways matter is shaped by factors such as academic discipline, country of origin, gender, and age. This insight is significant for several reasons. First, it offers a more precise lens through which institutions can interpret standardized test results or EAP completion. Rather than assuming that high IELTS or TOEFL scores guarantee uniform success, universities may consider tailoring their admissions thresholds and support services by accounting for discipline-specific demands, cultural or linguistic differences, and known gender gaps. In doing so, they can better identify at-risk students (e.g., men in EAP or Chinese students in TOEFL mid) and allocate resources—such as language support, peer mentoring, or bridging programs—accordingly.

Second, these findings can inform policy decisions regarding how to structure EAP programs or refine cut-off scores for different majors. For example, students in STEM were found to benefit disproportionately from higher ELP compared to those in Business at University A, thus institutions may need to revisit how they align admissions benchmarks with the unique linguistic and academic rigors of each field. Similarly, evidence that non-Chinese students outperform their Chinese peers in certain ELP pathways suggests that some groups may require additional cultural or academic scaffolding. By recognizing these nuanced interactions, universities can make more data-driven interventions—ranging from discipline-based writing clinics to gender-focused support—that elevate student performance more effectively than one-size-fits-all approaches.

Third, this dissertation's inclusion and analyses of moderator variables expands theoretical models of language proficiency and academic success by illustrating how students in different ELP pathways (e.g., EAP versus standardized tests) perform differently across specific

subgroups and institutional contexts. By refining our understanding of exactly where and for whom standardized test scores or EAP completion most strongly predict achievement, scholars and policymakers can evolve more nuanced frameworks for international student success—ones that fully acknowledge how culture, gender, age, and discipline work in tandem with language proficiency demonstrated through different pathways.

In essence, these findings articulate a more sophisticated and practically relevant picture of how ELP influences students' academic trajectories. Rather than viewing language proficiency as a monolithic predictor, institutions can harness these insights to foster more targeted admissions decisions, strengthen academic support programs, and ultimately improve graduation rates and student experiences. By recognizing that discipline and demographic backgrounds can magnify or mitigate the advantages of high ELP, universities become better equipped to support a diverse international student population, ensuring that language proficiency is leveraged—and complemented by appropriate resources—to maximize students' chances of academic success.

Significance and Contributions for Practice and Policy

The implications of these findings for higher education practice and policy are multifaceted as the findings of this dissertation highlight the multidimensionality of academic achievement among international students. The varied outcomes across proficiency pathways and institutions support the theoretical proposition that academic achievement in a second language context is moderated by contextual, institutional, and individual factors rather than predicted solely by linguistic proficiency measured via standardized tests or course-based assessments.

From a sociocultural perspective, these findings emphasize that linguistic competence may interact with institutional culture, curriculum design, and disciplinary conventions,

influencing students' academic trajectories differently. Additionally, social-cognitive theories on academic self-efficacy and motivation could further explain the observed variability across institutions, disciplines, and demographic groups (e.g., Schunk & Usher, 2012; Schunk & DiBenedetto, 2020). For instance, self-efficacy beliefs among high-achieving standardized-test students might mediate their superior academic performance, while context-specific interventions at University B might elevate efficacy and achievement among initially lower-scoring students.

The implications of these findings for higher education practice and policy suggest a need to re-examine how language proficiency is evaluated and supported within international student populations. First, the consistent advantage observed for students admitted via standardized ELP test scores (IELTS and TOEFL) over those admitted through EAP pathways—across both cumulative GPA and credits earned—suggests that universities must critically assess the design and delivery of their EAP programs. If EAP-admitted students continue to lag in academic achievement, institutions should consider revising these programs to integrate more context tailored content such as discipline-specific language training, academic writing support, and research skills development. Such enhancements could help close the achievement gap and ensure that EAP serves as a genuine bridge to academic success rather than a suboptimal alternative.

Furthermore, this dissertation's findings underscore the need for admissions policies to account for context-specific moderators such as discipline, country of origin, gender, and age. The differential outcomes observed between institutions—where University A, with a balanced disciplinary distribution and lower global rank, exhibited larger effect sizes than University B, which is highly ranked with a strong STEM focus—indicate that the predictive value of ELP

tests is not universal. This finding allows for a more comprehensive understanding of how institutional characteristics, such as disciplinary focus and academic culture may interact with language proficiency to shape international students' academic outcomes. However, more research is needed here to ascertain precisely the aspects that are the most predictive. Regardless, institutions with distinct academic profiles may need to tailor their cut-scores or placement criteria to align more closely with the language demands of specific universities and disciplines. For instance, because analyses revealed that high proficiency in STEM and Social Sciences & Humanities is strongly linked to increased credit accumulation, universities might consider setting higher ELP benchmarks for these fields or providing targeted supplemental instruction where necessary.

The role of demographic factors further highlights the importance of a nuanced approach. Data showed that country of origin and gender moderate academic outcomes in meaningful ways. Non-Chinese students, in many cases, tended to achieve higher CGPAs and earn more credits than their Chinese counterparts. Similarly, the consistent advantage observed among women suggests that gender-specific factors should be considered in both admissions and support programming. Given that men consistently underperformed relative to women—it is imperative that institutions develop targeted academic support programs specifically for men. Such interventions could include tailored academic advising, peer mentoring programs, study skills workshops, and initiatives designed to address the unique challenges that men may face in adapting to an English-medium academic environment. By proactively addressing these disparities, universities can help mitigate the disadvantage of men in higher education, ensuring that all students have access to the resources necessary to achieve their full academic potential. Recognizing these differences may allow universities to design targeted interventions, such as

mentorship programs or academic workshops, that address the unique needs of various other underperforming subgroups.

Theoretically, further refinement of predictive validity models to explicitly incorporate institutional characteristics, disciplinary linguistic demands, and sociocultural factors can advance our understanding of international student success. Such models should attempt to better capture the dynamic interplay of linguistic, cognitive, and sociocultural dimensions inherent in international students' educational trajectories, ultimately informing more nuanced, context-sensitive approaches to admission and student support policies.

These results are significant because they challenge the notion that standardized test scores alone can serve as a universal proxy for academic readiness in English medium universities. Instead, data reveal that the interplay of language proficiency with discipline-specific demands and demographic characteristics creates a complex landscape that demands more flexible, evidence-based policy responses. Policymakers and institutional leaders should therefore move beyond one-size-fits-all cut-offs and instead adopt admission and support strategies that are responsive to the varied contexts in which international students operate as well as the linguistic, cultural and demographic diversity of these students. This could involve periodic local validation studies, differential EAP programming tailored by discipline, and enhanced diagnostic assessments during orientation to identify students at risk of underperformance.

Ultimately, these findings are relevant to policy because they provide actionable insights that can lead to more effective educational practices. By accounting for the multifaceted nature of academic success—where language proficiency interacts with student background and institutional context, universities can better support the diverse needs of their international

student populations. This, in turn, may contribute to improved retention rates, higher graduation rates, and overall enhanced educational outcomes in a globalized higher education environment.

Limitations

While every effort was made to mitigate potential sources of error, several limitations inherent in the study's design, data collection processes, and analytical methods remain. Firstly, the lack of data standardization across institutions and significant issues encountered during the data collection phase posed considerable challenges. Specifically, extensive data cleaning was necessary due to inconsistencies and inaccuracies in how variables were initially recorded by the participating universities. This process inevitably resulted in the removal of numerous cases, which may have limited the overall representativeness of the final samples and introduced potential biases.

Secondly, important differences in sample characteristics between University A and University B should be considered. The institutional profiles varied substantially, with University A serving a more diverse student body across multiple disciplines and University B concentrating heavily on STEM disciplines. Such institutional differences may influence the extent to which the findings can be generalized to other academic contexts.

Additionally, this dissertation initially hypothesized convergence in the predictive validity of standardized proficiency tests and EAP pathways across universities. In other words, it was theorized that both universities would produce results favouring the test-based pathway. However, divergence rather than convergence was observed. Specifically, results demonstrated differing relationships between proficiency pathways and academic achievement at the two

institutions, complicating attempts to draw definitive conclusions or provide conclusive answers regarding the relative predictive validity of EAP versus standardized tests.

Furthermore, although critical moderator variables such as gender, age, discipline, and country of origin were included in the analyses, other potentially influential factors such as high school GPA, first language spoken, and EAP grades were not available due to institutional data limitations. The absence of these variables restricts a comprehensive understanding of baseline academic and linguistic preparedness, limiting the predictive power and interpretive depth of the dissertation's statistical models.

Lastly, this dissertation focused exclusively on first-year academic outcomes (CGPA and credits earned), excluding longitudinal analyses of student achievement over subsequent years or broader measures of academic success, such as retention rates or graduation outcomes. Longitudinal tracking would provide further insight into whether initial differences observed persist, narrow, or widen throughout the duration of students' academic careers.

Collectively, these limitations underline the complexity of evaluating academic achievement among international students and emphasize the need for further research incorporating comprehensive datasets, longitudinal designs, and additional background variables to more conclusively understand the predictors of academic success.

Recommendations for Future Research

Based on the limitations identified, several recommendations for future research emerge. First, future investigations should strive to secure additional comprehensive datasets. While this dissertation secured two institutions worth of data, the institutional variation found through analyses indicates that further exploration of ELP pathways at other institutions is warranted.

Considering the differences between University A and B in terms of institutional ranking, student demographics, disciplinary composition, and curricular approaches, future research could intentionally select for and attempt to isolate and control for these aspects. Such studies could examine how factors including institutional prestige, disciplinary mix, language support structures, and demographic composition independently and interactively influence the relationship between ELP pathways and academic achievement. Furthermore, expanding data collection to additional universities and perhaps colleges would allow for a more comprehensive picture of the variation in the effects of ELP pathways on academic achievement.

Additionally, researchers should aim to include additional academic background variables—such as high school GPA, first language spoken, and EAP grades—which would provide a more complete picture of students’ baseline academic and linguistic preparedness. Incorporating these measures would likely not only improve the predictive power of statistical models but also help disentangle the complex interplay between language proficiency and academic success.

In addition, future studies would benefit from employing a longitudinal design that tracks students beyond their first year. Longitudinal research could reveal whether the initial performance gap between standardized test pathways and EAP pathways persists, narrows, or widens over time. Such studies could also investigate how retention, graduation rates, and post-graduation outcomes are impacted by early ELP pathways and performance.

Finally, researchers should consider using a mixed methods approaches that combine quantitative with qualitative data, such as interviews with professors who teach a significant number of international students. Such research could explore the professoriates’ perceptions of readiness of international students admitted through different pathways for English medium

instruction and degree completion. Additionally, through work to explore the experiences of students admitted through different pathways at university and their perceptions of language support services, future studies could provide a richer context for understanding how and why different ELP pathways influence academic achievement. These insights could ultimately inform more nuanced, evidence-based admissions policies and tailored academic support programs in higher education, ensuring that the diverse needs of international students are effectively met.

References

- Abunawas, M. (2014). *A meta-analytic investigation of the predictive validity of the Test of English as a Foreign Language (TOEFL) scores on GPA* (Doctoral dissertation).
- Agresti, A. (2018). *Statistical methods for the social sciences* (5th ed.). Pearson.
- Adekitan, A. I., & Salau, O. (2019). The impact of engineering students' performance in the first three years on their graduation result using educational data mining. *Heliyon*, 5(2).
- Adelman, C. (1999). Answers in the Toolbox. Academic Intensity, Attendance Patterns, and Bachelor's Degree Attainment.
- Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 78(5), 218-223.
- Al-Malki, M. A. S. (2014). Testing the predictive validity of the IELTS test on Omani English candidates' professional competencies. *International Journal of Applied Linguistics and English Literature*, 3(5), 166-172.
- Alshammari, M. M. (2016). IELTS academic reading module test: Validity and reliability. *British Journal of English Linguistics*, 4(2), 26-37.
- Aluko, R. O., Daniel, E. I., Shamsideen Oshodi, O., Aigbavboa, C. O., & Abisuga, A. O. (2018). Towards reliable prediction of academic performance of architecture students using data mining techniques. *Journal of Engineering, Design and Technology*, 16(3), 385-397.

- Alyahyan, E., & Düşteğör, D. (2020). Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, 17(1), 3.
- Andrade, M. S. (2006). International students in English-speaking universities: Adjustment factors. *Journal of Research in International education*, 5(2), 131-154.
- Andrade, M. S. (2009). The effects of English language proficiency on adjustment to university life. *International Multilingual Research Journal*, 3(1), 16-34.
- Angelkoska, S., Stankovska, G., & Dimitrovski, D. (2016). The Personal Characteristics Predictors of Academic Success. *Bulgarian Comparative Education Society*.
- Arco-Tirado, J. L., Fernández-Martín, F. D., & Fernández-Balboa, J. M. (2011). The impact of a peer-tutoring program on quality standards in higher education. *Higher Education*, 62, 773-788.
- Arnold, I. J., & Rowaan, W. (2014). First-year study success in economics and econometrics: The role of gender, motivation, and math skills. *The Journal of Economic Education*, 45(1), 25-35.
- Arrigoni, E., & Clark, V. (2015). *Investigating the appropriateness of IELTS cut-off scores for admissions and placement decisions at an English-medium university in Egypt*. (IELTS Research Reports Online Series No. 3). IELTS. <https://www.ielts.org/en-us/for-researchers/research-reports/online-series-2015-3>

- Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & education, 113*, 177-194.
- Astin, A.W. (1993). *What matters in college? Four critical years revisited*. San Francisco: Jossey-Bass.
- Aulck, L., Velagapudi, N., Blumenstock, J., & West, J. (2016). Predicting student dropout in higher education. *International Journal of Educational Technology in Higher Education, 17*(3), 1-21.
- Aydın, G. (2017). Personal factors predicting college student success. *Eurasian Journal of Educational Research, 17*(69), 93-112.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford university press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press
- Baik, C., & Greig, J. (2009). Improving the academic outcomes of undergraduate ESL students: The case for discipline-based academic skills programs. *Higher Education Research & Development, 28*(4), 401–416. <https://doi.org/10.1080/07294360903067005>
- Bakker, A., Cai, J., English, L., Kaiser, G., Mesa, V., & Van Dooren, W. (2019). Beyond small, medium, or large: Points of consideration when interpreting effect sizes. *Educational Studies in Mathematics, 102*, 1-8.

