ENGLISH LANGUAGE TESTING AND THE INTERNATIONALIZATION OF THE AMERICAN UNIVERSITY

By

Nathan Michael Johnson

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Higher, Adult, and Lifelong Education—Doctor of Philosophy

2022

ABSTRACT

ENGLISH LANGUAGE TESTING AND THE INTERNATIONALIZATION OF THE AMERICAN UNIVERSITY

By

Nathan Michael Johnson

International students have comprised a steadily growing demographic in America's colleges and universities for the past seventy-five years. Over the many years and decades, the proportion of international students whose first language is not English has also increased. Subsequently, American higher education has come to rely on an important assessment for the English language proficiency of prospective international students, the Test of English as a Second Language (TOEFL). The TOEFL has become an influential exam not only within the United States, but also across institutions and educational contexts around the world.

This dissertation examines the history and development of the TOEFL within the historiography of American higher education. Taking into account the history of intelligence testing in the U.S. and advancements in large-scale, standardized testing made during both the first and second world wars, this dissertation also explores the creation of Educational Testing Service (ETS), the company that owns and administers the TOEFL. The history of ETS is intricately connected to the history of American higher education in the 20th century, and its assessment instruments have spurred a robust test preparation industry. As ETS's most widely administered test, the TOEFL has helped proliferate a global industry of English language instruction and test preparation.

Today, the English language holds immense linguistic capital across the world. The English language has played an important role in the evolution of the American university, and the American university, in turn, has helped amplify English's hegemonic status as a global

lingua franca. This dissertation integrates the story of the English language and English language testing into the historiography of American higher education, bringing the power and influence of the English language into fuller relief.

ACKNOWLEDGEMENTS

I am grateful to the College of Education at Michigan State University for generously supporting my doctoral studies. I also wish to thank Steven Weiland, Brendan Cantwell, Dongbin Kim, and Lynn Paine for their guidance.

I am indebted to my friends and family, who encouraged me along the way. Thank you, Joachim, Dennis, Royal, Chris & Annie, Mom & Dad, Megumi, Jamie, Jonathan, and Angela.

TABLE OF CONTENTS

KEY TO ABBREVIATIONS

ACE	American Council on Education
ACT	American College Testing
ASTP	Army Specialized Training Program
CAL	Center for Applied Linguistics
CBT	Computer-Based Test
CEEB	College Entrance Examination Board
CIA	Central Intelligence Agency
CIES	Council for International Exchange of Scholars
COE	Committee of Examiners
CPE	Certificate of Proficiency in English
ELI	English Language Institute
ELL	English Language Learner
ETS	Educational Testing Service
FSI	Foreign Service Institute
GRE	Graduate Record Examinations
IEP	Intensive English Program
IIE	Institute of International Education
IQ	Intelligence Quotient
KAIST	Korea Advanced Institute of Technology
MIT	Massachusetts Institute of Technology
MLA	Modern Languages Association

NAFSA	National Association of Foreign Student Advisors (1947–1964); National Association of Foreign Student Affairs (1964–present)
NEA	National Education Association
OPI	Oral Proficiency Interview
OSN	Online Scoring Network
SAT	Scholastic Aptitude Test (1926–1990); Scholastic Assessment Test (1990–1997); SAT (1997–present)
STEM	Science, Technology, Engineering, Mathematics
TESOL	Teachers of English to Speakers of Other Languages
TOEFL	Test of English as a Foreign Language
TOEIC	Test of English for International Communications
TSE	Test of Spoken English
TWE	Test of Written English
YMCA	Young Men's Christian Association
YWCA	Young Women's Christian Association

Introduction

In the 21st century, the English language rests securely as a singular *lingua franca* in a globalized world. For many people, particularly those in Anglophone countries, the English language does not immediately resonate with globalization. However, English has indeed become an important conduit for international communication. Despite an abundance of artifacts evincing a globalized world—or, perhaps because of them—globalization is often taken for granted as a familiar part of life. Similarly, the hegemonic status of the English language in international discourse frequently evades attention.

Studies in American higher education generally do not focus on the English language as a pressing or key subject. Despite continued academic interest in the internationalization of U.S. colleges and universities, including the steady growth in international student enrollments, scholars of American higher education have not regarded the English language as central to the story, identity, or impact of the American university. Nevertheless, the American university has played an instrumental role in elevating and maintaining the hegemonic position of the English language globally. As a principal source of global knowledge production, U.S. colleges and universities repeatedly reify the privileged position of the English language through academic publications and communications, particularly in the natural sciences (Gordin, 2015). And for international students whose first language in not English, studying at a world class university in the U.S. necessitates acquiring a high level of English proficiency. Thus, in order to participate at the highest levels of academic discourse in the world, whether as a scholar or a student, mastery of the English language is often a minimum requirement.

This dissertation is motivated by perceived gaps in the higher education literature, which largely ignores the relationships between American higher education and the English language. In order to address these relationships, I examine the history and development of the most successful and widely administered English proficiency exam in the world, the Test of English as a Foreign Language (TOEFL). Furthermore, I contextualize the story of the TOEFL within the foundational history of the American university, locating the story of the English language within the historiography of U.S. higher education. The story of the English language in American higher education is integral to the evolution of the university, its curriculum, and the outsized success of the American university system during the 20th century. Since the 1960s, the elevated reputation of American higher education has sparked growing admiration and interest from international students and scholars. As international student enrollments proceeded to rise in the decades that followed, the English language remained an indelible part of the history of U.S. higher education and an increasingly essential medium for navigating the international communications of commerce, scientific discovery, and foreign affairs.

The internationalization of the American university would not be possible without the robust and enthusiastic participation of students from abroad. In addressing the needs of admissions departments across the U.S., the invention and evolution of the TOEFL would help codify and quantify the linguistic expectations and requirements for international students. However, the power and importance of the TOEFL as a gatekeeping tool for university admissions would ultimately lend further credence to the status of the English language as a global *lingua franca*. In this dissertation, I argue that American higher education is complicit in the hegemonic power of the English language in today's globalized world. Furthermore, I demonstrate that high-stakes English language assessments, like the TOEFL, contribute to

establishing and legitimizing the exchange value of English in the global marketplace of ideas. I examine the TOEFL as an artifact of the history of American higher education, a history that begins with language learning at the center of the curriculum. The fortunes of the American university have followed closely with the fortunes of the United States, and I show that the fortunes of the English language as a global *lingua franca*, too, are essential to this matrix.

The dissertation begins with a review of the historiography of American higher education. Chapter 1 examines the historical evolution of the university curriculum and utilizes Laurence Veysey's competing conceptions of American higher education as a framework for understanding how the English language influenced the development of American higher learning. Drawing from the historiography of U.S. higher education, I identify the trajectory of second language learning within the American curriculum and explore shifts in second language learning goals during periods of curricular change or upheaval. Finally, Chapter 1 describes the historical development of international exchange in American higher education, linking the fortunes of the U.S. and U.S. higher education with growing interest in foreign studies. Chapter 1 provides the historical context for understanding the foundational importance of the English language and second language acquisition in American higher education. Likewise, the chapter sets the stage for an eventual imperative among U.S. colleges and universities to test the English language ability of prospective students from abroad.

In order to better understand the impact and influence of the TOEFL on higher education, international students, and the contemporary dynamics of English language hegemony, chapters 2, 3, and 4 offer a three-part historical overview of the TOEFL and its parent company, Educational Testing Service (ETS). Chapter 2 traces the origins of the TOEFL to the technological innovations of psychometric testing at the end of the 19th century and the

geopolitical forces that would ultimately pull the United States into global conflict. With a sense of extreme urgency, the U.S. Army brought together America's leading psychologists and psychometricians to develop instruments for assessing the mental fitness of new recruits at the cusp of World War I. The circumstances and immediate demands of wartime helped accelerate advancements in large-scale psychometric testing, eventually producing techniques and ideas that would find new applications in American higher education. After the war, psychometric testing soon fell out of favor with the U.S. military, but America's colleges and universities were just beginning to consider novel ways for assessing prospective students for admission. Chapter 2 illuminates the global pressures that hastened an explosion of interest and innovation in large-scale assessment technologies and follows the impending intersection of American higher education and a nascent testing industry, which would ultimately birth the TOEFL.

Shortly following the first world war, American colleges and universities began looking beyond their regional purview in order to expand their reach to a truly national scale. America's public secondary education system was gradually taking shape after the turn of the century, and many of the nation's leading universities, particularly those in the northeast, were considering ways to enroll more geographically diverse cohorts of students. Chapter 3 begins by describing the leading figures and unique circumstances after World War I that would ultimately ignite an entire industry and culture of college admissions testing in the U.S. Many of the psychometric testing experts that made innovative strides in the service of the U.S. military during the war would return to their academic positions in peacetime. Goaded by university imperatives to assess greater numbers of students for admission, these scholars, researchers, and university leaders pushed admissions testing in the direction of large-scale, multiple-choice instruments.

A second world war, even more destructive and devastating than the first, once again created a unique opportunity to push the limits of psychometric testing. As with the first world war, the U.S. military summoned the nation's leading experts to design assessment tools for the war effort. A culmination of urgency and military resources drove new advancements in psychometric testing, producing technologies that would find immediate application in college admissions after the war. Chapter 3 sheds light on this second period of remarkable innovation during unprecedented global conflict and explains how the convergence of these wartime experts precipitated the establishment of an independent testing company, ETS. Founded in 1947, ETS is the oldest test development organization in the United States, responsible for such large-scale tests as the Graduate Record Examination (GRE), the Praxis assessments (used for teacher licensure and certification), and the College Board's SAT. Today, ETS's most widely administered assessment is the TOEFL. Chapter 3 brings together the histories of psychometric testing—in war and peace—and the founding of ETS, which would apply wartime testing techniques to the immerging assessment needs of American higher education. By the mid-1950s, ETS was in a strong position to address the challenges of assessing international students' English language abilities, paving the way for the first iteration of the TOEFL.