- Bista, K. (2014). The influence of country of origin and academic level on Asian students' gains of learning. *Journal of International Students, 2015 Vol. 5 No. 3, 5(3)*, 300-305.
- Bo, W. V., Fu, M., & Lim, W. Y. (2023). Revisiting English language proficiency and its impact on the academic performance of domestic university students in Singapore. *Language Testing, 40(1)*, 133-152.
- Bowman, N. A. (2014). Conceptualizing openness to diversity and challenge: Its relation to college experiences, achievement, and retention. *Innovative Higher Education, 39*, 277-291.
- Bretag, T. (2007). The emperor's new clothes: Yes, there is a link between English language competence and academic standards. *People and Place, 15(1)*, 13
- Bretz Jr, R. D. (1989). College grade point average as a predictor of adult success: A meta-analytic review and some additional evidence. *Public Personnel Management, 18(1)*, 11-22.
- Bridgeman, B., Cho, Y., & DiPietro, S. (2016). Predicting grades from an English language assessment: The importance of peeling the onion. *Language Testing, 33(3)*, 307-318.
- British Council (2017). IELTS numbers rise to three million a year. Retrieved from: <https://www.britishcouncil.org/contact/press/ielts-numbers-rise-three-million-year#:~:text=IELTS%20is%20the%20International%20English,taken%20in%20the%20last%20year>
- Brown, J. D. (2011). *Testing in language programs: A comprehensive guide to English language assessment*. [Jamesa Dean Brown].

- Burgess, T. C., & Greis, N. A. (1970). English Language Proficiency and Academic Achievement among Students of English as a Second Language at the College Level.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied linguistics*, 1(1), 1-47.
- Carvalho, R. G. G. (2016). Gender differences in academic achievement: The mediating role of personality. *Personality and Individual Differences*, 94, 54-58
- Cassidy, S. (2012). Exploring individual differences as determining factors in student academic achievement in higher education. *Studies in Higher Education*, 37(7), 793-810.
- Chee, K. H., Pino, N. W., & Smith, W. L. (2005). Gender differences in the academic ethic and academic achievement. *College student journal*, 39(3), 604-619.
- Chemers, M. M., Hu, L. T., & Garcia, B. F. (2001). Academic self-efficacy and first year college student performance and adjustment. *Journal of Educational psychology*, 93(1), 55.
- Chen, X. (2005). *First-generation students in postsecondary education: A look at their college Transcripts*. U.S. Department of Education. Washington, DC: U.S. Government Printing Office.
- Chen, X., Saafir, A., & Graham, S. (2020). Ethnicity, peers, and academic achievement: who wants to be friends with the smart kids? *Journal of Youth and Adolescence*, 49, 1030-1042.
- Cheng, L., Myles, J., & Curtis, A. (2004). Targeting language support for non-native English speaking graduate students at a Canadian university. *TESL Canada Journal*, 22, 50–71.

- Cheng, L., & Fox, J. (2008). Towards a Better Understanding of Academic Acculturation: Second Language Students in Canadian Universities. *Canadian Modern Language Review*, 65(2), 307–333. <https://doi.org/10.3138/cmlr.65.2.307>
- Cho, Y., & Bridgeman, B. (2012). Relationship of TOEFL iBT® scores to academic performance: Some evidence from American universities. *Language Testing*, 29(3), 421–442.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. T. Press Cambridge.
- Clark, A. T., Lippincott, D., & Kim, J. (2021). More than Learning English? The Impact of University Intensive English Language Program Attendance on International Student Academic Achievement. *Education Policy Analysis Archives*, 29(43)
- Conger, D., & Long, M. C. (2010). Why are men falling behind? Gender gaps in college performance and persistence. *The Annals of the American Academy of Political and Social Science*, 627(1), 184–214.
- Corcoran, J. N., Williams, J., & Johnston, K. P. (2022). English for academic purposes in Canada: Results from an exploratory national survey. *BC TEAL Journal*, 7(1), 55–84.
- Credé, M., Roch, S. G., & Kieszczynka, U. M. (2010). Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research*, 80(2), 272–295.
- Crossman, E., Choi, Y., & Hou, F. (2022). International students as a source of labour supply: A summary of recent trends [Economic and Social Reports]. *Statistics Canada*.
- Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the

optimum age question and some other matters. *Working Papers on Bilingualism*, No. 19, 121-129.

Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada:

A reassessment. *Applied Linguistics*, 1, 132-149

Cummins, J. (1984). *Bilingualism and special education: Issues in assessment & pedagogy*.

Avon, U.K.: Multilingual Matters

Cummins, J. (2001). Negotiating identities: Education for empowerment in a diverse

society (2nd ed.). Los Angeles: California Association for Bilingual Education.

Cummins, J. (2008) BICS and CALP: Empirical and theoretical status of distinction. In Street, B.

& Hornberger, N. H. (Eds.). *Encyclopedia of Language and Education, 2nd Edition*,

Volume 2: Literacy. (pp. 71-83). New York: Springer Science + Business Media LLC

Cummins, J. & Oller, Jr. J.W. (1983) In Cummins, J., & Swain, M. (2014). *Bilingualism in*

education: Aspects of theory, research and practice (pp. 138-161). Routledge.

D'Agostino, J. V., & Powers, S. J. (2009). Predicting teacher performance with test scores and

grade point average: A meta-analysis. *American Educational Research Journal*, 46(1),

146-182.

Daller, M., & Yixin, W. (2017). Predicting study success of international students. *Applied*

Linguistics Review, 8(4), 355-374.

- Dang, C. N., & Dang, T. N. Y. (2023). The Predictive Validity of the IELTS Test and Contribution of IELTS Preparation Courses to International Students' Subsequent Academic Study: Insights from Vietnamese International Students in the UK. *RELC Journal*, 54(1), 84–98. <https://doi.org/10.1177/0033688220985533>
- Davies, A. and Criper, C. (1988) IELTS research report 1 (1), ELTS Validation Project Report. London: British Council and Cambridge: UCLES
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T. and McNamara, T. (1999). *Dictionary of Language Testing. Studies in Language Testing, Volume 7*. Cambridge, UK: UCLES/Cambridge University Press.
- Davies, S., & Pizarro Milian, R. (2020). Transfer student outcomes at the University of Toronto: GPA, access to STEM, and graduation. *Toronto: Ontario Council on Articulation and Transfer*.
- DeFreitas, S. C. (2012). Differences between African American and European American first-year college students in the relationship between self-efficacy, outcome expectations, and academic achievement. *Social Psychology of Education*, 15(1), 109-123.
- Dewey, M. (2007). English as a lingua franca and globalization: an interconnected perspective. *International Journal of Applied Linguistics*, 17(3).
- De Wit, K., Heerwegh, D., & Verhoeven, J. C. (2012). Do ICT competences support educational attainment at university?. *Journal of Information Technology Education: Research*, 11(1), 1-25.

- Diamond, M. R., & O'Brien-Malone, A. (2018). Pathways to performance: an examination of entry pathway and first-year university results. *Asia Pacific journal of education*, 38(1), 110-128.
- Dooley, P. (2010). Students' perspectives of an EAP pathway program. *Journal of English for Academic Purposes*, 9(3), 184–197. <https://doi.org/10.1016/j.jeap.2010.02.013>
- Dooley, P., & Oliver, R. (2002). An investigation into the predictive validity of the IELTS Test as an indicator of future academic success. *Prospect*, 17(1), 36-54.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., & Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27-46.
- Douglas, S. R., & Rosvold, M. (2018). Intercultural Communicative Competence and English for Academic Purposes: A Synthesis Review of the Scholarly Literature. *Canadian Journal of Applied Linguistics*, 21(1), 23–42. <https://doi.org/10.7202/1050809ar>
- Douglas, S. R., & Landry, M. (2021). English for academic purposes programs: Key trends across Canadian universities. *Comparative and International Education*, 50(1).
- Dye, D. A., & Reck, M. (1989). College grade point average as a predictor of adult success: A reply. *Public Personnel Management*, 18(2), 235-241.
- Enikő, M. E., & Szamosközi, I. (2017). Predictors of academic success in college: Actual learning skills. *Erdelyi Pszichologiai Szemle= Transylvanian Journal of Psychology*, 18(1).

ETS, 2010, Linking TOEFL iBT Scores to IELTS Scores. Retrieved from:

https://www.ets.org/s/toefl/pdf/linking_toefl_ibt_scores_to_ielts_scores.pdf

Fauria, R. M., & Fuller, M. B. (2015). Transfer Student Success: Educationally Purposeful

Activities Predictive of Undergraduate GPA. *Research & Practice in Assessment*, 10, 39-52.

Feast, V. (2002). The impact of IELTS scores on performance at university. *International Education Journal*, 3(4), 70-85.

Fedorova, N. (2021). Questioning EAP: a critique of the English for Academic Purposes courses at university. *Journal of Teaching English for Specific and Academic Purposes*, 401-408.

Floyd, C. (2015). Closing the gap: International student pathways, academic performance and academic acculturation. *Journal of Academic Language & Learning*, 9(2).

<https://pdfs.semanticscholar.org/3c60/52c325372e5f096fe2e620341f1d7ef18c06.pdf>

Fox, J. (2005). Rethinking Second Language Admission Requirements: Problems With

Language-Residency Criteria and the Need for Language Assessment and Support. *Language Assessment Quarterly*, 2(2), 85–115.

https://doi.org/10.1207/s15434311laq0202_1

Fox, J., Cheng, L., Berman, R., Song, X., & Myles, J. (2006). Costs and benefits: English for academic purposes instruction in Canadian universities. *Carleton Papers in Applied Language Studies [CPALS]*, 23, 1-108.

Fox, J., Cheng, L., & Zumbo, B. D. (2013). Do they make a difference? The impact of English

- language programs on second language students in Canadian universities. *TESOL Quarterly*, 48(1), 57-85.
- Friedman, B. A., & Mandel, R. G. (2011). Motivation predictors of college student academic performance and retention. *Journal of College Student Retention: Research, Theory & Practice*, 13(1), 1-15.
- Frost, K., & McNamara, T. (2018). Language tests, language policy and citizenship. *The Oxford handbook of language policy and planning*, 280-298.
- Gagen, T. (2019). The Predictive Validity of IELTS Scores: A Meta-Analysis (publication number 6406) [Doctoral dissertation, Western University, Electronic Thesis and Dissertation Repository].
<https://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=8762&context=etd>
- Gershenfeld, S., Ward Hood, D., & Zhan, M. (2016). The role of first-semester GPA in predicting graduation rates of underrepresented students. *Journal of College Student Retention: Research, Theory & Practice*, 17(4), 469-488.
- Ghenghesh, P. (2015). The relationship between English language proficiency and academic performance of university students—should academic institutions really be concerned?. *International Journal of Applied linguistics and English literature*, 4(2), 91-97.
- Gibbison, G. A., Henry, T. L., & Perkins-Brown, J. (2011). The chicken soup effect: The role of recreation and intramural participation in boosting freshman grade point average. *Economics of Education Review*, 30(2), 247-257.

- Ginther, A., & Elder, C. (2014). A comparative investigation into understandings and uses of the TOEFL iBT® Test, the International English Language Testing Service (Academic) test, and the Pearson Test of English for graduate admissions in the United States and Australia: A case study of two university contexts. *ETS Research Report Series*, 2014(2), 1-39.
- Ginther, A., & Yan, X. (2018). Interpreting the relationships between TOEFL iBT scores and GPA: Language proficiency, policy, and profiles. *Language Testing*, 35(2), 271-295
- Goldsmith, R., Havery, C., Edwards, E., James, N., Murphy, A., Mort, P., ... & Yeo, J. (2022). A multi-faceted evaluation of the impact on students of an Australian university-wide academic language development program. *Journal of English for Academic Purposes*, 60, 101192.
- Government of Canada (2019). *Building on success: Canada's International Education Strategy (2019-2024)*. Retrieved from: <https://www.international.gc.ca/education/strategy-2019-2024-strategie.aspx?lang=eng>
- Government of Canada (2024). *Canada to stabilize growth and decrease number of new international student permits issued to approximately 360,000 for 2024*. Retrieved from: <https://www.canada.ca/en/immigration-refugees-citizenship/news/2024/01/canada-to-stabilize-growth-and-decrease-number-of-new-international-student-permits-issued-to-approximately-360000-for-2024.html>
- Government of Canada (2024). *Renewing Canada's IES - Proposed pillar: Diversification in international education*. Retrieved from: https://www.international.gc.ca/education/strategy_research-strategie_recherche/diversification.aspx?lang=eng

- Government of Canada (2025). *2025 provincial and territorial allocations under the international student cap*. Retrieved from: <https://www.canada.ca/en/immigration-refugees-citizenship/news/notices/2025-provincial-territorial-allocations-under-international-student-cap.html>
- Graham, J. G. (1987). English language proficiency and the prediction of academic success. *TESOL quarterly*, 21(3), 505-521.
- Hagedorn, L. S. (2005). How to define retention. *College student retention formula for student success*, 90-105.
- Hamid, M. O., Hoang, N. T., & Kirkpatrick, A. (2019). Language tests, linguistic gatekeeping and global mobility. *Current Issues in Language Planning*, 20(3), 226-244.
- Hannon, B. (2014). Predicting college success: The relative contributions of five social/personality factors, five cognitive/learning factors, and SAT scores. *Journal of education and training studies*, 2(4), 46.
- Harsch, C., Ushioda, E., & Ladroue, C. (2017). Investigating the predictive validity of TOEFL iBT® test scores and their use in informing policy in a United Kingdom university setting. *ETS Research Report Series*, 2017(1), 1-80.
- Hashemi, A., & Daneshfar, S. (2018). A review of the IELTS test: Focus on validity, reliability, and washback. *IJELTAL (Indonesian Journal of English Language Teaching and Applied Linguistics)*, 3(1), 39-52.
- Hellas, A., Ihantola, P., Petersen, A., Ajanovski, V. V., Gutica, M., Hynninen, T., ... & Liao, S. N. (2018, July). Predicting academic performance: a systematic literature review.

In *Proceedings companion of the 23rd annual ACM conference on innovation and technology in computer science education* (pp. 175-199).

Heller, M. (2003). Globalization, the new economy, and the commodification of language and identity. *Journal of sociolinguistics*, 7(4), 473-492.

Hill, K., Storch, N., & Lynch, B. (1999). A comparison of IELTS and TOEFL as predictors of academic success. *International English Language Testing System (IELTS) Research Reports 1999, Volume 2*, 62.

Hirsh, D. (2006). English language, academic support and academic outcomes: A discussion paper. *University of Sydney Papers in TESOL*, 2(2), 193–211

Hoang, N. T. H. (2019). Building a validity argument for the use of academic language tests for immigration purposes: evidence from immigration-seeking test-takers. *Language Education & Assessment*, 2(3), 135-154

Hoekje, B., Harman, T., & Heitner, R. (2021). English language proficiency requirements in university conditional admission. *Journal of Academic Language and Learning*, 15(1), 144-161.

Holmes, J. R. (2021). An Examination of Student Success within Canadian Higher Education: Fifty years of findings and recommendations for the future. *Canadian Journal of Higher Education*, 51(3), 13-32.

Hu, R., & Trenkic, D. (2021). The effects of coaching and repeated test-taking on Chinese candidates' IELTS scores, their English proficiency, and subsequent academic

- achievement. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1486-1501.
- Hulstijn, J. H. (2011). Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly*, 8(3), 229-249.
- Hwang, K. Y., & Dizney, H. F. (1970). Predictive validity of the test of English as a foreign language for Chinese graduate students at an American university. *Educational and Psychological Measurement*, 30(2), 475-477.
- Ihlenfeldt, S. D., & Rios, J. A. (2023). A meta-analysis on the predictive validity of English language proficiency assessments for college admissions. *Language Testing*, 40(2), 276-299.
- Ingram, D., & Bayliss, A. (2007). IELTS as a predictor of academic language performance, Part 1. In *International English Language Testing System (IELTS) Research Reports 2007: Volume 7* (pp. 1-68). [Canberra]: IELTS Australia and British Council.
- International Organization for Migration. (2023). *World Migration Report 2024: Chapter 2 - International students*. World Migration Report. <https://worldmigrationreport.iom.int/what-we-do/world-migration-report-2024-chapter-2/international-students>
- Institute for Digital Research and Education. (n.d.). *Regression with Stata: Chapter 3 - Regression with categorical predictors*. UCLA: Statistical Consulting Group. <https://stats.oarc.ucla.edu/stata/webbooks/reg/chapter3/regression-with-statachapter-3-regression-with-categorical-predictors/>

- Jansen, E. P., & Suhre, C. J. (2010). The effect of secondary school study skills preparation on first-year university achievement. *Educational studies*, 36(5), 569-580.
- Jenkins, J. (2013). *English as a lingua franca in the international university: The politics of academic English language policy*. Routledge.
- Jordan, R. (2004). *English for academic purposes*. Cambridge, England: Cambridge University Press.
- Johnson, R. C., & Tweedie, M. G. (2021). "IELTS-out/TOEFL-out": Is the end of general English for academic purposes near? Tertiary student achievement across standardized tests and general EAP. *Interchange*, 52(1), 101-113.
- Karim, S., & Haq, N. (2014). An Assessment of IELTS Speaking Test. *International Journal of Evaluation and Research in Education*, 3(3), 152-157.
- Kerstjens, M., & Nery, C. (2000). Predictive validity in the IELTS test: A study of the relationship between IELTS scores and students' subsequent academic performance. In *International English Language Testing System (IELTS) Research Reports 2000: Volume 3* (pp. 85-108). [Melbourne]: IELTS Australia.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Pergamon Press.
- Kuh, G. D., Kinzie, J. L., Buckley, J. A., Bridges, B. K., & Hayek, J. C. (2006). *What matters to student success: A review of the literature* (Vol. 8). Washington, DC: National Postsecondary Education Cooperative.
- Lehmann, P. S., & Meldrum, R. C. (2021). School suspension in Florida: The interactive effects of race, ethnicity, gender, and academic achievement. *Justice quarterly*, 38(3), 479-512.

- Leppel, K. (2001). The impact of major on college persistence among freshmen. *Higher Education, 41*, 327-342.
- Letkiewicz, J., Lim, H., Heckman, S., Bartholomae, S., Fox, J. J., & Montalto, C. P. (2014). The path to graduation: Factors predicting on-time graduation rates. *Journal of College Student Retention: Research, Theory & Practice, 16*(3), 351-371.
- Li, J. (2019). An evaluation of IELTS speaking test. *Open Access Library Journal, 6*(12), 1.
- Light, R. L., Xu, M., & Mossop, J. (1987). English Proficiency and Academic Performance of international Students. *Tesol Quarterly, 21*(2), 251-261.
- Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician, 54*(3), 217-224.
- Makarova, E., & Birman, D. (2015). Cultural transition and academic achievement of students from ethnic minority backgrounds: A content analysis of empirical research on acculturation. *Educational Research, 57*(3), 305-330.
- Malm, J., Bryngfors, L., & Mörner, L. L. (2012). Supplemental instruction for improving first year results in engineering studies. *Studies in Higher education, 37*(6), 655-666.
- Martirosyan, N. M., Hwang, E., & Wanjohi, R. (2015). Impact of English proficiency on academic performance of international students. *Journal of International Students, 5*(1), 60-71.
- Mattern, K. D., & Shaw, E. J. (2010). A look beyond cognitive predictors of academic success: Understanding the relationship between academic self-beliefs and outcomes. *Journal of college student development, 51*(6), 665-678.

- Metcalfe, A. S. (2010). Revisiting academic capitalism in Canada: No longer the exception. *The Journal of Higher Education*, 81(4), pp. 489-514.
- Moore, T., Morton, J., & Price, S. (2012). *Construct validity in the IELTS academic reading test* (Vol. 11). IELTS Research Report 4.
- Müller, A., & Daller, M. (2019). Predicting international students' clinical and academic grades using two language tests (IELTS and C-test): A correlational research study. *Nurse Education Today*, 72, 6–11. <https://doi.org/10.1016/j.nedt.2018.10.007>
- Murray, N. (2013). Widening participation and English language proficiency: A convergence with implications for assessment practices in higher education. *Studies in Higher Education*, 38(2), 299-311.
- Murray, N. (2016). *Standards of English in higher education: Issues, challenges and strategies*. Cambridge University Press.
- Neil, M. (2018). University gatekeeping tests: What are they really testing and what are the implications for EAP provision?. *Jacet journal*, 62, 15-27
- Murray, N., & Hicks, M. (2016). An institutional approach to English language proficiency. *Journal of Further and Higher Education*, 40(2), 170-187.
- Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear statistical models*.
- Neumann, H., Padden, N., & McDonough, K. (2019). Beyond English language proficiency scores: Understanding the academic performance of international undergraduate students during the first year of study. *Higher Education Research & Development*, 38(2), 324-338.

- Noble, J., & Sawyer, R. (2002). Predicting Different Levels of Academic Success in College Using High School GPA and ACT Composite Score. ACT Research Report Series.
- Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: big five correlates of GPA and SAT scores. *Journal of personality and social psychology*, 93(1), 116
- O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual differences*, 43(5), 971-990.
- O'Dwyer, J., Kantarcioğlu, E., & Thomas, C. (2018). An investigation of the predictive validity of the TOEFL iBT® test at an english-medium university in Turkey. *ETS Research Report Series*, 2018(1), 1-13.
- Oguntunde, P. E., Okagbue, H. I., Oguntunde, O. A., Opanuga, A. A., & Oluwatunde, S. J. (2018). Analysis of the inter-relationship between students' first year results and their final graduating grades. *International Journal of Advanced and Applied Sciences*, 5(10), 1-6.
- Oliver, R., Vanderford, S., & Grote, E. (2012). Evidence of English language proficiency and academic achievement of non-English-speaking background students. *Higher Education Research & Development*, 31(4), 541-555.
- Parasnis, J., & Swan, J. (2020). Differences in educational attainment by country of origin: Evidence from Australia. *Migration Studies*, 8(4), 530-553.
- Park, C. L., Boman, J., Care, W. D., Edwards, M., & Perry, B. (2008). Persistence and attrition: What is being measured? *Journal of College Student Retention: Research, Theory & Practice*, 10(2), 223-233.

- Pascarella, E. T., & Terenzini, P. T. (2005). *How College Affects Students: A Third Decade of Research. Volume 2*. Jossey-Bass, An Imprint of Wiley. 10475 Crosspoint Blvd, Indianapolis, IN 46256.
- Pearce II, J. A. (2017). How employers can stanch the hemorrhaging of collegiate GPA credibility. *Business Horizons*, 60(1), 35-43.
- Pearson, W. S. (2020). The effectiveness of pre-sessional EAP programmes in UK higher education: A review of the evidence. *Review of Education*, 8(2), 420-447.
- Pearson, W. S. (2021). Policies on minimum English language requirements in UK higher education, 1989-2021. *Journal of Further and Higher Education*, 45(9), 1240-1252.
- Peltekov, P. (2021). The International English Language Testing System (IELTS): A Critical Review. *Journal of English Language Teaching and Linguistics*, 6(2), 395.
- Pilcher, N., & Richards, K. (2017). Challenging the power invested in the International English Language Testing System (IELTS): Why determining 'English' preparedness needs to be undertaken within the subject context. *Power and Education*, 9(1), 3-17.
- Piller, I., & Bodis, A. (2024). Marking and unmarking the (non) native speaker through English language proficiency requirements for university admission. *Language in Society*, 53(1), 1-23.
- Pirmohamed, S., Debowska, A., & Boduszek, D. (2017). Gender differences in the correlates of academic achievement among university students. *Journal of Applied Research in Higher Education*, 9(2), 313-324.

- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language learning*, 64(4), 878-912.
- Pochon-Berger, E., & Lenz, P. (2014). Language requirements and language testing for immigration and integration purposes: a synthesis of academic literature: Scientific Centre of Expertise in Multilingualism (SCEM).
- Pommerich, M. (2007). Concordance: The Good, the Bad, and the Ugly. In: Dorans, N.J., Pommerich, M., Holland, P.W. (eds) *Linking and Aligning Scores and Scales. Statistics for Social and Behavioral Sciences*. Springer, New York, NY. https://doi.org/10.1007/978-0-387-49771-6_11
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological bulletin*, 135(2), 322.
- Read, J. (2022). Test review: The international English language testing system (IELTS). *Language testing*, 39(4), 679-694.
- Reason, R. D. (2009). An examination of persistence research through the lens of a comprehensive conceptual framework. *Journal of College Student Development*, 50(6), 659-682.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological bulletin*, 138(2), 353.

- Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological bulletin*, 130(2), 261.
- Roessingh, H., & Douglas, S. R. (2012). Educational outcomes of English language learners at university. *Canadian Journal of Higher Education*, 42(1), 80–97.
<https://doi.org/10.47678/cjhe.v42i1.182449>
- Rose, H., Curle, S., Aizawa, I., & Thompson, G. (2020). What drives success in English medium taught courses? The interplay between language proficiency, academic skills, and motivation. *Studies in Higher Education*, 45(11), 2149-2161.
- Ruegg, R., Petersen, N., Hoang, H., & Marianne. (2020). Effects of pathways into university on the academic success of international undergraduate students. *Higher Education Research & Development*, 40(6), 1283-1297.
- Ruegg, R., Hoang, H., & Petersen, N. (2024). The influence of English language proficiency test scores on the academic success of ESL undergraduate students. *Educational Research and Evaluation*, 29(3-4), 111-129.
- Sackett, P. R., Kuncel, N. R., Arneson, J. J., Cooper, S. R., & Waters, S. D. (2009). Socioeconomic Status and the Relationship between the SAT® and Freshman GPA: An Analysis of Data from 41 Colleges and Universities. Research Report No. 2009-1. *College Board*.
- Sadeghi, B., Kashanian, N. M., Maleki, A., & Haghdooost, A. (2013). English Language Proficiency as a Predictor of Academic Achievement among Medical Students in Iran. *Theory & Practice in Language Studies (TPLS)*, 3(12).