More than 35 million people around the world have taken the TOEFL and the test is accepted at over 11,500 institutions across more than 160 countries (ETS, n.d.-a; ETS, n.d.-b). The importance and impact of the TOEFL in higher education cannot be overstated, especially when coupled with the continued growth in international student populations at U.S. colleges and universities. Chapter 4 explains how the remarkable growth in American higher education after World War II impacted the fortunes and international reputations of both the United States and its colleges and universities. Further, the chapter draws focus on the expanding interest in

international education and the rising numbers of foreign students, providing a fertile landscape for English language testing. Chapter 4 provides a historical overview of the how the TOEFL came into being and describes the decades-long process of revising the test to assess complex constructs like communicative competence. The sophistication and evaluative power of the modern TOEFL has made it a ubiquitous tool for quantifying English ability and making admissions decisions about prospective international students at U.S. universities. Chapter 4 reveals the TOEFL's potential for reifying and maintaining English as a prestige language and global *lingua franca*. The TOEFL has commercial and cultural implications that help to reinforce English language hegemony in the world today.

The final chapter delves into the implications of the English language in a globalized world, bringing the outsized influence of American higher education and the TOEFL into relief. Chapter 5 considers the complex historical and economic power structures that have ushered the English language into its position as a global *lingua franca*. Within these contexts of history and economy, Chapter 5 weighs the influence of the TOEFL as both a premier English language test and as a reifying force in English language dominance. The importance of the TOEFL as a gatekeeping tool for U.S. colleges and universities also elevates its status as an English language assessment instrument and as a marker of prestige for English language acquisition. Today's linguistic landscape, with English dominating international lines of communication, can be attributed in part to the enormous international success of the American higher education system, as well as large-scale English language tests like the TOEFL.

This dissertation considers the internationalization of the American university through the legacy of language learning and English language testing. Accelerated by the unique demands of wartime, the technological advancements in psychometric testing engendered a lasting potency

and credibility to standardized testing. The TOEFL has become an influential measure of English language ability, and that ability in turn has currency around the world. English language teaching and English language testing have become massive enterprises that cater to a diverse global clientele often seeking admission to American colleges and universities. International students have come to hold a prominent place on many U.S. campuses in the 21st century. Their efforts to learn English and tackle the TOEFL, to live and study in another country, and to contribute to their newfound communities in the U.S. are an ongoing and important part of the American university story.

Chapter 1:

English and the History of the American University

In the historiography of American higher education, rarely does the English language attract attention or even garner a mention. Instead, the English language is usually regarded as a given, assuming it is regarded at all. Perhaps this is to be expected in a country with no official national language, but whose domestic-born population trends toward the monolingual. For most Americans, English is more of a well-worn habit than a state-sanctioned doctrine. The primacy of English in the U.S. is more often de facto than de jure, and consequently Americans routinely take their common tongue for granted. To write about the English language in the history of American higher education, then, would be to give consideration to something that seems inherently clear, if not altogether obvious.

However, discretely peppered within the history of American higher education is a story of English as a signifier of curricular and institutional change. Those changes, in tandem with America's evolving dominance as an economic and military superpower, would subsequently play a pivotal role in elevating the esteem of American higher education in the eyes of the world. The reputed quality and status of America's leading colleges and universities would eventually attract growing numbers of students from abroad—students who in turn would need to successfully navigate their American schooling in the medium of English.

This chapter surveys the historiography of American higher education to reveal the story of the English language embedded in the American college curriculum. The historical context provided in this chapter sets the stage for understanding the role of the English language in

shaping American higher education and, subsequently, the powerful position of American colleges and universities in positioning the English language as global *lingua franca* today.

The Foundations of American Higher Education

As an entry point to the colonial era and the origins of the American college, the historian John Thelin begins *A History of American Higher Education* (2011) with a contradiction tied to the turn of the 20th century. During the last decade of the 19th century, colleges with a colonial legacy rediscovered and, in many ways, refocused their origin stories. The pattern of linking a university or college to its colonial past was an exercise in romanticism and nostalgia, meant to conjure idyllic images of college life and elevate the university's purpose and mission. Naturally, the stories and artifacts that spoke to an institution's history were often repurposed in the service of institutional genealogy, leaning not so subtly on the trappings of prestige associated with pedigree. Thelin points out that this colonial revivalism was intriguing in its novelty at the time—the colonial past had not been lauded quite like this before—but it also "fostered standards of academic honor and imitation" (Thelin, 2011, p. 5). As America was slipping into the 20th century, universities with colonial roots were cementing their status into their romanticized legacies, while newer institutions took note and emulated the characteristics and ornamentation of their time-honored peers.

In the decades that followed, American colleges and universities continued to reify the glorious past into everything from the architecture to the honor code. The process of colonial revivalism, sustained through much of the 20th century, has all but pushed the realities of the colonial era and the burgeoning colonial colleges out of the public imagination. Instead, we are left with a romantic and often incomplete picture of the foundations of American higher

education. Herein lies the contradiction upon which John Thelin begins his opening chapter (2011).

The most pronounced misconception about the earliest American colleges is the idea that the emulation and transplantation of England's universities at Oxford and Cambridge onto the American landscape was more or less effective and complete. The Oxbridge model was, in fact, the starting point for many of America's colonial colleges, but the process of replicating the organization, structure, and curriculum from Oxford and Cambridge was far more complicated, erratic, and strained than latter day revivalists would care to concede. More often than not, the legal, financial, political, and societal realities of the American colonies pressured the early colleges to adjust and refine their execution of the Oxbridge model.

According to Thelin, three key differences between the colonial colleges and the Oxbridge model stood out. First, the organization of the colonial colleges and the Oxbridge model had important differences. Oxford and Cambridge were universities that housed numerous colleges, and while the colleges were responsible for instruction, it was the universities that had the power to confer degrees. The colonial schools, on the other hand, were discrete colleges that coalesced the instructional and degree-granting responsibilities into a single institution (Thelin, 2011, p. 8). More radically, the colonial schools shifted governance out of the faculty's hands. Unlike Oxford and Cambridge, the colonial colleges located school governance within an external governing board. Diverging from the Oxbridge model, the early colleges looked to the universities at Edinburgh and Glasgow for inspiration on how to structure and govern. Coupled with a strong college president, the resulting system of institutional governance allowed for greater accountability and established a resilient model for American colleges and universities to come (Thelin, 2011, p. 11).

A second difference that Thelin acknowledges is the curious likelihood that higher education wasn't particularly necessary in the American colonies. The universities at Oxford and Cambridge were already long-established and had a clear purpose in England. In the colonies, however, the need for higher education, especially among the children of the affluent merchant class, seemed far less transparent (Thelin, 2011., p. 23). Professional careers required apprenticeships, not college degrees, and there was very little emphasis on actually completing college once started (Thelin, 2011, p. 19). Furthermore, and contrary to stereotype, the colonial colleges were not especially concerned with preparing young men for the clergy (Thelin, 2011., p. 26). Interestingly, this absence of a defining purpose for higher education in the American colonies most likely created a vacuum that needed to be filled. Simply aping the higher education priorities of the English elite would only serve to "identify and ratify a colonial elite… essential to transmitting a relatively fixed social order" (Thelin, 2011, p. 24). While the maintenance of class structure remained durable, the colonial colleges also began to embrace an ethos of preparing young men of means for the responsibilities of leadership and public service (Thelin, 2011, p. 26). This emerging purpose of the colonial college would ultimately have consequences for shaping a generation of men at the cusp of the American Revolution (Thelin, 2011, p. 35).

The last difference Thelin addresses is the gaping chasm between England and the American colonies with regards to college preparation. Both Oxford and Cambridge benefited from a coherent system of primary and secondary schools that helped channel young men into the ranks of higher education. In the American colonies, however, no such system of public or private education was in place for school-aged boys. In fact, American public K-12 education wouldn't truly begin gaining traction until the early 19th century, some two-hundred years after the first colonial college was founded (Herbst, 1989). Thus, the standards for gaining admittance

to any of the American colleges in the colonial period were necessarily pliable. In many instances the colleges themselves assumed the role of prep-school, admitting teenage boys and providing them with foundational instruction. This had the dual benefit of raising additional revenue and securing capable students for future college admission (Thelin, 2011, p. 18).

Surprisingly, the contrasts between the idealized Oxbridge model and the realized colonial colleges had very limited bearing on the tone or purpose of the colonial revivalism started in the 1890s. While it was the case that a certain measure of Anglophilia played a part in the colonial colleges' origin stories, the true thrust of the recruiting and public relations effort was anchored in attributes framed as distinctly American. The colonial colleges were invariably linked to the leadership and heroes of America's founding and the American Revolution. In the absence of any clear need or direction for colleges in the American colonies, the fact that these early institutions of higher education found a purpose and mission distinct from those in England—one that cultivated civic leaders in a fledgling democracy—provided a credible and inspiring legacy upon which to hang their laurels.