- Saunders-Scott, D., Braley, M. B., & Stennes-Spidahl, N. (2018). Traditional and psychological factors associated with academic success: Investigating best predictors of college retention. *Motivation and emotion*, 42, 459-465.
- Sax, L. J., & Harper, C. E. (2007). Origins of the gender gap: Pre-college and college influences on differences between men and women. *Research in Higher Education*, 48, 669-694
- Schoepp, K. (2018). Predictive validity of the IELTS in an English as a medium of instruction environment. *Higher Education Quarterly*, 72(4), 271-285.
- Schoepp, K., & Garinger, D. (2016). IELTS and academic success in higher education: A UAE perspective. *International Journal of Applied Linguistics and English Literature*, 5(3), 145-151.
- Schunk, D. H., & Usher, E. L. (2012). Social cognitive theory and motivation. *The Oxford handbook of human motivation*, 2, 11-26.
- Schunk, D. H., & DiBenedetto, M. K. (2020). Motivation and social cognitive theory. *Contemporary educational psychology*, 60, 101832.
- Schwager, I. T., Hülshager, U. R., Bridgeman, B., & Lang, J. W. (2015). Graduate Student Selection: Graduate record examination, socioeconomic status, and undergraduate grade point average as predictors of study success in a western European University. *International Journal of Selection and Assessment*, 23(1), 71-79.
- Severiens, S. E., & Schmidt, H. G. (2009). Academic and social integration and study progress in problem based learning. *Higher education*, 58(1), 59-69.

- Shakibaei, G., & Memari, M. (2019). On the predictive validity (Generalizability) of IELTS academic tests. *Teaching English Language, 13*(2), 299-317.
- Shao, X., Hufnagel, P., & Karp, R. (2010). The effect of clarion transitions on student academic performance and retention. *Journal of College Student Retention: Research, Theory & Practice, 11*(4), 435-457.
- Shaw, E. J., Mattern, K. D., & Patterson, B. F. (2011). Discrepant SAT critical reading and writing scores: Implications for college performance. *Educational Assessment, 16*(3), 145-163.
- Sheard, M. (2009). Hardiness commitment, gender, and age differentiate university academic performance. *British Journal of Educational Psychology, 79*(1), 189-204.
- Shohamy, E. (2007). Language tests as language policy tools. *Assessment in education, 14*(1), 117-130.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of educational research, 75*(3), 417-453.
- Soldner, M., Smither, C., Parsons, K., & Peek, A. (2016). Toward improved measurement of student persistence and completion. *Text. American Institutes for Research. May, 16*.
- Solórzano, R. W. (2008). High Stakes Testing: Issues, Implications, and Remedies for English Language Learners. *Review of Educational Research, 78*(2), 260–329.
- <https://doi.org/10.3102/0034654308317845>

- Sonnert, G., & Fox, M. F. (2012). Women, men, and academic performance in science and engineering: The gender difference in undergraduate grade point averages. *The Journal of Higher Education*, 83(1), 73-101.
- Soto, C., & Pérez-Milans, M. (2021). Language, neoliberalism, and the commodification of pedagogy. In *Education and the Discourse of Global Neoliberalism* (pp. 20-36). Routledge.
- Statistics Canada (2024). *Is the recent spike in international students affecting domestic university enrolment at Canadian public postsecondary institutions?* Retrieved from: <https://www.statcan.gc.ca/o1/en/plus/5435-recent-spike-international-students-affecting-domestic-university-enrolment-canadian>
- Steinmayr, R., Meiner, A., Weideinger, A. F., & Wirthwein, L. (2014). *Academic achievement* (pp. 9780199756810-0108). Oxford, UK:: Oxford University Press.
- Stoynoff, S. (1997). Factors associated with international students' academic achievement. *Journal of instructional psychology*, 24(1), 56.
- Strayhorn, T. L. (2007). Factors influencing the academic achievement of first-generation college students. *Journal of Student Affairs Research and Practice*, 43(4), 1278-1307.
- Tan, L. (2021). Performance comparison between IB and non-IB school students on the International Schools' Assessment. *Australian Council for Educational Research*
- Thomas, M. (1994). Assessment of L2 proficiency in second language acquisition research. *Language learning*, 44, 307-307.
- Thorpe, A., Snell, M., Davey-Evans, S., & Talman, R. (2017). Improving the Academic

- Performance of Non-native English-Speaking Students: The Contribution of Pre-sessional English Language Programmes: Contribution of Pre-sessional English Language Programmes. *Higher Education Quarterly*, 71(1), 5–32.
<https://doi.org/10.1111/hequ.12109>
- Tinto, V. (1999). Taking student retention seriously: Rethinking the first year of college. *NACADA journal*, 19(2), 5-9.
- Tinto, V. (2017). Reflections on student persistence. *Student Success*, 8(2), 1-8.
- Tomlinson, B. (2020). Which test of which English and why? In *The Routledge handbook of world Englishes* (pp. 641-658). Routledge.
- Torenbeek, M., Jansen, E., & Hofman, A. (2010). The effect of the fit between secondary and university education on first-year student achievement. *Studies in higher education*, 35(6), 659-675.
- Trapmann, S., Hell, B., Hirn, J. O. W., & Schuler, H. (2007). Meta-analysis of the relationship between the Big Five and academic success at university. *Zeitschrift für Psychologie/Journal of Psychology*, 215(2), 132-151.
- Tweedie, M. G., & Chu, M. W. (2019). Challenging equivalency in measures of English language proficiency for university admission: data from an undergraduate engineering programme. *Studies in Higher Education*, 44(4), 683-695.
- Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational psychologist*, 39(2), 111-133.
- Vaquera, E., & Kao, G. (2012). Educational achievement of immigrant adolescents in Spain: Do gender and region of origin matter?. *Child development*, 83(5), 1560-1576

- Van Herpen, S. G., Meeuwisse, M., Hofman, W. A., Severiens, S. E., & Arends, L. R. (2017). Early predictors of first-year academic success at university: pre-university effort, pre-university self-efficacy, and pre-university reasons for attending university. *Educational Research and Evaluation*, 23(1-2), 52-72.
- Van der Zanden, P. J., Denessen, E., Cillessen, A. H., & Meijer, P. C. (2018). Domains and predictors of first-year student success: A systematic review. *Educational Research Review*, 23, 57-77.
- Vella, E. J., Turesky, E. F., & Hebert, J. (2016). Predictors of academic success in web-based courses: Age, GPA, and instruction mode. *Quality Assurance in Education*, 24(4), 586-600.
- Vulperhorst, J., Lutz, C., de Kleijn, R., & van Tartwijk, J. (2018). Disentangling the predictive validity of high school grades for academic success in university. *Assessment & Evaluation in Higher Education*, 43(3), 399-414.
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press
- Waluyo, B., & Panmei, B. (2021). English Proficiency and Academic Achievement: Can Students' Grades in English Courses Predict Their Academic Achievement?. *Mextesol Journal*, 45(4), n4.
- Warikoo, N., & Carter, P. (2009). Cultural explanations for racial and ethnic stratification in academic achievement: A call for a new and improved theory. *Review of Educational Research*, 79(1), 366-394.

- Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. M., & Schmidt, F. L. (2015). College performance and retention: A meta-analysis of the predictive validities of ACT® scores, high school grades, and SES. *Educational Assessment*, 20(1), 23-45
- Wharrad, H. J., Chapple, M., & Price, N. (2003). Predictors of academic success in a Bachelor of Nursing course. *Nurse Education Today*, 23(4), 246-254.
- Wilder, S. (2023). Effects of parental involvement on academic achievement: a meta-synthesis. In *Mapping the field* (pp. 137-157). Routledge.
- Wilson, C. A., Babcock, S. E., & Saklofske, D. H. (2019). Sinking or swimming in an academic pool: A study of resiliency and student success in first-year undergraduates. *Canadian Journal of Higher Education*, 49(1), 60–84. <https://doi.org/10.47678/cjhe.v49i1.188220>
- Woodrow, L. (2006). Academic success of international postgraduate education students and the role of English proficiency. *University of Sydney papers in TESOL*, 1(1), 51-70.
- Wongtrirat, R. (2010). English language proficiency and academic achievement of international students: A meta-analysis.
- Wonnacott, K. M. (2024). Language Proficiency Requirements and International Enrolment: A Systematized Review.
- Yen, D., & Kuzma, J. (2009). Higher IELTS score, higher academic performance? The validity of IELTS in predicting the academic performance of Chinese students. *Worcester Journal of Learning and Teaching*, (3).
- Yao, D., & Wallace, M. P. (2021). Language assessment for immigration: A review of validation research over the last two decades. *Frontiers in Psychology*, 12, 773132.

- Yixin, W., & Daller, M. (2014, September). Predicting Chinese students' academic achievement in the UK. In *Proceedings of the 47th annual meeting of the British association for applied linguistics, learning, working and communicating in a global context* (pp. 217-227).
- York, T. T., Gibson, C., & Rankin, S. (2015). Defining and measuring academic success. *Practical Assessment, Research, and Evaluation*, 20(1), 5.
- Zheng, Z., & Mustapha, S. M. (2022). A literature review on the academic achievement of college students. *Journal of Education and Social Sciences*, 20(1), 11-18.
- Zimmermann, J., Brodersen, K. H., Heinemann, H. R., & Buhmann, J. M. (2015). A Model-Based Approach to Predicting Graduate-Level Performance Using Indicators of Undergraduate-Level Performance. *Journal of Educational Data Mining*, 7(3), 151-176.
- Zhou, Y. X., Ou, C. Q., Zhao, Z. T., Wan, C. S., Guo, C., Li, L., & Chen, P. Y. (2015). The impact of self-concept and college involvement on the first-year success of medical students in China. *Advances in Health Sciences Education*, 20, 163-179.
- Zwick, R. (2013). Disentangling the role of high school grades, SAT® scores, and SES in predicting college achievement. *ETS Research Report Series*, 2013(1), 1-20.

**Appendix A: Pairwise Comparison for CGPA by ELP Pathway and Discipline for
University A**

Margin Predicted CGPA –by ELP Pathway and Discipline for University A

Pathway Group	Discipline	Margin (Predicted CGPA)	Std. Error	t	P>t	95% Confidence Interval
EAP	SS and Humanities	5.248139	0.110883	47.330	0.000	[5.030,5.465]
	Business	5.388294	0.080134	67.240	0.000	[5.231,5.545]
	STEM	5.622663	0.136938	41.060	0.000	[5.354,5.891]
IELTS Low	SS and Humanities	5.686845	0.060574	93.880	0.000	[5.568,5.805]
	Business	5.858681	0.050777	115.380	0.000	[5.759,5.958]
	STEM	5.834805	0.074041	78.800	0.000	[5.689,5.979]
TOEFL Low	SS and Humanities	5.329233	0.125115	42.590	0.000	[5.083,5.574]
	Business	5.75287	0.120657	47.680	0.000	[5.516,5.989]
	STEM	5.8217	0.132844	43.820	0.000	[5.561,6.082]
IELTS Mid	SS and Humanities	6.017556	0.050874	118.280	0.000	[5.917,6.117]
	Business	6.036983	0.057306	105.350	0.000	[5.92,6.149]
	STEM	6.074493	0.048169	126.110	0.000	[5.984,6.168]
TOEFL Mid	SS and Humanities	5.821776	0.074211	78.450	0.000	[5.676,5.967]
	Business	4.700011	0.083218	56.480	0.000	[4.536,4.863]
	STEM	5.936947	0.100379	59.150	0.000	[5.740,6.133]

IELTS High	SS and Humanities	6.500224	0.141544	45.920	0.000	[6.222,6.777]
	Business	6.1772	0.253299	24.390	0.000	[5.680,6.673]
	STEM	6.809992	0.178938	38.060	0.000	[6.459,7.160]
TOEFL High	SS and Humanities	6.584723	0.222345	29.610	0.000	[6.143,7.020]
	Business	5.071427	0.319272	15.880	0.000	[4.445,5.697]
	STEM	6.666978	0.343526	19.410	0.000	[5.993,7.340]

Significant ($p < .05$) Pairwise Comparisons for CGPA across ELP Pathways and Disciplines for University A