Thelin (2011, p. 40) closes his first chapter with a reference to a Harvard admission report from 1960 to illustrate an eventual shift away from the language and imagery of the colonial revivalism of the late 19th century, a shift that recognizes the creative potential and vitality of a new generation of college students. By the 1960s, the romanticism and vestiges of the colonial past felt stodgy and overly serious. The college curriculum, even by the end of the 19th century, had moved well away from the classical studies and requirements of Greek and Latin that defined it in the colonial days. The extant colonial colleges of the 1960s had more than secured their status among higher education institutions in the U.S. and around the world. Catering to the dreams, talents, and ambitions of a younger generation would come at zero cost

to their cemented prestige. Rather, these schools and their high-ranking peer institutions would thrive by attracting the very best and brightest students America and the world had to offer—not with romantic tales of an antiquated colonial past, but with the promise and lure of a future containing boundless opportunities.

The American University "Emerges"

For those dwelling in the present, there is a natural tendency to survey the past with a wide-angle lens. The big picture provides space to see trends and patterns, change and stasis, and cause and effect. Through careful research and analysis, historians attach meaning and importance to people, events, and stretches of time that can inform all that follows. In the historiography of American higher education, Laurence Veysey provides one of the most important and durable theses about how and when the American university became what it is today.

Veysey's *The Emergence of the American University* (1965), an abridged version of his doctoral dissertation, pinpoints a transformative period in the history of American higher education. Veysey convincingly argues that the American university, as we recognize it today, "emerged" at the end of the 19th century—a span stretching roughly from 1865 to 1910. The first half of Veysey's book identifies and examines four competing approaches to higher learning during this period that would ultimately reshape the curriculum and the university itself. The innovations and transformations of the university curriculum at the cusp of the 20th century would play a critical role in the success of American colleges and universities and their widening appeal to students from abroad.

Much of Veysey's discussion about the emergence of the American university revolves around a unique concentration of important and transformational university presidents who held

the reins at what would become some of the world's most important and transformational universities. Charles Eliot at Harvard (from 1869–1909), Andrew D. White at Cornell (1866–1885), William Rainey Harper at Chicago (1891–1906), James B. Angell at Michigan (1871–1909), Daniel Coit Gilman at Johns Hopkins (1875–1901), and David Starr Jordan at Stanford (1891–1913) are several key examples. Although the competing purposes and philosophical underpinnings of the American university had already been percolating for generations, the leadership at many of these large and prominent institutions began incorporating them into the curriculum, often in radical and novel ways.

Among Veysey's four competing concepts of higher learning, the first represents the old guard of the university and hearkens back to the Oxbridge model of the colonial past. Veysey calls this approach to higher learning *discipline and piety*. The poster child for the *discipline and piety* perspective toward the end of the 19th century was Noah Porter, president of Yale University (1871–1886). Those who believed in cultivating mental discipline in higher education, like Porter, held an "interlocking set of psychological, theological, and moral convictions" that placed primary value on difficult mental work in abstract subjects (Veysey, 1965, pp. 22–24). The rigorous study of Greek or Latin, for example, was not a means to an end, but the end itself. This kind of labor of the mind was understood as an exercise to build not only mental acumen, but also moral character.

Furthermore, as an inherently conservative approach to higher learning, *discipline and piety* looked askance at new and emerging areas of study. Resistance to science was particularly acute, as it threatened to overtake or displace any order or understanding of the universe to which pious Christians long held claim. For the old guard, then, imbedded within positivism and scientific inquiry was a "synonym for atheistic materialism" (Veysey, 1965, p. 41). The holdouts

of *discipline and piety* regarded science as a menace to the status quo and a danger to the character-building qualities of the established curriculum and its pedagogic rituals. But by the turn of the 20th century, the march of science through the halls of academe had already marked a path of no return. The remaining three conceptions of higher learning that Veysey identifies each accounted for the role science was to play in the university.

The second competing concept on Veysey's list is *utility*. The philosophical grounding of *utility* was not merely an affinity for the pragmatic, the practical, or some vocational ambition, but also a moral imperative that drew from democratic principles. Under the university leaders who fell into the *utility* camp, higher learning functioned both as a space where equity and equality were actualized and as a wellspring from which knowledge could flow freely into society (Veysey, 1965, pp. 63–64). In place of the character-building ethos of *discipline and piety*, *utility* rested its moral calling on the promise of higher education to transform civic life. The university "would make each of its graduates into a force of civic virtue" and encourage political leaders to serve the social good (Veysey, 1965, p. 72).

Unlike *discipline and piety*, the concept of *utility* elevated science as a rational and useful endeavor. Scientific research could be used for practical purposes and innovation, but also had the idealized potential to find "rational substitutes… for political procedures subject to personal influence" (Veysey, 1965, p. 72). Science had the possibility of lending efficiency to the social order, while buttressing democratic institutions against chaos and irrational diversions. Thus, "efficiency"—not "pragmatism"—became the slogan for those who championed *utility*, connotating "collective effort" and implying more "immediate organization" (Veysey, 1965, pp. 116–117).

Many of the most important university presidents during this period were supporters of higher education's shift toward *utility*. Cornell, Harvard, Michigan, Stanford, and Columbia, to name several institutions with *utility*-minded leadership, all experienced curricular changes that ameliorated or introduced subjects and content deemed useful. Perhaps the most radical curricular change came from Charles Eliot, president of Harvard, who implemented "extreme curricular anarchy" in the form of a free-for-all elective system (Veysey, 1965, p. 99). For Eliot, the presumed individual talent naturally associated with Harvard's student body entitled students to absolute individual freedom and self-determination. The students at Harvard could freely pick and choose coursework they deemed most useful and worthwhile. David Starr Jordan, at Stanford, described this position in laissez-faire terms, "It is not for the university to decide on the relative values of knowledge. Each man makes his own market, controlled by his own standards" (Veysey, 1965, p. 114). Thus, Eliot, Jordon, and many of their *utility*-leaning peers embraced the usefulness of higher education but did so without abandoning the inherent elitism their prestigious institutions conferred.

"The Wisconsin Idea," in another example, overtly intertwined the utility of higher education with the efficiency of the state, proposing that the University of Wisconsin function as a kind of feeder system for the state's experts, leaders, and civil servants. Although the Wisconsin Idea never truly took off, facing too much suspicion and pushback from state legislators, its premise and principles highlighted an idealized relationship between a useful university curriculum and democratic society (Veysey, 1965, p. 108). The language of *utility* and useful higher education resonated at the end of the 19th century with many university presidents and, subsequently, across entire university systems.

However, *utility* was not the only concept that elevated the importance of scientific inquiry at this time. The third competing viewpoint of higher education, according to Veysey, was a focus on *research*. The distinction between *research* and *utility* hinged, in part, on the notion of "pure" science—that is, the pursuit of science solely for the advancement of knowledge and truth, rather than for some utilitarian purpose. The *research* camp favored an approach to science that cared far more about the intellectual processes of knowledge-building, learning, and scholarship than its eventual application or usefulness. This point of view extended most directly from the universities in Germany during the 19th century and materialized most clearly at Johns Hopkins University.

Despite being the earliest actualization of the American research university, Johns Hopkins was not the first to award the degree of PhD. That distinction would belong to Yale University in 1861. But by 1871, with the ascendency of Noah Porter as Yale's president, any nascent movement to advance Yale's curriculum in a progressive *research*- or *utility*-minded fashion would soon be tempered by Porter's *discipline and piety* principles (Veysey, 1965, p. 50).

Johns Hopkins University opened in 1876 with Daniel Coit Gilman as its first president. With the support of Johns Hopkins' faculty and fellowship holders, Gilman ushered in a space in American higher education that resembled the ethos and ideals of the German universities at the time (Veysey, 1965, p. 161). The effect was not only to prioritize "pure" research and science, but also to shift the ways in which university students were instructed. Old-fashioned lectures and recitations made room for laboratories and seminars, guiding young scholars beyond rote memorization and toward productive scholarship (Veysey, 1965, p. 153). Johns Hopkins would prove to be tremendously influential and important in shaping the academic, intellectual, and

curricular expectations at America's leading universities. Even Harvard had to play catch-up to John Hopkins, finally matching the Baltimore school in advanced instruction by the early 1890s (Veysey, 1965, p. 165).

Johns Hopkins, under Gilman as president, made great strides in higher education pedagogy and introduced a robust model for graduate education and research. The research output of faculty would eventually offer a quantitative indicator of productivity, prestige, quality, or some other such implied or correlated measure of excellence. By 1901 Stanford began publishing lists of faculty output in its annual reports to hit home the point (Veysey, 1965, p. 177). However, the ideology of "pure" scientific inquiry, as reflected in the German model and ethos, had limits on financing and funding (Veysey, 1965, p. 171). Where money was concerned, the *utility* approach to higher education had certain advantages over the concept of *research* and "pure" science. Simply put, *utility* and usefulness were more marketable to outside donors and external funders than the idealized vision of scientific inquiry. While *utility* would ultimately win out over "pure" *research* during the "emergence" of the American university, the tension between applied and "pure" science in academe would persist.

Veysey's fourth, and final, competing concept of higher learning during the period of "emergence" was *liberal culture*. This perspective was grounded in humanistic ideals and championed "distinct connotations" of culture: "aesthetic, moral and emotional, and social" (Veysey, 1965, p. 184). Science, too, had a place among *liberal culture*, where inspiration or beauty could be attained through fact. Although the *liberal culture* humanists represented a relatively small cadre at colleges and universities, they were particularly vocal—even strident—in pressing their views on American higher education (Veysey, 1965, p. 181). The study of literature only began in American universities after 1865, and modern languages did not appear

as distinct fields until the 1870s and 1880s (Veysey, 1965, p. 182). For more than two centuries Greek and Latin, the curricular cornerstones of American higher education, had only been taught from a functional, grammar-translation standpoint, not as literature. The end of the 19th century brought a humanistic slant to the pedagogy and priorities of the university classroom, producing professors—particularly those in the humanities—who recognized the importance of process in their classroom teaching (Veysey, 1965, p. 221).