Pathway_Groups#Discipline	Contrast	Std. err.	t	P>t	[95% conf. interval]	
(1 1) vs (0 0)	0.611	0.12	5.07	0.000	0.168	1.053
(1 2) vs (0 0)	0.587	0.133	4.42	0.002	0.099	1.074
(3 0) vs (0 0)	0.769	0.123	6.27	0.000	0.318	1.221
(3 1) vs (0 0)	0.789	0.125	6.31	0.000	0.329	1.248
(3 2) vs (0 0)	0.826	0.122	6.76	0.000	0.377	1.276
(4 0) vs (0 0)	0.574	0.134	4.28	0.004	0.081	1.066
(4 1) vs (0 0)	-0.548	0.138	-3.97	0.015	-1.056	-0.041
(4 2) vs (0 0)	0.689	0.151	4.57	0.001	0.135	1.243
(5 0) vs (0 0)	1.252	0.181	6.92	0.000	0.587	1.917
(5 2) vs (0 0)	1.562	0.212	7.38	0.000	0.784	2.34
(6 0) vs (0 0)	1.337	0.249	5.37	0.000	0.422	2.252
(6 2) vs (0 0)	1.419	0.361	3.93	0.018	0.09	2.748
(1 1) vs (0 1)	0.47	0.092	5.11	0.000	0.132	0.809
(1 2) vs (0 1)	0.447	0.108	4.15	0.007	0.051	0.842
(3 0) vs (0 1)	0.629	0.097	6.52	0.000	0.274	0.984
(3 1) vs (0 1)	0.649	0.099	6.56	0.000	0.285	1.012
(3 2) vs (0 1)	0.686	0.096	7.17	0.000	0.334	1.038
(4 0) vs (0 1)	0.433	0.11	3.93	0.018	0.028	0.839
(4 1) vs (0 1)	-0.688	0.114	-6.02	0.000	-1.108	-0.268
(4 2) vs (0 1)	0.549	0.13	4.21	0.005	0.07	1.028
(5 0) vs (0 1)	1.112	0.165	6.75	0.000	0.507	1.717
(5 2) vs (0 1)	1.422	0.198	7.19	0.000	0.695	2.148
(6 0) vs (0 1)	1.196	0.237	5.05	0.000	0.325	2.068
(4 1) vs (0 2)	-0.923	0.16	-5.76	0.000	-1.511	-0.334

(5 0) vs (0 2)	0.878	0.197	4.45	0.002	0.153	1.602
(5 2) vs (0 2)	1.187	0.225	5.27	0.000	0.359	2.016
(6 0) vs (0 2)	0.962	0.261	3.68	0.049	0.002	1.922
(3 0) vs (1 0)	0.331	0.081	4.09	0.009	0.034	0.628
(3 1) vs (1 0)	0.35	0.084	4.18	0.006	0.042	0.658
(3 2) vs (1 0)	0.388	0.08	4.82	0.000	0.092	0.683
(4 1) vs (1 0)	-0.987	0.102	-9.69	0.000	-1.361	-0.612
(5 0) vs (1 0)	0.813	0.156	5.22	0.000	0.24	1.386
(5 2) vs (1 0)	1.123	0.191	5.89	0.000	0.422	1.824
(6 0) vs (1 0)	0.898	0.231	3.89	0.022	0.048	1.747
(2 0) vs (1 1)	-0.529	0.136	-3.88	0.022	-1.031	-0.028
(4 1) vs (1 1)	-1.159	0.096	-12.08	0.000	-1.511	-0.806
(5 0) vs (1 1)	0.642	0.153	4.2	0.006	0.08	1.203
(5 2) vs (1 1)	0.951	0.188	5.06	0.000	0.26	1.643
(4 1) vs (1 2)	-1.135	0.11	-10.28	0.000	-1.541	-0.729
(5 0) vs (1 2)	0.665	0.161	4.13	0.008	0.073	1.258
(5 2) vs (1 2)	0.975	0.195	5	0.000	0.259	1.691
(3 0) vs (2 0)	0.688	0.134	5.13	0.000	0.195	1.182
(3 1) vs (2 0)	0.708	0.138	5.15	0.000	0.202	1.213
(3 2) vs (2 0)	0.745	0.133	5.6	0.000	0.256	1.234
(4 1) vs (2 0)	-0.629	0.151	-4.17	0.006	-1.184	-0.074
(4 2) vs (2 0)	0.608	0.159	3.82	0.029	0.022	1.193
(5 0) vs (2 0)	1.171	0.188	6.24	0.000	0.481	1.861
(5 2) vs (2 0)	1.481	0.217	6.82	0.000	0.682	2.279
(6 0) vs (2 0)	1.255	0.255	4.93	0.000	0.319	2.192
(4 1) vs (2 1)	-1.053	0.146	-7.19	0.000	-1.591	-0.515
(5 0) vs (2 1)	0.747	0.186	4.01	0.013	0.063	1.432
(5 2) vs (2 1)	1.057	0.216	4.89	0.000	0.263	1.851

(4 1) vs (2 2)	-1.122	0.157	-7.12	0.000	-1.701	-0.543
(5 2) vs (2 2)	0.988	0.221	4.46	0.002	0.174	1.802
(4 1) vs (3 0)	-1.318	0.099	-13.36	0.000	-1.68	-0.955
(5 2) vs (3 0)	0.792	0.185	4.29	0.004	0.113	1.472
(4 1) vs (3 1)	-1.337	0.101	-13.2	0.000	-1.709	-0.965
(5 2) vs (3 1)	0.773	0.188	4.12	0.008	0.083	1.463
(4 1) vs (3 2)	-1.374	0.097	-14.12	0.000	-1.732	-1.017
(5 2) vs (3 2)	0.735	0.184	4.01	0.013	0.061	1.41
(4 1) vs (4 0)	-1.122	0.112	-10	0.000	-1.534	-0.709
(5 0) vs (4 0)	0.678	0.159	4.27	0.004	0.095	1.262
(5 2) vs (4 0)	0.988	0.193	5.13	0.000	0.28	1.697
(4 2) vs (4 1)	1.237	0.131	9.42	0.000	0.754	1.72
(5 0) vs (4 1)	1.8	0.165	10.88	0.000	1.192	2.408
(5 1) vs (4 1)	1.477	0.267	5.52	0.000	0.494	2.46
(5 2) vs (4 1)	2.11	0.198	10.64	0.000	1.381	2.839
(6 0) vs (4 1)	1.885	0.238	7.92	0.000	1.01	2.759
(6 2) vs (4 1)	1.967	0.354	5.56	0.000	0.667	3.267
(5 2) vs (4 2)	0.873	0.203	4.3	0.004	0.126	1.62
(6 1) vs (5 0)	-1.429	0.35	-4.09	0.009	-2.714	-0.144
(6 1) vs (5 2)	-1.739	0.366	-4.75	0.000	-3.085	-0.392

The variable "Pathway" is coded as follows: 0 = EAP, 1 = IELTS Low, 2 = TOEFL Low, 3 = IELTS Mid, 4 = TOEFL Mid, 5 = IELTS High, and 6 = TOEFL High. The "Discipline" variable is coded as: 0 = Social Sciences & Humanities, 1 = Business, and 2 = STEM.

**Appendix B: Pairwise Comparison for CGPA by ELP Pathway and Discipline for
University B**

Margin Predicted CGPA - by Pathway and Discipline for University B

Pathway Group	Discipline	Margin (Predicted CGPA)	Std. Error	t	P>t	95% Confidence Interval
EAP	SS and Humanities	2.659	0.059	44.870	0.000	2.543
	Business	3.146	0.153	20.620	0.000	2.847
	STEM	2.723	0.040	67.740	0.000	2.644
IELTS Low	SS and Humanities	2.774	0.084	32.950	0.000	2.609
	Business	2.583	0.591	4.370	0.000	1.423
	STEM	2.772	0.088	31.600	0.000	2.600
TOEFL Low	SS and Humanities	2.400	0.482	4.980	0.000	1.455
	Business	No observations				
	STEM	2.840	0.231	12.280	0.000	2.386
IELTS Mid	SS and Humanities	2.664	0.054	49.460	0.000	2.559
	Business	2.967	0.167	17.780	0.000	2.640
	STEM	2.763	0.028	99.070	0.000	2.709
TOEFL Mid	SS and Humanities	2.827	0.164	17.290	0.000	2.506
	Business	No observations				
	STEM	2.919	0.086	33.860	0.000	2.750

IELTS High	SS and Humanities	2.781	0.265	10.490	0.000	2.261
	Business	No observations				
	STEM	2.979	0.086	34.720	0.000	2.811
TOEFL High	SS and Humanities	3.464	0.416	8.330	0.000	2.648
	Business	3.930	0.838	4.690	0.000	2.287
	STEM	3.378	0.197	17.130	0.000	2.991

Only statistically significant pairwise comparisons are presented in the appendices. Therefore, interactions are not reported here as testing did not reach statistical significance and were omitted for clarity and conciseness.

Appendix C: Pairwise Comparison for CGPA by ELP Pathway and Origin for University

A

Margin Predicted CGPA - by Pathway by Origin (University A)

Pathway Group	Country of Origin	Margin (Predicted CGPA)	Std. Error	t	P>t	95% Confidence Interval
EAP	China	5.298301	0.0657286	80.610	0.000	[5.169,5.427]
	Other Countries	5.530281	0.1280815	43.180	0.000	[5.279,5.781]
IELTS Low	China	5.728277	0.0356022	160.900	0.000	[5.658,5.798]
	Other Countries	5.702196	0.0984808	57.900	0.000	[5.509,5.895]
TOEFL Low	China	5.542644	0.1152874	48.080	0.000	[5.316,5.768]
	Other Countries	5.747648	0.0933564	61.570	0.000	[5.564,5.930]
IELTS Mid	China	6.064608	0.0514242	117.930	0.000	[5.963,6.163]
	Other Countries	6.096703	0.0360482	169.130	0.000	[6.026,6.167]
TOEFL Mid	China	4.730658	0.07272	65.050	0.000	[4.588,4.873]
	Other Countries	6.116704	0.0650215	94.070	0.000	[5.989,6.241]
IELTS High	China	7.262057	0.4005608	18.130	0.000	[6.476,8.043]
	Other Countries	6.621772	0.1042662	63.510	0.000	[6.417,6.826]
TOEFL High	China	5.157096	0.2662814	19.370	0.000	[4.635,5.679]

	Other Countries	6.904893	0.2017793	34.220	0.000	[6.509,7.300]
--	--------------------	----------	-----------	--------	-------	---------------

Significant ($p < .05$) Pairwise Comparisons for CGPA across ELP Pathways and Origin for University A

Pathway_Groups#Origin	Contrast	Std. err.	t	P>t	[95% conf. interval]	
(1 0) vs (0 0)	0.43	0.074	5.82	0.000	0.174	0.685
(2 2) vs (0 0)	0.449	0.115	3.92	0.008	0.053	0.846
(3 0) vs (0 0)	0.766	0.083	9.2	0.000	0.478	1.054
(3 2) vs (0 0)	0.798	0.076	10.5	0.000	0.536	1.061
(4 0) vs (0 0)	-0.568	0.097	-5.85	0.000	-0.903	-0.232
(4 2) vs (0 0)	0.818	0.093	8.77	0.000	0.496	1.141
(5 0) vs (0 0)	1.964	0.406	4.84	0.000	0.56	3.368
(5 2) vs (0 0)	1.323	0.124	10.69	0.000	0.895	1.751
(6 2) vs (0 0)	1.607	0.213	7.56	0.000	0.872	2.341
(3 0) vs (0 2)	0.534	0.138	3.87	0.010	0.057	1.012
(3 2) vs (0 2)	0.566	0.133	4.27	0.002	0.107	1.025
(4 0) vs (0 2)	-0.8	0.147	-5.42	0.000	-1.309	-0.29
(4 2) vs (0 2)	0.586	0.143	4.09	0.004	0.09	1.082
(5 0) vs (0 2)	1.732	0.42	4.12	0.003	0.278	3.185
(5 2) vs (0 2)	1.091	0.165	6.61	0.000	0.521	1.662
(6 2) vs (0 2)	1.375	0.239	5.75	0.000	0.549	2.201
(3 0) vs (1 0)	0.336	0.062	5.39	0.000	0.121	0.552
(3 2) vs (1 0)	0.368	0.052	7.13	0.000	0.19	0.547
(4 0) vs (1 0)	-0.998	0.08	-12.4	0.000	-1.276	-0.72
(4 2) vs (1 0)	0.388	0.075	5.19	0.000	0.13	0.647
(5 0) vs (1 0)	1.534	0.402	3.81	0.013	0.143	2.924
(5 2) vs (1 0)	0.893	0.111	8.08	0.000	0.511	1.276
(6 2) vs (1 0)	1.177	0.205	5.74	0.000	0.468	1.886
(3 2) vs (1 2)	0.395	0.104	3.78	0.015	0.033	0.756
(4 0) vs (1 2)	-0.972	0.123	-7.91	0.000	-1.396	-0.547
(4 2) vs (1 2)	0.415	0.118	3.52	0.039	0.007	0.822
(5 0) vs (1 2)	1.56	0.412	3.78	0.014	0.134	2.985
(5 2) vs (1 2)	0.92	0.143	6.42	0.000	0.424	1.415
(6 2) vs (1 2)	1.203	0.224	5.36	0.000	0.427	1.979
(3 0) vs (2 0)	0.522	0.126	4.14	0.003	0.086	0.958
(3 2) vs (2 0)	0.554	0.121	4.57	0.000	0.135	0.973
(4 0) vs (2 0)	-0.812	0.136	-5.98	0.000	-1.282	-0.342
(4 2) vs (2 0)	0.574	0.133	4.32	0.001	0.115	1.033
(5 0) vs (2 0)	1.719	0.417	4.12	0.003	0.278	3.161
(5 2) vs (2 0)	1.079	0.156	6.93	0.000	0.541	1.618
(6 2) vs (2 0)	1.362	0.233	5.86	0.000	0.558	2.166

(3 2) vs (2 2)	0.349	0.1	3.5	0.042	0.005	0.693
(4 0) vs (2 2)	-1.017	0.119	-8.57	0.000	-1.427	-0.607
(5 0) vs (2 2)	1.514	0.411	3.68	0.021	0.093	2.935
(5 2) vs (2 2)	0.874	0.14	6.26	0.000	0.392	1.357
(6 2) vs (2 2)	1.157	0.222	5.21	0.000	0.389	1.925
(4 0) vs (3 0)	-1.334	0.089	-14.99	0.000	-1.642	-1.026
(5 2) vs (3 0)	0.557	0.116	4.79	0.000	0.155	0.959
(6 2) vs (3 0)	0.84	0.208	4.03	0.005	0.12	1.56
(4 0) vs (3 2)	-1.366	0.082	-16.64	0.000	-1.65	-1.082
(5 2) vs (3 2)	0.525	0.11	4.78	0.000	0.145	0.905
(6 0) vs (3 2)	-0.94	0.269	-3.49	0.043	-1.869	-0.01
(6 2) vs (3 2)	0.808	0.205	3.95	0.007	0.1	1.516
(4 2) vs (4 0)	1.386	0.098	14.13	0.000	1.047	1.725
(5 0) vs (4 0)	2.531	0.407	6.22	0.000	1.124	3.939
(5 2) vs (4 0)	1.891	0.128	14.83	0.000	1.45	2.332
(6 2) vs (4 0)	2.174	0.215	10.13	0.000	1.432	2.916
(5 2) vs (4 2)	0.505	0.122	4.13	0.003	0.082	0.928
(6 0) vs (4 2)	-0.96	0.274	-3.5	0.043	-1.907	-0.012
(6 2) vs (4 2)	0.788	0.212	3.72	0.018	0.057	1.52
(6 0) vs (5 0)	-2.105	0.481	-4.38	0.001	-3.767	-0.442
(6 0) vs (5 2)	-1.465	0.286	-5.12	0.000	-2.453	-0.476
(6 2) vs (6 0)	1.748	0.334	5.23	0.000	0.593	2.903

The "Pathway" variable is coded as follows: 0 = EAP, 1 = IELTS Low, 2 = TOEFL Low, 3 = IELTS Mid, 4 = TOEFL Mid, 5 = IELTS High, and 6 = TOEFL High. The "Origin" variable is coded as: 0 = China, and 1 = All Other Countries

Appendix D: Pairwise Comparison for CGPA by ELP Pathway and Origin for University

B

Margin Predicted CGPA – by Pathway by Origin (University B)

Pathway Group	Country of Origin	Margin (Predicted CGPA)	Std. Error	t	P>t	95% Confidence Interval
EAP	China	2.698961	0.0426331	60.760	0.000	[2.615,2.782]
	Other Countries	2.747578	0.0477856	55.250	0.000	[2.653,2.841]
IELTS Low	China	2.794624	0.0617486	44.760	0.000	[2.673,2.915]
	Other Countries	2.723830	0.4804079	5.750	0.000	[1.781,3.665]
TOEFL Low	China	2.817883	0.2310677	12.120	0.000	[2.364,3.270]
	Other Countries	2.502053	0.4818709	4.950	0.000	[1.557,3.446]
IELTS Mid	China	2.748278	0.0287372	92.350	0.000	[2.691,2.804]
	Other Countries	2.733328	0.0481109	56.070	0.000	[2.638,2.827]
TOEFL Mid	China	2.905979	0.1076232	26.850	0.000	[2.694,3.117]
	Other Countries	2.903249	0.1074386	26.940	0.000	[2.692,3.113]
IELTS High	China	3.513229	0.3715939	9.440	0.000	[2.784,4.241]
	Other Countries	2.932974	0.0807811	36.050	0.000	[2.774,3.091]
TOEFL High	China	3.649544	0.4814365	6.990	0.000	[2.705,4.593]

	Other Countries	3.370978	0.1857192	18.170	0.000	[3.006,3.735]
--	--------------------	----------	-----------	--------	-------	---------------

Significant ($p < .05$) Pairwise Comparisons for CGPA across ELP Pathways and Origin for University B

Pathway	Groups#Origin	Contrast	Std. err.	t	P>t	[95% conf. interval]
(6 1) vs (0 1)		0.687	0.178	3.87	0.010	0.072 1.302

The "Pathway" variable is coded as follows: 0 = EAP, 1 = IELTS Low, 2 = TOEFL Low, 3 = IELTS Mid, 4 = TOEFL Mid, 5 = IELTS High, and 6 = TOEFL High. The "Origin" variable is coded as: 0 = China, and 1 = All Other Countries.

Appendix E: Pairwise Comparison for CGPA by ELP Pathway and Gender for University

A

Margin Predicted CGPA – by Pathway by Gender (University A)

Pathway Group	Gender	Predicted CGPA	Std. Error	t	P>t	95% Confidence Interval
EAP	Men	5.2222	0.0826	47.330	0.000	[5.060, 5.384]
	Women	5.6317	0.0844	67.240	0.000	[5.466, 5.797]
IELTS Low	Men	5.6784	0.0504	41.060	0.000	[5.579, 5.777]
	Women	5.9848	0.0483	93.880	0.000	[5.890, 6.079]
TOEFL Low	Men	5.4052	0.1124	115.380	0.000	[5.184, 5.625]
	Women	5.8128	0.0964	78.800	0.000	[5.623, 6.001]
IELTS Mid	Men	5.8268	0.0438	42.590	0.000	[5.740, 5.912]
	Women	6.1818	0.0415	47.680	0.000	[6.100, 6.263]
TOEFL Mid	Men	5.2156	0.0680	43.820	0.000	[5.082, 5.348]
	Women	5.7097	0.0697	118.280	0.000	[5.573, 5.846]
IELTS High	Men	6.3811	0.1582	105.350	0.000	[6.070, 6.691]
	Women	6.6252	0.1349	126.110	0.000	[6.360, 6.889]
TOEFL High	Men	6.4391	0.2313	78.450	0.000	[5.985, 6.892]
	Women	5.9765	0.2277	56.480	0.000	[5.530, 6.422]

Significant ($p < .05$) Pairwise Comparisons for CGPA across ELP Pathways and Gender for University A

Pathway_Groups#Gender	Contrast	Std. err.	t	P>t	[95% conf. interval]	
(0 1) vs (0 0)	0.41	0.117	3.5	0.0420	0.005	0.814
(1 0) vs (0 0)	0.456	0.095	4.81	0.0000	0.129	0.784
(1 1) vs (0 0)	0.763	0.094	8.12	0.0000	0.438	1.087
(2 1) vs (0 0)	0.591	0.128	4.63	0.0000	0.149	1.032
(3 0) vs (0 0)	0.605	0.095	6.38	0.0000	0.277	0.932
(3 1) vs (0 0)	0.96	0.094	10.24	0.0000	0.636	1.284
(4 1) vs (0 0)	0.488	0.109	4.48	0.0010	0.111	0.864
(5 0) vs (0 0)	1.159	0.18	6.44	0.0000	0.537	1.781
(5 1) vs (0 0)	1.403	0.16	8.78	0.0000	0.851	1.955
(6 0) vs (0 0)	1.217	0.246	4.95	0.0000	0.367	2.067
(1 1) vs (0 1)	0.353	0.096	3.69	0.0210	0.022	0.684
(3 1) vs (0 1)	0.55	0.095	5.78	0.0000	0.221	0.879
(4 0) vs (0 1)	-0.416	0.108	-3.84	0.0110	-0.791	-0.041
(5 0) vs (0 1)	0.749	0.18	4.15	0.0030	0.126	1.373
(5 1) vs (0 1)	0.993	0.16	6.19	0.0000	0.439	1.548
(1 1) vs (1 0)	0.306	0.067	4.59	0.0000	0.076	0.537
(3 1) vs (1 0)	0.503	0.067	7.46	0.0000	0.27	0.737
(4 0) vs (1 0)	-0.463	0.085	-5.45	0.0000	-0.756	-0.169
(5 0) vs (1 0)	0.703	0.168	4.19	0.0030	0.122	1.283
(5 1) vs (1 0)	0.947	0.146	6.47	0.0000	0.441	1.453
(2 0) vs (1 1)	-0.58	0.123	-4.72	0.0000	-1.004	-0.155
(4 0) vs (1 1)	-0.769	0.084	-9.19	0.0000	-1.058	-0.48
(5 1) vs (1 1)	0.64	0.145	4.41	0.0010	0.138	1.143
(3 0) vs (2 0)	0.422	0.12	3.5	0.0420	0.006	0.838
(3 1) vs (2 0)	0.777	0.12	6.5	0.0000	0.363	1.19
(5 0) vs (2 0)	0.976	0.194	5.04	0.0000	0.306	1.646
(5 1) vs (2 0)	1.22	0.175	6.96	0.0000	0.615	1.825
(6 0) vs (2 0)	1.034	0.257	4.02	0.0050	0.145	1.922
(3 1) vs (2 1)	0.369	0.104	3.54	0.0360	0.009	0.729
(4 0) vs (2 1)	-0.597	0.118	-5.07	0.0000	-1.004	-0.19
(5 1) vs (2 1)	0.812	0.165	4.93	0.0000	0.243	1.382
(3 1) vs (3 0)	0.355	0.059	6.03	0.0000	0.152	0.558
(4 0) vs (3 0)	-0.611	0.081	-7.55	0.0000	-0.891	-0.332
(5 1) vs (3 0)	0.798	0.141	5.68	0.0000	0.313	1.284
(4 0) vs (3 1)	-0.966	0.079	-12.16	0.0000	-1.241	-0.691
(4 1) vs (3 1)	-0.472	0.08	-5.89	0.0000	-0.749	-0.195
(4 1) vs (4 0)	0.494	0.097	5.09	0.0000	0.159	0.83
(5 0) vs (4 0)	1.165	0.172	6.77	0.0000	0.571	1.76
(5 1) vs (4 0)	1.41	0.151	9.35	0.0000	0.889	1.931
(6 0) vs (4 0)	1.223	0.241	5.08	0.0000	0.391	2.056
(5 0) vs (4 1)	0.671	0.172	3.9	0.0090	0.077	1.266
(5 1) vs (4 1)	0.916	0.151	6.08	0.0000	0.395	1.436

The "Pathway" variable is coded as follows: 0 = EAP, 1 = IELTS Low, 2 = TOEFL Low, 3 = IELTS Mid, 4 = TOEFL Mid, 5 = IELTS High, and 6 = TOEFL High. The "Gender" variable is coded as follows: 0 = men, and 1 = women.

Appendix F: Pairwise Comparison for CGPA by ELP Pathway and Gender for University

B

Margin Predicted CGPA - Pathway by Gender (University B)

Pathway Group	Gender	Predicted CGPA	Std. Error	t	P>t	95% Confidence Interval
EAP	Men	2.6720	0.0415	61.110	0.000	[2.590, 2.753]
	Women	2.7839	0.0499	54.630	0.000	[2.685, 2.881]
IELTS Low	Men	2.8972	0.0900	32.060	0.000	[2.720, 3.073]
	Women	2.7372	0.0823	33.090	0.000	[2.575, 2.898]
TOEFL Low	Men	2.4705	0.2783	8.700	0.000	[1.924, 3.016]
	Women	3.1429	0.3139	9.990	0.000	[2.527, 3.758]
IELTS Mid	Men	2.6975	0.0324	81.080	0.000	[2.634, 2.761]
	Women	2.8141	0.0381	72.630	0.000	[2.739, 2.888]
TOEFL Mid	Men	2.8874	0.1007	28.460	0.000	[2.690, 3.084]
	Women	2.9295	0.1166	25.060	0.000	[2.700, 3.158]
IELTS High	Men	2.9228	0.0986	29.300	0.000	[2.729, 3.116]
	Women	2.9980	0.1404	21.220	0.000	[2.722, 3.273]
TOEFL High	Men	3.2743	0.2230	14.670	0.000	[2.837, 3.711]
	Women	3.6022	0.2773	12.620	0.000	[3.058, 4.146]

Only statistically significant results are presented in appendices. Therefore, contrasts are not reported here as testing did not reach statistical significance and were omitted for clarity and conciseness.