Adherents of *liberal culture* tended to admonish the *discipline and piety* approach for its rigidity and moralizing, but at the same time *liberal culture* was partly a continuation of mental discipline in a new guise (Veysey, 1965, p. 194). The humanistic perspective ported some of the vestiges of *discipline and piety*, especially where consensus on canonical worth was concerned. *Liberal culture* in the American university overwhelmingly preferred high culture and often bolstered itself in the same elitist trappings of mental discipline. At the same time, it "evinced a cosmopolitanism which set it sharply apart from the insulated… piety of the mid-century college divines" (Veysey, 1965, p. 197).

Ultimately, *liberal culture* lacked the power or influence to genuinely tilt the university against the tide of *utility* and *research*. While *liberal culture* thrived in the classroom, it was poorly represented among administrators (Veysey, 1965, p. 233). More importantly, faculty in the sciences had little to gain by engaging the humanists or any other champions of *liberal culture*. Armed with facts, logic, and the scientific method, those on the side of *utility* and *research* were securely ensconced in a position of strength (Veysey, 1965, p. 256). *Utility* and *research* would neither stoop nor join the fray. Quite simply (and dismissively), those in the service of science had better things to do.

Veysey's competing concepts of American higher learning would continue to tussle well after the defining dates of the American university's "emergence." However, it was already becoming clear that *utility*'s position was virtually unassailable. The American university in the 20th century would continue to have a place for *liberal culture* and *research*, and, at times, cling nostalgically to *discipline and piety*. But in many ways these positions would need to be defended from the standpoint of their usefulness. As the costs of higher education soared from the 1980s onward, *utility* for the sake of vocational viability became even more pronounced and important. After all, how can one expect to pay off escalating student loans without skills that are immediately valuable in the job market? Metaphors of higher education as a commodity or investment would continue to secure *utility*'s importance in academe, not only for academics and administrators, but also for students in their role as customers and consumers.

The Curriculum

The curricular debates that Veysey identified in his "emergence" thesis neither started in 1865 nor truly concluded in 1910. The defining arguments over *discipline and piety*, *utility*, *research*, and *liberal culture* were gestating long before Noah Porter, Charles Eliot, and Daniel Coit Gilman assumed the presidencies at Yale, Harvard, and Johns Hopkins, respectively. Furthermore, these concepts continued to circulate in one form or another within American higher education long after these men left their posts and endure today in continuing discussions about the purpose of American higher learning.

One of the earliest and most important comprehensive histories on American higher education was Frederick Rudolph's *The American College and University: A History* (1962/1990). Rudolph's *History* had both temporal and institutional scope, capturing the breadth of America's colleges and universities from the colonial period through the end of World War II.

In addition to addressing key topics, like the evolution of the university curriculum and administration, Rudolph shone a light on areas that had previously escaped scholarly attention. Novel for the time, Rudolph gave emphasis to student life and drew attention to the importance of the extra-curriculum, including intercollegiate athletics. For more than two decades after its initial release, *The American College and University: A History* was considered essential reading in the study of American higher education, and it left a pronounced vacuum in 1986 when it temporarily went out of print (Thelin, 1990). Within several years Rudolph's *History* returned to publication and remains an invaluable examination of America's evolving colleges and universities, holding its own with more recent evaluations of American higher education (e.g., and Bok, 2013; Geiger, 2015; and Thelin, 2011).

Fifteen years after the initial publication of *The American College and University: A History*, Rudolph returned to the subject of American higher education with a comprehensive history of the university curriculum. *Curriculum: A History of the American Undergraduate Course of Study Since 1636* (1977) examined the origins and evolution of higher education with a focus on the changing values and shifting priorities toward the curriculum, pedagogy, and institutional mission. Taken together, Rudolph's *A History* and *Curriculum* provide an exceptional primer on the American university, from its beginnings with the founding of Harvard in 1636 to the last years of the so-called "golden age" of higher education (approximately, 1945–1970) (Thelin, 2011, p. 260). Woven into this history is a story of the arguments and ideas that culminated in the period of "emergence" (1865–1910) and continues to resonate in American higher education today.

Although not a centerpiece of either of Rudolph's texts, there exists an unattended story of language education running through the history of the American university curriculum. The

role of ancient languages, the English language, and, eventually, modern languages in the evolution of the curriculum is hopelessly entangled with the competing concepts Veysey (1965) identified in the "emergence": *discipline and piety*, *utility*, *research*, and *liberal culture*. Neither Rudolph nor Veysey, nor any other historian of the American university, have singled out the importance of language education in their histories. However, imbedded within the historiography of American higher education is a story of language education that is deeply entwined with the discussions and debates that shaped the university.

As much as any other shifting priority or purpose in American higher education, the role of language education has been an instrumental driver of curricular change. The story of language use and language learning has been an unassuming and somewhat fragmented part of the historiography of American higher education. Language is taught, language is read, language is translated, language is spoken—and, in tandem with the competing concepts of higher learning, the role of language education in the American university repeatedly influenced the priorities driving curricular change.

The curriculum of the colonial colleges, like their Oxbridge models in England, were originally grounded in the study of ancient Latin and Greek. These two dead languages were important as links to the great civilizations of the past and as signifiers of mental discipline. Hebrew, too, was studied at Harvard, and in the same way Latin provided a link to the Reformation and Greek to the Renaissance, the study of Hebrew helped tie the college to the ancient school of Hebrew prophets (Rudolph, 1962/1990, p. 24–25). For most college boys, the study of Latin, Greek, and, to a lesser extent, Hebrew lacked practical application, and the pedagogical approach of the times focused on rote memorization and precise translations through lectures and recitation. The utility of these ancient languages existed almost exclusively within

the confines of the Latin or Greek classroom, where original texts could be translated and interpreted. From the perspective of mental discipline, studying these languages strengthened and attuned one's mind while also building character. Supposedly, enduring the sheer repetition and tedium of Latin and Greek would cultivate in young men an industrious work ethic and moral disposition.

The results of this sort of curriculum, of course, were often far from the idealized outcome attributed to *discipline and piety*. College students in the colonial era were notoriously rowdy and often subject to disciplinary action. As the sons of colonial merchants and traders, the college experience offered very little besides the network of an elite class. Many students in the colonial colleges never attained their diplomas, nor truly needed the credential for finding viable and fruitful career paths (Thelin, 2011).

From the beginnings of the American college, the study of Latin and Greek was not immediately useful. Although the historical record on the early curriculum is opaque at best, it is unlikely that Latin was used as the language of instruction at Harvard—or, at least not for very many years—and Greek most certainly had little use outside of the Greek classroom. Latin had a place at Harvard and other colonial colleges in ceremonies, rituals, and debate, and there is evidence in the Harvard laws that Latin was required for conversing on campus at least until 1686. But by 1692, the requirement for students to converse in Latin had disappeared from the laws completely (Rudolph, 1977, p. 36). The influence of the Scottish universities likely had some bearing on the role of Latin and Greek outside the classroom in the 17th century. Even during Harvard's formative years, these institutions in Scotland had already had a profound influence on the American colonies. The intellectual developments in Calvinist Scotland, which had a nonconformist bent, helped shift priorities toward "the use of English as the language of

instruction, a stress on science and politics, [and] a friendliness to experimental inquiry" in the colonial colleges (Rudolph, 1977, p. 26).

English began to encroach on the rituals at Harvard as early as 1653 when a commencement debate was conducted in English, but it would prove to be a short-lived innovation and not repeated for another century. In 1754, Harvard allowed for debates in English during the commencement ceremonies, and by 1760 the ceremonies themselves would be conducted in English, eventually leading to the inclusion of English orations during the ceremonies. The ascendancy of English in the public rituals at colonial colleges reflected a gradual intellectual shift that favored scientific and enlightened thinking. The curriculum itself would begin to favor inductive reasoning over deduction, ethics over theology, and English over Latin (Rudolph, 1977, p. 37).

After the American Revolution, the new country was invigorated with independence and moving forward with an optimism in the democratic principles inspired by the Enlightenment. For the visionaries of the time, the future of the country was indelibly linked to notions of progress and innovation. Science, and the utility of science, opened untold possibilities. Communication across the Atlantic would be well-served by studies in modern languages, and French became an early favorite after the Revolution. At the college level, French appeared in the curriculum as early as 1779 (Rudolph, 1962/1990, p. 37). Other modern languages would eventually find their way into the American university curriculum, but by the end of the 18th century, English had firmly secured its place as the most important language for study, scholarship, ritual, debate, and for guiding the nascent United States into a prosperous future.

In 1799, the American Philosophical Society called for proposals for plans for an American education system. The call was competitive, and the Society offered a prize for the

best plan. Many great minds of the day proffered papers in response, from Benjamin Rush to Noah Webster. Rush was particularly forceful in his warnings about the study of ancient languages, describing it as a threat to the country's future development, "To spend four or five years in learning two dead languages, is to turn our backs upon a gold mine, in order to amuse ourselves catching butterflies" (Rudolph, 1962/1990, p. 43). Rush concisely illustrates the argument in favor of *utility* by stressing the opportunity costs of a *discipline and piety* approach to the curriculum. A country had been born, a new century was on the horizon, and the place of Latin and Greek in the American university curriculum was utterly suspect.