Appendix G: Pairwise Comparison for CCE by ELP Pathway and Gender for University A

Margin Predicted CCE - Pathway by Discipline (University A)

Pathway Group	Discipline	Margin (Predicted CCE)	Std. Error	t	P>t	95% Confidence Interval
EAP	SS and Humanities	8.471	0.265	31.950	0.000	[7.951, 8.990]
	Business	8.423	0.192	43.960	0.000	[8.048, 8.799]
	STEM	8.425	0.327	25.730	0.000	[7.783, 9.066]
IELTS Low	SS and Humanities	10.400	0.145	71.810	0.000	[10.117,10.684]
	Business	10.321	0.121	85.020	0.000	[10.083,10.559]
	STEM	9.501	0.177	53.670	0.000	[9.154, 9.848]
TOEFL Low	SS and Humanities	10.350	0.299	34.600	0.000	[9.763, 10.936]
	Business	10.254	0.288	35.550	0.000	[9.689, 10.820]
	STEM	9.953	0.318	31.340	0.000	[9.330, 10.576]
IELTS Mid	SS and Humanities	11.807	0.122	97.070	0.000	[11.568,12.045]
	Business	10.924	0.137	79.730	0.000	[10.656,11.193]
	STEM	10.361	0.115	89.960	0.000	[10.135,10.586]
TOEFL Mid	SS and Humanities	10.836	0.177	61.070	0.000	[10.489,11.184]
	Business	8.480	0.199	42.620	0.000	[8.090, 8.870]
	STEM	10.461	0.240	43.590	0.000	[9.991, 10.931]
IELTS High	SS and Humanities	12.311	0.338	36.380	0.000	[11.648,12.974]

	Business	10.406	0.606	17.180	0.000	[9.219, 11.593]
	STEM	11.151	0.428	26.070	0.000	[10.313, 11.990]
TOEFL High	SS and Humanities	11.501	0.532	21.630	0.000	[10.459, 12.543]
	Business	9.152	0.763	11.990	0.000	[7.655, 10.648]
	STEM	11.050	0.821	13.450	0.000	[9.440, 12.660]

Significant ($p < .05$) Pairwise Comparisons for CCE across ELP Pathways and Discipline for University A

Pathway_Groups#Discipline	Contrast	Std. err.	t	P>t	[95% conf. interval]	
(1 0) vs (0 0)	3.728	0.437	8.54	0.000	2.12	5.335
(1 2) vs (0 0)	3.451	0.451	7.65	0.000	1.79	5.113
(5 2) vs (0 0)	2.019	0.449	4.5	0.001	0.368	3.67
(1 0) vs (0 1)	3.127	0.734	4.26	0.004	0.426	5.828
(1 2) vs (0 1)	2.851	0.742	3.84	0.026	0.121	5.58
(1 0) vs (0 2)	4.004	0.407	9.84	0.000	2.506	5.502
(1 2) vs (0 2)	3.728	0.42	8.87	0.000	2.182	5.274
(3 2) vs (0 2)	1.205	0.216	5.57	0.000	0.409	2.001
(5 2) vs (0 2)	2.296	0.396	5.8	0.000	0.838	3.754
(3 0) vs (1 0)	-4.635	0.413	-11.23	0.000	-6.154	-3.116
(3 1) vs (1 0)	-4.164	0.778	-5.35	0.000	-7.026	-1.302
(3 2) vs (1 0)	-2.799	0.372	-7.53	0.000	-4.167	-1.432
(4 0) vs (1 0)	-3.642	0.774	-4.71	0.001	-6.491	-0.794
(4 2) vs (1 0)	-2.602	0.512	-5.09	0.000	-4.486	-0.719
(3 0) vs (1 2)	-4.359	0.427	-10.2	0.000	-5.931	-2.787
(3 1) vs (1 2)	-3.888	0.786	-4.94	0.000	-6.782	-0.994
(3 2) vs (1 2)	-2.523	0.385	-6.55	0.000	-3.94	-1.107
(4 0) vs (1 2)	-3.366	0.783	-4.3	0.004	-6.246	-0.486
(4 2) vs (1 2)	-2.326	0.521	-4.46	0.002	-4.244	-0.408
(3 0) vs (2 2)	-3.812	1	-3.81	0.030	-7.491	-0.132
(3 2) vs (3 0)	1.836	0.253	7.27	0.000	0.906	2.765
(4 2) vs (3 0)	2.032	0.431	4.71	0.001	0.446	3.619
(5 2) vs (3 0)	2.926	0.436	6.71	0.000	1.322	4.53

The variable "Pathway" is coded as follows: 0 = EAP, 1 = IELTS Low, 2 = TOEFL Low, 3 = IELTS Mid, 4 = TOEFL Mid, 5 = IELTS High, and 6 = TOEFL High. The "Discipline" variable is coded as: 0 = Social Sciences & Humanities, 1 = Business, and 2 = STEM.

Appendix H: Pairwise Comparison for CCE by ELP Pathway and Discipline for University

B

Margin Predicted CCE - Pathway by Discipline (University B)

Pathway Group	Discipline	Margin (Predicted CCE)	Std. Error	t	P>t	95% Confidence Interval
EAP	SS and Humanities	12.595	0.251	50.270	0.000	12.103
	Business	13.196	0.645	20.450	0.000	11.931
	STEM	12.318	0.170	72.480	0.000	11.985
IELTS Low	SS and Humanities	16.322	0.356	45.870	0.000	15.625
	Business	15.093	2.500	6.040	0.000	10.190
	STEM	16.046	0.371	43.270	0.000	15.319
TOEFL Low	SS and Humanities	10.470	2.037	5.140	0.000	6.477
	Business	No observations				
	STEM	15.499	0.978	15.850	0.000	13.582
IELTS Mid	SS and Humanities	11.687	0.228	51.330	0.000	11.241
	Business	12.158	0.706	17.230	0.000	10.774
	STEM	13.523	0.118	114.680	0.000	13.292
TOEFL Mid	SS and Humanities	12.680	0.691	18.340	0.000	11.324
	Business	No observations				
	STEM	13.720	0.364	37.640	0.000	13.005
IELTS High	SS and Humanities	13.200	1.120	11.780	0.000	11.003

	Business	No observations				
	STEM	14.614	0.363	40.290	0.000	13.902
TOEFL High	SS and Humanities	13.395	1.758	7.620	0.000	9.947
	Business	14.425	3.542	4.070	0.000	7.480
	STEM	14.585	0.834	17.490	0.000	12.950

Significant ($p < .05$) Pairwise Comparisons for CCE across ELP Pathways and Discipline for University B

Pathway_Groups#Discipline	Contrast	Std. err.	t	P>t	[95% conf. interval]	
(1 0) vs (0 0)	3.728	0.437	8.54	0.000	2.12	5.335
(1 2) vs (0 0)	3.451	0.451	7.65	0.000	1.79	5.113
(5 2) vs (0 0)	2.019	0.449	4.5	0.001	0.368	3.67
(1 0) vs (0 1)	3.127	0.734	4.26	0.004	0.426	5.828
(1 2) vs (0 1)	2.851	0.742	3.84	0.026	0.121	5.58
(1 0) vs (0 2)	4.004	0.407	9.84	0.000	2.506	5.502
(1 2) vs (0 2)	3.728	0.42	8.87	0.000	2.182	5.274
(3 2) vs (0 2)	1.205	0.216	5.57	0.000	0.409	2.001
(5 2) vs (0 2)	2.296	0.396	5.8	0.000	0.838	3.754
(3 0) vs (1 0)	-4.635	0.413	-11.23	0.000	-6.154	-3.116
(3 1) vs (1 0)	-4.164	0.778	-5.35	0.000	-7.026	-1.302
(3 2) vs (1 0)	-2.799	0.372	-7.53	0.000	-4.167	-1.432
(4 0) vs (1 0)	-3.642	0.774	-4.71	0.001	-6.491	-0.794
(4 2) vs (1 0)	-2.602	0.512	-5.09	0.000	-4.486	-0.719
(3 0) vs (1 2)	-4.359	0.427	-10.2	0.000	-5.931	-2.787
(3 1) vs (1 2)	-3.888	0.786	-4.94	0.000	-6.782	-0.994
(3 2) vs (1 2)	-2.523	0.385	-6.55	0.000	-3.94	-1.107
(4 0) vs (1 2)	-3.366	0.783	-4.3	0.004	-6.246	-0.486
(4 2) vs (1 2)	-2.326	0.521	-4.46	0.002	-4.244	-0.408
(3 0) vs (2 2)	-3.812	1	-3.81	0.030	-7.491	-0.132
(3 2) vs (3 0)	1.836	0.253	7.27	0.000	0.906	2.765
(4 2) vs (3 0)	2.032	0.431	4.71	0.001	0.446	3.619
(5 2) vs (3 0)	2.926	0.436	6.71	0.000	1.322	4.53

The variable "Pathway" is coded as follows: 0 = EAP, 1 = IELTS Low, 2 = TOEFL Low, 3 = IELTS Mid, 4 = TOEFL Mid, 5 = IELTS High, and 6 = TOEFL High. The "Discipline" variable is coded as: 0 = Social Sciences & Humanities, 1 = Business, and 2 = STEM.

Appendix I: Pairwise Comparison for CCE by ELP Pathway and Origin for University A

Margin Predicted CCE - Pathway by Origin (University A)

Pathway Group	Country of Origin	Margin (Predicted CCE)	Std. Error	t	P>t	95% Confidence Interval
EAP	China	7.841	0.157	49.910	0.000	[7.533,8.149]
	Other Countries	9.249	0.306	30.210	0.000	[8.648,9.849]
IELTS Low	China	9.668	0.085	113.610	0.000	[9.502,9.835]
	Other Countries	10.340	0.235	43.930	0.000	[9.878,10.801]
TOEFL Low	China	9.330	0.276	33.860	0.000	[8.790, 9.870]
	Other Countries	11.077	0.223	49.640	0.000	[10.639,11.514]
IELTS Mid	China	10.900	0.123	88.680	0.000	[10.659,11.141]
	Other Countries	11.518	0.086	133.680	0.000	[11.350,11.687]
TOEFL Mid	China	8.058	0.174	46.360	0.000	[7.717, 8.399]
	Other Countries	11.556	0.155	74.350	0.000	[11.251,11.860]
IELTS High	China	12.365	0.957	12.910	0.000	[10.488,14.242]
	Other Countries	12.112	0.249	48.600	0.000	[11.623,12.600]
TOEFL High	China	8.259	0.636	12.980	0.000	[7.011, 9.506]
	Other Countries	12.385	0.482	25.680	0.000	[11.439,13.330]

Significant ($p < .05$) Pairwise Comparisons for CCE across ELP Pathways and Origin for University A

Pathway_Groups#Origin	Contrast	Std. err.	t	P>t	[95% conf. interval]	
(0 2) vs (0 0)	1.408	0.345	4.08	0.004	0.217	2.6
(1 0) vs (0 0)	1.828	0.177	10.35	0.000	1.217	2.438
(1 2) vs (0 0)	2.499	0.284	8.8	0.000	1.518	3.481
(2 0) vs (0 0)	1.49	0.316	4.71	0.000	0.397	2.582
(2 2) vs (0 0)	3.236	0.274	11.81	0.000	2.289	4.184
(3 0) vs (0 0)	3.059	0.199	15.37	0.000	2.371	3.748
(3 2) vs (0 0)	3.678	0.182	20.24	0.000	3.05	4.306
(4 2) vs (0 0)	3.715	0.223	16.66	0.000	2.944	4.486
(5 0) vs (0 0)	4.524	0.971	4.66	0.000	1.169	7.88
(5 2) vs (0 0)	4.271	0.296	14.43	0.000	3.248	5.294
(6 2) vs (0 0)	4.544	0.508	8.94	0.000	2.788	6.3
(2 2) vs (0 2)	1.828	0.378	4.83	0.000	0.52	3.136
(3 0) vs (0 2)	1.651	0.33	5	0.000	0.511	2.792
(3 2) vs (0 2)	2.27	0.317	7.15	0.000	1.173	3.367
(4 2) vs (0 2)	2.307	0.343	6.73	0.000	1.122	3.493
(5 2) vs (0 2)	2.863	0.395	7.26	0.000	1.499	4.227
(6 2) vs (0 2)	3.136	0.571	5.49	0.000	1.162	5.11
(2 2) vs (1 0)	1.408	0.24	5.87	0.000	0.579	2.238
(3 0) vs (1 0)	1.232	0.149	8.26	0.000	0.716	1.747
(3 2) vs (1 0)	1.85	0.124	14.98	0.000	1.423	2.277
(4 0) vs (1 0)	-1.61	0.192	-8.37	0.000	-2.275	-0.946
(4 2) vs (1 0)	1.887	0.179	10.56	0.000	1.269	2.505
(5 2) vs (1 0)	2.443	0.264	9.24	0.000	1.53	3.357
(6 2) vs (1 0)	2.716	0.49	5.54	0.000	1.022	4.411
(3 2) vs (1 2)	1.179	0.25	4.72	0.000	0.316	2.042
(4 0) vs (1 2)	-2.282	0.293	-7.78	0.000	-3.296	-1.268
(4 2) vs (1 2)	1.216	0.281	4.32	0.001	0.243	2.189

(5 2) vs (1 2)	1.772	0.342	5.17	0.000	0.588	2.955
(6 2) vs (1 2)	2.045	0.536	3.81	0.013	0.19	3.899
(2 2) vs (2 0)	1.747	0.355	4.92	0.000	0.519	2.974
(3 0) vs (2 0)	1.57	0.302	5.21	0.000	0.528	2.612
(3 2) vs (2 0)	2.188	0.29	7.55	0.000	1.187	3.19
(4 0) vs (2 0)	-1.272	0.325	-3.92	0.008	-2.395	-0.15
(4 2) vs (2 0)	2.226	0.317	7.01	0.000	1.129	3.323
(5 2) vs (2 0)	2.781	0.372	7.47	0.000	1.494	4.068
(6 2) vs (2 0)	3.055	0.556	5.49	0.000	1.133	4.977
(4 0) vs (2 2)	-3.019	0.284	-10.65	0.000	-3.999	-2.038
(6 0) vs (2 2)	-2.818	0.675	-4.18	0.003	-5.15	-0.486
(3 2) vs (3 0)	0.618	0.151	4.11	0.004	0.098	1.139
(4 0) vs (3 0)	-2.842	0.213	-13.36	0.000	-3.577	-2.107
(5 2) vs (3 0)	1.212	0.278	4.36	0.001	0.25	2.173
(6 0) vs (3 0)	-2.641	0.648	-4.07	0.004	-4.882	-0.401
(4 0) vs (3 2)	-3.46	0.196	-17.64	0.000	-4.139	-2.782
(6 0) vs (3 2)	-3.26	0.643	-5.07	0.000	-5.482	-1.038
(4 2) vs (4 0)	3.498	0.234	14.92	0.000	2.687	4.308
(5 0) vs (4 0)	4.307	0.973	4.43	0.001	0.943	7.671
(5 2) vs (4 0)	4.054	0.305	13.3	0.000	3	5.107
(6 2) vs (4 0)	4.327	0.513	8.43	0.000	2.553	6.1
(6 0) vs (4 2)	-3.297	0.655	-5.03	0.000	-5.562	-1.032
(6 0) vs (5 0)	-4.106	1.15	-3.57	0.032	-8.08	-0.133
(6 0) vs (5 2)	-3.853	0.684	-5.64	0.000	-6.216	-1.49
(6 2) vs (6 0)	4.126	0.799	5.17	0.000	1.366	6.886

The "Pathway" variable is coded as follows: 0 = EAP, 1 = IELTS Low, 2 = TOEFL Low, 3 = IELTS Mid, 4 = TOEFL Mid, 5 = IELTS High, and 6 = TOEFL High. The "Origin" variable is coded as: 0 = China, and 1 = All Other Countries

Appendix J: Pairwise Comparison for CCE by ELP Pathway and Origin for University B

Margin Predicted CCE - Pathway by Origin (University B)

Pathway Group	Country of Origin	Margin (Predicted CCE)	Std. Error	t	P>t	95% Confidence Interval
EAP	China	12.261	0.182	67.540	0.000	[11.905,12.617]
	Other Countries	12.891	0.203	63.350	0.000	[12.492,13.290]
IELTS Low	China	16.151	0.263	61.430	0.000	[15.636,16.667]
	Other Countries	12.965	2.046	6.340	0.000	[8.954, 16.977]
TOEFL Low	China	14.619	0.984	14.860	0.000	[12.690,16.549]
	Other Countries	13.179	2.052	6.420	0.000	[9.156, 17.203]
IELTS Mid	China	12.724	0.122	103.980	0.000	[12.484,12.964]
	Other Countries	13.790	0.205	67.310	0.000	[13.388,14.192]
TOEFL Mid	China	12.847	0.458	28.030	0.000	[11.948,13.746]
	Other Countries	14.356	0.457	31.380	0.000	[13.459,15.253]
IELTS High	China	13.761	1.582	8.700	0.000	[10.658,16.864]

	Other Countries	14.954	0.344	43.470	0.000	[14.279,15.629]
TOEFL High	China	15.248	2.050	7.440	0.000	[11.228,19.268]
	Other Countries	14.722	0.791	18.620	0.000	[13.172,16.273]

Significant ($p < .05$) Pairwise Comparisons for CCE across ELP Pathways and Origin for University B

Pathway_Groups#Origin	Contrast	Std. err.	t	P>t	[95% conf. interval]	
(1 0) vs (0 0)	3.89	0.325	11.96	0.000	2.764	5.016
(3 1) vs (0 0)	1.529	0.281	5.44	0.000	0.557	2.501
(4 1) vs (0 0)	2.095	0.496	4.23	0.002	0.38	3.81
(5 1) vs (0 0)	2.693	0.395	6.83	0.000	1.328	4.058
(1 0) vs (0 1)	3.26	0.346	9.42	0.000	2.062	4.458
(5 1) vs (0 1)	2.063	0.403	5.12	0.000	0.669	3.457
(3 0) vs (1 0)	-3.427	0.284	-12.06	0.000	-4.41	-2.444
(3 1) vs (1 0)	-2.361	0.334	-7.06	0.000	-3.519	-1.204
(4 0) vs (1 0)	-3.304	0.524	-6.3	0.000	-5.118	-1.491
(3 1) vs (3 0)	1.066	0.236	4.51	0.001	0.248	1.883
(4 1) vs (3 0)	1.632	0.472	3.45	0.051	-0.003	3.266
(5 1) vs (3 0)	2.23	0.363	6.14	0.000	0.974	3.486
(5 1) vs (4 0)	2.107	0.572	3.68	0.021	0.128	4.086

The "Pathway" variable is coded as follows: 0 = EAP, 1 = IELTS Low, 2 = TOEFL Low, 3 = IELTS Mid, 4 = TOEFL Mid, 5 = IELTS High, and 6 = TOEFL High. The "Origin" variable is coded as: 0 = China, and 1 = All Other Countries

Appendix K: Pairwise Comparison for CCE by ELP Pathway and Gender for University A

Margin Predicted CCE - Pathway by Gender (University A)

Pathway Group	Gender	Predicted CCE	Std. Error	t	P>t	95% Confidence Interval
EAP	Men	8.026	0.197	40.780	0.000	[7.641, 8.412]
	Women	8.918	0.201	44.290	0.000	[8.523, 9.313]
IELTS Low	Men	9.782	0.120	81.450	0.000	[9.546,10.017]
	Women	10.615	0.115	92.090	0.000	[10.389,10.84]
TOEFL Low	Men	9.696	0.268	36.180	0.000	[9.170, 10.221]
	Women	10.617	0.230	46.180	0.000	[10.166,11.06]
IELTS Mid	Men	10.724	0.104	102.670	0.000	[10.519,10.92]
	Women	11.276	0.099	113.900	0.000	[11.082,11.47]
TOEFL Mid	Men	9.353	0.162	57.700	0.000	[9.035, 9.671]
	Women	10.355	0.166	62.340	0.000	[10.029,10.681]
IELTS High	Men	11.463	0.377	30.390	0.000	[10.724,12.202]
	Women	11.536	0.322	35.870	0.000	[10.906,12.162]
TOEFL High	Men	11.339	0.551	20.560	0.000	[10.258,12.424]
	Women	9.933	0.543	18.300	0.000	[8.869, 10.997]

Significant ($p < .05$) Pairwise Comparisons for CCE across ELP Pathways and Gender for University A

Pathway_Groups#Gender	Contrast	Std. err.	t	P>t	[95% conf. interval]	
(1 0) vs (0 0)	1.755	0.226	7.77	0.000	0.974	2.537
(1 1) vs (0 0)	2.588	0.224	11.56	0.000	1.814	3.362
(2 0) vs (0 0)	1.669	0.333	5.02	0.000	0.519	2.82
(2 1) vs (0 0)	2.59	0.304	8.51	0.000	1.538	3.642
(3 0) vs (0 0)	2.698	0.226	11.94	0.000	1.917	3.478
(3 1) vs (0 0)	3.25	0.223	14.55	0.000	2.478	4.022
(4 0) vs (0 0)	1.327	0.255	5.2	0.000	0.445	2.208
(4 1) vs (0 0)	2.328	0.26	8.97	0.000	1.431	3.225
(5 0) vs (0 0)	3.437	0.429	8.01	0.000	1.955	4.92
(5 1) vs (0 0)	3.51	0.381	9.21	0.000	2.193	4.827
(6 0) vs (0 0)	3.313	0.586	5.65	0.000	1.287	5.339
(1 0) vs (0 1)	0.864	0.231	3.74	0.017	0.065	1.662
(1 1) vs (0 1)	1.696	0.228	7.43	0.000	0.907	2.486
(2 1) vs (0 1)	1.699	0.307	5.53	0.000	0.637	2.76
(3 0) vs (0 1)	1.806	0.23	7.87	0.000	1.012	2.599
(3 1) vs (0 1)	2.358	0.227	10.4	0.000	1.574	3.142
(4 1) vs (0 1)	1.436	0.262	5.48	0.000	0.53	2.343
(5 0) vs (0 1)	2.545	0.43	5.92	0.000	1.058	4.033
(5 1) vs (0 1)	2.618	0.382	6.85	0.000	1.296	3.94
(6 0) vs (0 1)	2.421	0.588	4.12	0.003	0.39	4.452
(1 1) vs (1 0)	0.833	0.159	5.23	0.000	0.283	1.383
(3 0) vs (1 0)	0.942	0.164	5.75	0.000	0.376	1.508
(3 1) vs (1 0)	1.494	0.161	9.29	0.000	0.938	2.05
(5 0) vs (1 0)	1.682	0.4	4.2	0.002	0.298	3.065
(5 1) vs (1 0)	1.754	0.349	5.03	0.000	0.548	2.96
(3 1) vs (1 1)	0.662	0.157	4.22	0.002	0.119	1.204
(4 0) vs (1 1)	-1.262	0.2	-6.32	0.000	-1.951	-0.572

(3 0) vs (2 0)	1.028	0.287	3.58	0.031	0.037	2.02
(3 1) vs (2 0)	1.581	0.285	5.54	0.000	0.595	2.566
(5 0) vs (2 0)	1.768	0.462	3.83	0.012	0.171	3.365
(5 1) vs (2 0)	1.841	0.418	4.41	0.001	0.397	3.284
(4 0) vs (2 1)	-1.264	0.281	-4.5	0.001	-2.234	-0.293
(3 1) vs (3 0)	0.552	0.14	3.94	0.008	0.067	1.037
(4 0) vs (3 0)	-1.371	0.193	-7.11	0.000	-2.038	-0.704
(4 0) vs (3 1)	-1.923	0.19	-10.15	0.000	-2.578	-1.268
(4 1) vs (3 1)	-0.922	0.191	-4.82	0.000	-1.583	-0.261
(4 1) vs (4 0)	1.002	0.231	4.33	0.001	0.202	1.802
(5 0) vs (4 0)	2.111	0.41	5.14	0.000	0.692	3.529
(5 1) vs (4 0)	2.183	0.359	6.07	0.000	0.941	3.425
(6 0) vs (4 0)	1.986	0.575	3.46	0.050	0	3.972

The "Pathway" variable is coded as follows: 0 = EAP, 1 = IELTS Low, 2 = TOEFL Low, 3 = IELTS Mid, 4 = TOEFL Mid, 5 = IELTS High, and 6 = TOEFL High. The "Gender" variable is coded as follows: 0 = men, and 1 = women.

Appendix L: Pairwise Comparison for CCE by ELP Pathway and Gender for University B

Margin Predicted CCE - Pathway by Gender (University B)

Pathway Group	Gender	Predicted CCE	Std. Error	t	P>t	95% Confidence Interval
EAP	Men	12.273	0.177	69.520	0.000	[11.927,12.61]
	Women	12.738	0.212	59.950	0.000	[12.321,13.154]
IELTS Low	Men	16.613	0.383	43.380	0.000	[15.862,17.364]
	Women	16.144	0.350	46.140	0.000	[15.458,16.830]
TOEFL Low	Men	13.207	1.184	11.150	0.000	[10.885,15.528]
	Women	16.140	1.335	12.090	0.000	[13.521,18.758]
IELTS Mid	Men	13.330	0.138	96.770	0.000	[13.060,13.600]
	Women	12.723	0.162	78.410	0.000	[12.404,13.041]
TOEFL Mid	Men	13.993	0.428	32.680	0.000	[13.153,14.833]
	Women	12.765	0.496	25.730	0.000	[11.792,13.738]
IELTS High	Men	14.457	0.419	34.480	0.000	[13.635,15.279]
	Women	14.299	0.597	23.950	0.000	[13.128,15.470]
TOEFL High	Men	14.274	0.948	15.050	0.000	[12.414,16.134]
	Women	14.473	1.180	12.270	0.000	[12.159,16.786]

Significant ($p < .05$) Pairwise Comparisons for CCE across ELP Pathways and Gender for University B

Pathway_Groups#Gender	Contrast	Std. err.	t	P>t	[95% conf. interval]	
(1 0) vs (0 0)	4.34	0.435	9.98	0.000	2.836	5.844
(1 1) vs (0 0)	3.871	0.402	9.63	0.000	2.48	5.262
(3 0) vs (0 0)	1.057	0.233	4.53	0.001	0.249	1.864
(4 0) vs (0 0)	1.72	0.464	3.71	0.020	0.115	3.325
(5 0) vs (0 0)	2.184	0.453	4.82	0.000	0.615	3.752
(1 0) vs (0 1)	3.875	0.446	8.69	0.000	2.332	5.418
(1 1) vs (0 1)	3.407	0.412	8.28	0.000	1.983	4.831
(5 0) vs (0 1)	1.719	0.475	3.62	0.028	0.075	3.364
(3 0) vs (1 0)	-3.283	0.403	-8.15	0.000	-4.676	-1.89
(3 1) vs (1 0)	-3.89	0.408	-9.54	0.000	-5.302	-2.479
(4 0) vs (1 0)	-2.62	0.577	-4.54	0.001	-4.615	-0.624
(4 1) vs (1 0)	-3.848	0.623	-6.18	0.000	-6.002	-1.694
(5 0) vs (1 0)	-2.156	0.575	-3.75	0.016	-4.145	-0.167
(3 0) vs (1 1)	-2.814	0.375	-7.5	0.000	-4.113	-1.516
(3 1) vs (1 1)	-3.422	0.379	-9.04	0.000	-4.732	-2.112
(4 0) vs (1 1)	-2.151	0.558	-3.86	0.011	-4.081	-0.221
(4 1) vs (1 1)	-3.379	0.603	-5.6	0.000	-5.466	-1.292
(5 0) vs (3 1)	1.734	0.455	3.81	0.013	0.16	3.309

The "Pathway" variable is coded as follows: 0 = EAP, 1 = IELTS Low, 2 = TOEFL Low, 3 = IELTS Mid, 4 = TOEFL Mid, 5 = IELTS High, and 6 = TOEFL High. The "Gender" variable is coded as follows: 0 = men, and 1 = women.