For American higher education, the 19th century began with numerous hints at the possibilities of a *utility*-minded curriculum. America's first technical college, West Point, opened in 1802 and was intently focused on science, engineering, and all things practical in the service of the military (Rudolph, 1977, p. 62). Thomas Jefferson's University of Virginia opened in 1819 and, while not quite able to immediately live up to its innovations, was amazingly ahead of its time in both its curriculum and administration (Thelin, 2011, p. 50). Rensselaer Polytechnic, focusing on applied science and teacher education was founded in 1824 and anticipated the land-grant colleges by forty years (Rudolph, 1977, p. 62).

In addition to the scientific and technical emphases at these schools, modern languages began taking a firmer root in the curriculum. Harvard founded a professorship of modern languages in 1816, with language courses beginning in 1819 (Rudolph, 1977, p. 64). While modern languages were immensely useful, particularly in a young country still brimming with immigrants from Europe, Rudolph (1977, p. 63) argues that the favored status of French, Italian, Spanish, and German in the curriculum was more due to a general dislike of Latin and Greek than a particular affection for the modern languages. Harvard's modern language classes were

not required courses, but students often selected them in place of an ancient language requirement. Within two decades, Princeton, Yale, Bowdoin, Dartmouth, and Wesleyan would add modern languages to their curriculum as well (Rudolph, 1977, p. 64).

The curricular shifts away from Latin and Greek, and away from *discipline and piety*, faced a forceful rebuke in 1828 in the form of the Yale Report. In a broad sense, the Yale Report was a response to larger curricular trends in American higher education at the time, but the local catalyst was a revolt among the Yale students in 1825. The students' actions prompted the Connecticut legislature to openly criticize "the impracticality and unprogressive nature of the curriculum" at Yale (Rudolph, 1977, p. 66). Written by Yale faculty, the Yale Report attempted to articulate and prescribe the underlying pedagogical philosophy and curricular priorities in American higher education. The Yale Report relied heavily on the arguments of *discipline and piety* to justify the outsized role of Latin and Greek in the curriculum. At its core, the Yale faculty argued that quality higher learning required hard work that taxed and disciplined the mind. This rigid adherence to time-honored standards in the curriculum was an expression of principle and a genuine belief that mental discipline produced superior character and superior students. The Yale Report was also a last defiant stand against the swelling currents of change.

The Yale Report was widely circulated and read, and for a period its reactionary message held some of the progressive curricular trends at bay. Yale itself doubled down on Latin and Greek, moving to raise admission standards in those subject (Rudolph, 1977, p. 70). But a revolution in science and industry was already underway, and mastery of the new subjects and emerging disciplines was becoming necessary for any young man of vision hoping to participate in the rapidly changing economy.

In the midst of the Civil War, the Morrill Act was passed, establishing a model for technical and agricultural colleges in the United States. The Land-Grant Act of 1862 raised funds for these new practical colleges by selling federal lands and positioned the federal government to affect policy in American higher education (Thelin, 2011). The land-grant colleges prepared young men in areas of agriculture, industry, engineering, and military sciences, offering a practical education for the demands of a modern age.

As Veysey (1965) articulated in his thesis, the competing concepts of American higher education helped usher in the "emergence" of the American college and university. At the end of the 19th century, the curricular debates intersected with a stretch of rapid institutional growth and a new crop of inventive and inspired university leadership. Veysey makes clear that *utility* would eventually gain and maintain primacy over the others, but each concept would retain some measure of importance moving into the 20th century.

Before the elective free-for-all implemented by Charles Eliot at Harvard at the end of the 19th century, students often turned to the extracurriculum to supplement their educations and college experiences. Extracurricular activities consisted of far more than just sports, avocations, social functions, and recreation. From the time of the colonial colleges, students used the extracurriculum to append their educations and, by the mid-1800s, the extracurriculum "with which the students took charge of their own education" was an integral part of the American higher education system. Combined with college chapel, the classical course of study, and a capstone course in moral philosophy, the extracurriculum was the fourth component in a system of higher learning at mid-century (Rudolph, 1977, p. 98).

Intellectual pursuits within the extracurriculum covered a range of interests, including the sciences, the arts, public oration and debate, essay writing, and geography. But it was the literary

society that first secured a firm hold in the American college. The literary society appeared early, albeit briefly, at Harvard in 1728 and again, more durably, at Yale in 1753 (Rudolph, 1977, p. 95). The most spectacular aspect of the literary societies is that they anticipated, and most likely inspired, the eventual inclusion of the humanities within the college curriculum toward the end of the 19th century. Despite the core role of Latin and Greek in the American curriculum since 1636, these ancient languages and their appendant texts had not been studied as literature in the college classroom (Veysey, 1965, p. 182). The literary societies contended with the curricular vacuum by providing a space for students to explore literature, literary criticism, belles-lettres, and humanistic approaches to classical texts (often in English translation) (Rudolph, 1977, p. 96). Housed for over a century in the literary societies of the extracurriculum, the stage was set for the competing concept of *liberal cultural* to contend with the old guard's insistence on *discipline and piety*. With the exception of a handful of early "pioneers," like James Russell Lowell at Harvard, Francis A. March at Lafayette College, and John Bascom at Williams, the study of literature would not join the American curriculum until after 1865 (Veysey, 1965, p. 182; Rudolph, 1977, p. 140). But the enduring popularity and support of college literary societies helped carve inroads into the established curriculum, providing a place for the humanities and a platform for *liberal culture* in the American system of higher education.

The curricular inclusion of English literature also brought the residual benefit of donor-supported libraries, which had been established, curated, and expanded for decades by college literary societies. These literary society libraries often far exceeded the college libraries in size, breadth, and appeal, routinely outpacing their competing campus libraries in circulation (Rudolph, 1977, p. 96). The literary societies would eventually be absorbed as a discipline into

the curriculum and significantly expand their universities' libraries with voluminous collections endowed to the schools.

In his inaugural address at Harvard in 1869, Charles Eliot bemoaned the absence of a "systematic study of the English language." But by 1900, even Yale had supplanted the classics with English language and literature as "the backbone of the humanities" (Rudolph, 1977, p. 140). The rapid expansion of the humanities in the American college curriculum also coincided with the establishment of modern languages as a discipline. The literary societies had never restricted their humanistic subject matter to English only, and many societies had long established an enclave for exploring French literature and other modern languages. Appearing as distinct fields in the curriculum by the 1870s and 1880s, the utility of modern languages was gaining wider acceptance, which the military and civil engineers had always understood (Veysey, 1965, p. 182; Rudolph, 1977, p. 64–65). French, for example, had been required at West Point from the outset because of its relevance to scientific inquiry (Rudolph, 1962/1990, p. 229).

The usefulness of modern languages helped secure their place in the university among *utility*-minded leadership, but the humanistic qualities endeared modern languages to the *liberal culture* crowd as well. An affection for the world and appreciation of global cultures naturally entwined modern languages with the ethos of *liberal culture*. Complemented with rational arguments of usefulness from the adherents of *utility*, modern languages occasioned a united front in the face of *discipline and piety* and the classics curriculum.

At the dawn of the 20th century, *utility* seemed to be securely settled atop the competing concepts of American higher learning. However, the jostling for position would continue for another couple of decades, with instances of reversals and reprisals in curricular progress

(Rudolph, 1977, p. 244). By the mid-1920s, the conflicting ideas would resolve themselves into positions that largely resemble their place in the American university today, with a more outsized emphasis on *utility* and *research*, a subordinate role for *liberal culture*, and an increasingly vestigial existence for *discipline and piety*. After "the emergence," the curriculum in American higher education would veer closer and closer to the values of *utility* in the service of America's rapidly expanding industry and economy, the United States military and national security, the applied sciences and medicine, and, for consumers of higher education, attaining gainful postbaccalaureate employment. The college leadership at the end of the 19th century who effectively positioned their institutions for progressive curricular change understood higher learning as a market. Students, particularly those at elite universities, could be trusted to self-select a curriculum that best served their goals, needs, and, it was hoped, society at large.

The evolution of the American curriculum began with a shift toward English. English was practical and efficient for communicating on campus, and it served as a key marker of American identity and culture. As American higher education gradually aligned with the arguments and values of *utility*, the curriculum proceeded to abandon the ancient language requirements completely. A survey in 1923 revealed that among leading colleges and universities, only Amherst and Williams had no alternative degree options to avoid a Latin requirement for graduation (Rudolph, 1977, p. 244). Of course, today American institutions of higher education no longer restrict students to Latin or Greek in order to graduate, and most institutions offer alternatives to the Bachelor of Arts degree (e.g., Bachelor of Science), allowing undergraduates to circumvent second language study altogether. The utility of English and other modern languages, particularly in the pursuit of science and technology, helped shape the curriculum and the arguments that supported its continued progress and expansion.

"The emergence" of the American university at the end of the 19th century also coincided with major advancements in American agriculture and industry, steadily elevating the importance of the United States economy in the world. By the end of the second world war, the United States was a clear and distinct global superpower. American influence and power were linked directly to its scientific, industrial, agricultural, and military innovations and advancements. And the source for many of these achievements could be traced back to the American university. In addition, the economic and geopolitical primacy of the United States helped magnify the practical and functional significance of the English language, which had already been marked as linguistically important in the wake of the British Empire. Soon the eyes of the world would turn admiringly toward the American college and university. Increasingly, students from around the globe would seek a world class education in the United States. While students from the United Kingdom, Australia, and Canada might not find the threshold to American higher education particularly daunting, international students who did not share English as a first language would quickly discover additional barriers to participation.

International Higher Education Exchange

The system of American higher learning was influenced first by the universities in England and Scotland, and then from those in Germany and France. But until American higher education could hold its own with the European schools, the pattern of student movement for international exchange was almost invariably unidirectional. For more than two-hundred years, the American notion of studying abroad usually meant a journey east, across the Atlantic, to study at one of the established Old-World schools in Europe, rather than an importation of foreign students keen on an American university education (e.g., McCullough, 2011).

America's first foreign exchange student arrived sometime around 1784 from Venezuela to attend Yale University (Bevis, 2019, p. 72). The attendance of foreigners at American institutions of higher education would remain limited and uneven throughout the century that followed. In addition to students from Latin America, American universities in the 19th century attracted occasional students from China, Japan, Europe, and the Middle East (Bevis, 2016; 2019; Bevis & Lucas, 2007). These students were usually sponsored by the governments of their homelands and might find support in the United States in the form of local sponsors, churches, the YMCA and YWCA, the Boards of Mission, and kindly townspeople and faculty. Before the 20th century, American universities offered no formal support services to their foreign students (Bevis, 2019, p. 103).

Throughout most of the 19th century, the European perspective on American higher education was largely critical. Many complaints focused on underdeveloped facilities, including inadequate campus buildings and relatively scant library resources, but also took issue with apparent differences in the curriculum and academic standards, which were judged inferior to the expectations and experiences of European university education (Bevis, 2019, p. 70). This opinion would take a gradual turn for the better toward the end of the century, as Europeans began to recognize improvements in American higher education. The shift in Europeans' attitudes toward the quality of American higher education closely coincided with the period of growth and development that Laurence Veysey identified as the "emergence" of the American university, 1865–1910. The curricular innovations and changes during the end of the 19th century, which elevated the concepts of *utility* and *research* in the American university, were recognized and favorably regarded by European observers and prospective European students. In tandem with

America's economic, military, and geopolitical fortunes at the close of the 19th century, the American university was on the ascent in the eyes of the world.

The period of "emergence" would also overlap with America's earliest immigration policies. Prior to the Chinese Exclusion Act of 1882, the few foreign students who chose to study in the United States could enter the country without federal restrictions. The technical innovation of the steamship helped accelerate immigration in the U.S. from the 1840s onward, and periods of economic stress abroad, such as the potato famine in Ireland, prompted numerous waves of immigrants to settle in the U.S. throughout the 19th century. By the 1880s, in response to escalating immigration, Congress started developing legislation to cope with the ever-increasing flow of immigrants. However, this legislative work happened to coincide with a period of growing resentment among Americans toward Chinese immigrants and their communities. The anti-Chinese sentiment in the U.S. at this time dramatically influenced the focus and tenor of the first federal immigration legislation. The Chinese Exclusion Act, an overtly racist approach to immigration, sought to stem the tide of immigrants by singling out the Chinese. Although possible exceptions could be made for Chinese merchants, diplomats, and students, the message was clear: the Chinese were not welcome in the U.S. (Bevis, 2019; Bevis & Lucas, 2007).

The Chinese Exclusion Act was eventually repealed sixty-one years later in 1943. Restrictions on immigration based on race were removed in 1952, and exclusions to immigrants based on national origin and ethnicity were abolished in 1965. Perhaps the most enduring legacy of the Chinese Exclusion Act of 1882 was the evolving administration and bureaucracy tasked with enforcing U.S. immigration and customs laws. The immigration laws established at the end of the 19th century necessitated the creation of agencies that could evaluate and monitor the flow of immigrants to the U.S. and enforce newly established immigration policy. Despite the new

restrictions and exclusionary policies of the U.S., these nascent immigration agencies routinely permitted foreign students—including students from China—to enter the U.S. for the purpose of pursuing higher education (Bevis, 2019). However, the threshold for entry to the U.S. could only be met with the proper paperwork and documentation. After 1882, immigration to the U.S. was no longer an open door, but entry remained possible when one's papers were in order.

Established in 1867, fifteen years before the Chinese Exclusion Act, the U.S. Department of Education maintained a perennial interest in examining and documenting trends in international education. The Department refined its focus on comparative education at the end of the 19th century, establishing a division to study the education systems of foreign countries. The first decade of the 20th century brought increased federal involvement in international education exchange, which helped incorporate intercultural relations into the priorities of U.S. foreign policy (Bevis, 2019, p. 104). U.S. government interest in foreign students would steadily increase, and eventually attempts were made to take account of the foreign student population and their experiences. However, the first accurate census of foreign students would not be available until 1949, when the Institute of International Education (IIE) initiated its annual survey of foreign students in U.S. higher education. Though imprecise before midcentury, extant estimates indicate that the number of foreign students enrolled at U.S. colleges and universities were very small prior to World War I (Bevis, 2019, p. 115).

Formalized support services for international students began to appear in 1903 in the form of Cosmopolitan Clubs. These clubs were initiated by international students themselves to provide a support network and social outlet on the American campus. As Cosmopolitan Clubs spread around the country, their influence extended well beyond the hosting of social events and cultural exchange activities (Bevis, 2019, p. 110). The organizational potential of the

Cosmopolitan Clubs eventually allowed international students to leverage university support for their unique needs and concerns. By the 1930s, many American universities had created official positions to support their international student populations. The foreign student advisor, for example, would assist foreign students with everything from personal finances to matters of immigration (Bevis, 2019, p. 112). With U.S. immigration laws getting increasingly complex and restrictive during the first decades of the 20th century, foreign student advisors and other burgeoning campus support services would become indispensable for the international student population.

Between the arrival of the first international student in the U.S. in 1784 and the outbreak of World War II in 1939, institutions of American higher education evolved in ways that vastly improved their objective quality and global reputations. Many of the most dramatic changes at U.S. colleges and universities happened during the period of "emergence" at the end of the 19th century, emphasizing a shift in curricular priorities toward *utility* and *research*. International students who, for the better part of a century, were an anomaly on U.S. campuses, began appearing in greater numbers at the start of the 20th century. Gradually, colleges and universities recognized that foreign students required different kinds of support and services, establishing official positions and offices that could cater to the needs of students from abroad.

The first half of the 20th century is an important time in the progression of international higher education in the U.S., a period marked by the development of institutional support for foreign students, a strengthened and stabilized reputation for U.S. higher education, and a steady increase of foreign students coming to the U.S. By the end of World War II, American higher education would begin experiencing a surge in international students. However, the students coming to study after the war years would start to face greater scrutiny in their academic

preparation and ability to participate on America's campuses. For those students whose first language was not English, that scrutiny would shine a spotlight on language ability, leading to the development of English language assessment instruments that could accommodate large pools of applicants. The barriers to U.S. university admission and attendance for international students, much as their American peers were already discovering, would increasingly require overcoming a growing and oftentimes daunting testing regime. In the case of international students who learned English as a foreign language, the battery of potential examinations would inevitably come to include a test of their English proficiency.

Chapter 2:

Global Forces and the Urgency to Test Intelligence

The curricular changes and competing concepts of American higher education were deeply linked to the seismic shifts in political philosophy, industrial innovation, and scientific inquiry in the 19th century. A series of "revolutions"—American, Industrial, and Scientific—provided the *utility* camp with increasingly powerful arguments for moving the university curriculum away from *discipline and piety* and toward a course of education that, at its core, would be immediately useful. Democracy, industry, and science required skilled college graduates who could lead and innovate within the respective domains of government, commerce, and research. The 19th century heralded not only an "emergence" of the American university, but also a rapidly changing landscape for new college graduates to traverse.

At the turn of the 20th century, American colleges and universities began exploring ways to assess prospective students for admission and, at the same time, psychologists in France were developing innovative ways to describe and measure intelligence. The emergence of intelligence testing in France, much like the "emergence" of the American university, occurred at the end of the 19th century and was greatly influenced by revolutions in political thought, industrialization, and scientific research. The initial work in intelligence testing in France would soon be imported and adapted in the U.S. in unpredictable ways, eventually contributing to the design of large-scale standardized tests used for American college admissions. These early intelligence tests provided the foundations for psychometric testing and test development that would ultimately inform entrance exams like the Scholastic Aptitude Test (SAT), the Graduate Record Examinations (GRE), and the Test of English as a Foreign Language (TOEFL).

The origins of the TOEFL begin with the history of intelligence testing in the 20th century. The unique and disruptive geopolitical circumstances of the first world war vastly accelerated technical advancements in large-scale psychometric testing, generating new tools and techniques that would find applications in peacetime. This chapter describes the history of intelligence testing at the onset and aftermath of war to contextualize the extraordinary relationships being formed between American higher education, the U.S. government, and global events. The teams of experts who developed testing instruments for the U.S. military would eventually find themselves back in academe after the war. Their experiences would ultimately inform and shape the American university and the ways in which higher education assesses prospective students, including the English language ability of international students.

Assessing Intelligence in France and the United States

In the wake of their respective revolutionary wars, both the United States and France grappled with the implications and contradictions of their nascent democratic societies and forms of democratic governance, which espoused an ethos of equality but were also demonstrably inequitable. In purportedly equal societies, with liberty and justice for all, the Americans and the French sought ways to rationalize and explain the seemingly inherent differences between individuals and groups within their republics (Carson, 2007). In the American context, the focus on individual and group differences invariably turned to race, whereas the French generally maintained a broader socioeconomic interest in explaining inequality (Carson, 2007, p. 97). But on both sides of the Atlantic, a new construct was evolving to help address and identify the differences in human capability. By the end of the 19th century, the concept of "intelligence" had taken shape.

The notion of intelligence had existed in the French language since at least 1500, and its meanings continued to expand into the early 1800s. Prior to 1835, the range of definitions for "*intelligence*" in French included, for example, one's knowledge, comprehension, or capacity for understanding. Earlier definitions of "*intelligence*" included not only one's ability to know, but also the mere possession of knowledge. The evolution of the dictionary definition of "*intelligence*" gradually constrained the knowledge-possession aspects of the word in favor of the knowledge-ability denotations. In the sixth edition of *Le Dictionnaire de l'Académie françois* in 1835, "*intelligence*" no longer simply denoted knowledge, but instead suggested either a universal ability or some relative capacity that could be acquired in varied degrees (Carson, 2007, p. 78). In addition, the Academy's 1835 volume introduced the word's application to animals, a reflection on recent developments in natural history and the notion that intelligence could exist in measures (Carson, 2007, p. 79). The meaning of "*intelligence*" in French had already begun to indicate ability, and even hierarchy, by the mid-19th century.

The 18th century dictionary definition of intelligence in English, both in Britain and the U.S., was rather different from the French. Samuel Johnson's definition in 1755 emphasized the exchange of knowledge or information between individuals rather than any particular faculty or ability (Carson, 2007, p. 78). In 1841, with the second edition of Noah Webster's *American Dictionary of the English Language*, intelligence expanded its meaning in English to include the idea of "gift" or "ability," similar to the French denotations at the time (Carson, 2007, p. 79). But in both the U.S. and France, the dictionary definitions would be merely a starting point for modeling the meaning of intelligence for measuring human difference.

The development of intelligence as a scientific construct began in the 19th century among anthropologists and natural historians. The discussions and inquiry into differences in

intelligence tended to focus on the physiology and physical differences of humans. The racially diverse context of the U.S., with its indigenous peoples and Black slaves, provided a political and moral incentive to explain away or rationalize inequality in racial terms. French anthropologists, on the other hand, were far less interested than the Americans in establishing a hierarchy of human races. But together, the French and American approaches to modeling and quantifying human differences contributed to the earliest understandings of intelligence as a psychological concept (Carson, 2007).

French anthropologists, like Paul Broca (1824–1880), experimented with the measurement of human skulls and skull structure to make inferences and draw conclusions about human intelligence. Craniometry and craniology in France provided not only a quantifiable avenue for understanding intelligence, but also a positivist epistemology that privileged instrumental precision. Although the explanatory power of craniometry had failed by the end of the 19th century, its approach and ideas for measuring intelligence helped solidify the notion that human intelligence is quantifiable and instrumentally measurable (Carson, 2007).

In its own way, the American anthropological approach to intelligence also contributed to defining the construct of intelligence in the 19th century. The emphasis on race in the U.S. greatly impacted the way intelligence was understood. By focusing on racial difference, American anthropologists—and eventually psychologists—helped shape intelligence into a construct with a singular dimension. Rather than recognizing a variety of mental faculties or potential talents within a given individual or group, intelligence could be understood as a single, global indicator of human ability. More to the point of racial determination, a simpler view of intelligence removed any complications when ordering human hierarchy by race (Carson, 2007, p. 90–91).

As a unidimensional construct, intelligence allowed scientists and even the general public to easily ascribe and explain social position by race.

French psychologists of the late 19th century eventually appropriated the dictionary definitions of intelligence, but with some important changes and points of emphasis. The most significant denotative shift was a prioritization on positivism for understanding intelligence as a domain for scientific inquiry. This focus on the scientific helped distance the definition of intelligence from any previous metaphysical denotations, such as notions of human agency, transcendental morality, or divinity (Carson, 2007, pp. 114–120). Added to this definition in France was the idea of difference, which contributed to the exploration of intelligence in terms of pathology (Carson, 2007, p. 121). In order to understand "normal" intelligence, French psychologists sought to examine intellectual and behavioral anomalies that were considered deviations from the norm (Carson, 2007, p. 126). Thus, the definition of intelligence in the hands of French psychologists allowed the construct to be more readily explored through scientific inquiry, where researchers and clinicians could frame their investigations in terms of pathology and difference.

By the end of the 19th century, French psychology had transformed into an "instrument-oriented experimental science" (Carson, 2007, p. 131). The first scale for measuring intelligence, the Binet-Simon Scale, was developed in 1904, and its first iteration was completed in 1905. French psychologists Alfred Binet and Théodore Simon created the scale with the clinical purpose of helping physicians diagnose individuals with subnormal intelligence. The Binet-Simon Scale was not a simple multiple-choice exam, but rather a battery of questions and problems that could be presented to a subject by trained clinicians who then would closely observe and evaluate the responses. Although the scale was intended to classify those with

subnormal intelligence, the Binet-Simon Scale was referenced to what was considered "normal intelligence" on a continuum of age. Subsequently, the Binet-Simon Scale could also be administered to "normal" individuals who may be merely experiencing a degree of arrested development (Carson, 2007, p. 140–141). The 1905 version of the Binet-Simon Scale relied primarily on a subject's linguistic and reasoning capabilities, such as explaining distinctions between pairs of words. The 1905 version understood intelligence to be a singular entity but explored its manifestation through multiple modes of investigation. This allowed for both the "arbitrariness…and richness" of clinical observation, rather than reducing intelligence categorization to a simple score. (Carson, 2007, p. 144).

The observational and clinical aspects of the original 1905 version of the Binet-Simon Scale necessarily compromised the scope of its administration. Multiple psychologists were required to evaluate a single individual, and the focus was on identifying symptoms of pathology. However, Binet and Simon continued to refine their scale until 1911, when Alfred Binet died. The subsequent iterations steadily reduced the clinical attributes of the Binet-Simon Scale while simultaneously magnifying the importance of precision and objectivity. By 1911, the Binet-Simon Scale had refined its output to a single numeric score (Carson, 2007). Thus, with the 1911 version, the unidimensional interpretation of intelligence also gained prominence. In addition, four key features of the 1911 Binet-Simon Scale emerged: (1) the scale equated normal intelligence with average intelligence; (2) intelligence was seen as both developmental and universal; (3) intelligence could be leveraged for a variety of uses (e.g., placement, categorization); and (4) the Binet-Simon Scale never implied that intelligence was biological, hereditary, or fixed (Carson, 2007, pp. 145–147). Nevertheless, interest in the Binet-Simon Scale in France remained rather limited. In clinical circles, physicians did not appreciate being

supplanted by psychologists in the administration of the scale, and therefore refused to give up their authoritative ground. And in the domain of French education, a well-established system of competitive school exams left little room for the Binet-Simon Scale to add value (Carson, 2007, p. 151).

Despite the limited interest in the Binet-Simon Scale in France, as well as French case studies revealing the scale's weaknesses and limitations, psychologists in the United States would soon gravitate to the instrument. In 1910, the American psychologist Henry H. Goddard, who had worked at the New Jersey Training School for Feebleminded Girls and Boys, published his experiences using the 1908 version of the Binet-Simon Intelligence Scale. Goddard's article suggested the promise of an empirically explorable and quantifiably measurable intelligence construct (Carson, 2007, p. 161). Furthermore, and contrary to French psychologists, many American psychologists understood as a key feature of the Binet-Simon Scale that intelligence was indeed biological in origin, hereditary, and fixed (Carson, 2007, p. 147). Coinciding with America's racial divisions, recent nativist resistance to immigration, and social applications of Charles Darwin's theories of natural selection, the U.S. offered a fertile environment for assessing intelligence—rather than investigating it. The American psychological community and the broader American culture were acutely receptive to a notion of intelligence that could quantifiably separate the wheat from the chaff (Carson, 2007, p. 162). Assessing intelligence not only offered the possibility of weeding out groups and individuals seen as unfit, but also had the potential for serving managerial and bureaucratic efficiencies in an emerging professional class (Carson, 2007, p. 160). Thus, in the hands of the Americans, Binet-Simon offered a metrically oriented view of intelligence that helped solidify a hierarchical and unidimensional interpretation of the intelligence construct in the U.S.

In 1916, Stanford University psychologist Lewis M. Terman recalibrated the Binet-Simon Intelligence Scale on a population of Americans. Terman's version of the test, known as the Stanford-Binet, soon became the standard for assessing intelligence in the U.S. (Carson, 2007, p. 179). The Stanford-Binet primarily tested vocabulary, memory, observation, reasoning, and reactions to complex or moral situations (Carson, 2007, p. 188). Like the original Binet-Simon, Stanford-Binet relied heavily on the linguistic capacities of its subjects. A subject's command and breadth of the English language inevitably correlated with Terman's measures of intelligence. Unlike Binet-Simon, however, the Stanford-Binet was not restricted to isolating deficiencies and identifying pathologies. Instead, the Stanford-Binet assessed a broad range of intelligence, classifying everything from the subnormal to the superlative. Perhaps Terman's most indelible contribution to American psychology, as well as American culture, was his coining of the term IQ. Through the Stanford-Binet test, Lewis Terman equated the construct of intelligence with his measurement of an intelligence quotient (IQ). With a single number, the Stanford-Binet could classify anyone with an innate, fixed measure of intelligence, precisely locating them on a hierarchical scale of ability.

The American iteration of the Binet-Simon Intelligence Scale shifted the psychological priorities from the exploration of intelligence and diagnosis of pathologies to the assessment of intelligence and scaling of abilities (Carson, 2007, p. 162). The Stanford-Binet tested intelligence, as measured by IQ, but its expansive hierarchy of intelligence opened new avenues to talk about merit. The applications of intelligence testing in the U.S. would eventually extend to identifying those considered most capable and deserving of admission to American colleges and universities. In other words, merit would play a role in college admissions, and the results of intelligence tests offered an intriguing correlation with merit.

Terman's intelligence test—like the Binet-Simon before it—was a complex instrument, both time-consuming and labor-intensive to administer. Thus, the Stanford-Binet was too expensive and cumbersome to administer to large groups of people at one time. But the path to large-scale merit-based testing was already approaching. Geopolitical conflict and violence in Europe would soon bring the world to war, and the circumstances of America's involvement would present a unique opportunity to develop assessment instruments that could be administered to thousands of people at a time.

The World at War: Large-scale Testing

The year after Lewis Terman introduced the Stanford-Binet test, the United States declared war on Germany. The U.S. Army was far from battle-ready on April 6, 1917, when the U.S. entered World War I. In terms of equipment and personnel, the U.S. military was in sorry shape and short supply. The urgency presented by global conflict hastened massive recruitment and conscription efforts, which would require assessments for military fitness and extensive combat training for new recruits. Prior to World War I, the United States military had very little interest in evaluating the intelligence of its personnel, preferencing assessments of character instead (Carson, 2007, p. 199). But within weeks of the American entrance into the war, a Harvard psychologist named Robert Yerkes began promoting a plan for intelligence testing across the entire military apparatus. Eventually receiving approval from the U.S. Army, Yerkes embarked on a test development and administration program never before seen (Carson, 2007, p. 202).

Yerkes recruited and worked with a number of prominent American psychologists at the time, including Lewis Terman, Henry Goddard, and Edward Thorndike (Carson, 2007). Yerkes and his colleagues were the leading experts in intelligence testing in the U.S., and their

development of what eventually became known as Army Alpha and Army Beta led to a number of testing innovations that continue to inform psychometric testing today. Army Alpha was the principal testing instrument developed for the U.S. Army, whereas Army Beta was an alternate version of the test created to evaluate illiterate recruits (Carson, 2007, p. 199). The pilot program used to develop Army Alpha was known as Army *a* (Carson, 2007, p. 201).

The relationship between Yerkes and the U.S. military created a novel dynamic between psychometrician and client. The psychologist experts developing the test found themselves navigating a different set of needs and mores embedded in the U.S. Army establishment. To satisfy the Army-as-customer, the test developers needed to create an assessment that economized time and could be administered to every recruit (Carson, 2007, p. 203). The process exposed differences of opinion among the phycologist experts and the U.S. Army alike. Army *a* began as an assessment for mental deficiencies, to identify and expunge unfit recruits with "intellectual deficiency, psychopathic tendencies, nervous instability, and inadequate self-control," but gradually shifted to include a focus on superior results as well (Yerkes, April 1917, as cited in Carson, 2007, p. 201).

Novel issues in test development, not yet considered in civilian testing also emerged. The stakes of the Army Alpha were considerably higher than most psychological assessments up to this point. Generally, lower stakes assessments would not tempt test-takers to cheat or intentionally underperform. But Army Alpha changed that dynamic by raising the stakes of the outcomes. For a new Army recruit, success or failure on Army Alpha could mean the difference between life and death, civic duty and self-preservation, or courage and cowardice. Innovations like test security would have to be created and implemented. The use of multiple forms of the test, for example, helped mitigate cheating, and the use of multiple-choice responses helped

reduce the time it took to administer the test. Army *a* integrated the established tools used in civilian intelligence testing, but "transformed them as an overall package by placing a premium on economy of time, security, measurement across a broad spectrum, and adaptability to group administration" (Carson, 2007, pp. 204–205). Thanks to Yerkes and his team of intelligence-testing experts, Army *a* had eliminated the need for a professionally trained psychologist to administer the tests, assess the responses, and analyze the outcomes. Instead, with Army *a*, the results could be reduced to a single number with immediate interpretive power and the impression of objectivity (Carson, 2007).

The development of Army *a* also raised concerns about construct validity. The test designers were deeply concerned about whether or not they were, in fact, testing the constructs they claimed to be examining. To address these concerns, Yerkes needed to (1) demonstrate that the items selected for the test were appropriate in content and difficulty for the test-taker to demonstrate their understanding, (2) show that the test correlated with an already established instrument that measured the same construct, and (3) show that the test correlated with some other non-test form of evaluation (Carson, 2007, p. 205). Army *a*'s design and assembly by a team of intelligence-testing experts gave credence to the care and appropriateness of item selection. A more important victory for Army *a*'s validity was the correlational studies with established intelligence test, like the Stanford-Binet. The high correlation with intelligence tests helped validate Army *a* as a tool for assessing psychological intelligence. Finally, Army *a* was correlated with the "industrial and military history" of the examinees (Yerkes, July 1917, as cited in Carson, 2007, p. 205). This last correlation helped Yerkes address his perception of "serviceability" required by the Army. For Yerkes and his team, Army *a* was a carefully designed instrument, correlated with established civilian tests for intelligence and military

evaluations of individual test-takers. The “serviceability” angle helped Yerkes promote Army Alpha as not only an indicator of intelligence, but also as a predictor of serviceable performance in the U.S. Army (Carson, 2007, p. 207).

At its height, as many as 10,000 Army Alpha tests were processed per day during World War I (Carson, 2007, p. 197). The ease and expediency in which the tests could be administered helped the Army keep pace with its wartime demands. In the month prior to the United States’ declaration of war against Germany, the U.S. Army had approximately 6,000 officers and 200,000 soldier, but by the war’s conclusion in November 1918, the number of officers and soldiers had increased to 200,000 and 3.5 million, respectively (Carson, 2007, p. 200). Despite the test-developers’ intent for Army Alpha to be used to evaluate recruits on a spectrum of inferiority to superiority, in practice the test was only used to determine fitness to serve. Although Army Alpha was widely used throughout the war, its reception by the Army was decidedly mixed. Many officers looked askance at its descriptive or predictive power, and others were immediately critical or dismissive of the test. Nevertheless, by war’s end, more than 1.75 million recruits had taken the Army Alpha (Carson, 2007, p. 218).

The most important and enduring aspects of the Army Alpha program are its innovations for large-scale testing and its broad cultural implications for understanding intelligence and intelligence assessments. The military exigencies of World War I introduced speed, economy, and security to test design. Complex psychological constructs, like intelligence, could be reduced to timed, multiple-choice exams, administered in groups and evaluated by anyone with an answer key. In addition, Yerkes’ assessment program elevated the status of intelligence evaluations in the culture of the U.S. Army and, eventually, the broader American public. Army Alpha introduced the concept of intelligence to a generation of officers who would subsequently

consider the role of intelligence in their own evaluations of soldiers. Although Army Alpha was meant to evaluate military fitness, officers often used the information to confirm their own assessments about soldiers. And unit commanders were known to sometimes "pass the cull," retaining high scoring soldiers while transferring low scorers to other units (Carson, 2007, p. 218–219). Even after the war, the U.S. Army continued to apply forms of intelligence testing to its systems of recruitment and promotion. Army Alpha contributed to an understanding among psychologists and segments of the American public that intelligence was unitary, quantifiable, and measurable on a unidimensional scale (Carson, 2007, p. 197).

The civilian applications of these lessons and techniques would eventually extend beyond the field of psychology, seeping into high stakes domains where concepts of intellect are entangled with expectations of privilege and merit. American colleges and universities would soon apply the new technologies of testing to admissions decisions, generating an industry and culture of testing in U.S. higher education. The power of the new testing regimes to quantify students' academic ability and potential would lend further legitimacy and authority to the evaluative decisions of America's universities. By the time U.S. colleges and universities slipped into the arena of evaluating English language ability in the 1960s, American higher education would hold a surfeit of reputational capital. Commanding an abundance of influence and wielding a high-stakes English language test, American universities would carve another avenue for reinforcing English language hegemony around in the world.

Chapter 3:

Expanding Admissions and the Advent of Educational Testing Service

By the middle of the 20th century, American colleges and universities had turned to large-scale testing to inform their admissions decisions. Innovations in psychometric testing during the first World War, which included test security features and multiple-choice responses, were further enhanced through advancements in computing technologies. In addition, the "emergence" of the American university at the end of the 19th century helped precipitate an expansion of America's nascent secondary education system during the first decades of the 20th century.

Coordinating the curriculum of America's public high schools with the expectations of America's colleges and universities presented numerous problems for college and high school administrators alike. Mismatches between high school education and college-readiness revealed a need for objective indicators of students' preparation for college, opening new avenues for admissions testing. Furthermore, a growing demand for college education in the aftermath of World War II, and the subsequent increase in the number of applications, necessitated a shift in the design priorities of admissions tests. In order to evaluate prospective students' preparedness and accommodate an increasing volume of applicants, college admissions tests would have to be efficient and cost-effective, allowing for large-scale administration, expeditious scoring, and straightforward interpretation. Naturally, the validity and legitimacy of the tests would require buy-in from a whole swath of stakeholders, ranging from university administrators to the general public.

The highly technical demands of test development required teams of experts and, eventually, organizing bodies to provide structure and boundaries for the projects. With guidance

and influence from the leadership at America's most prominent universities, particularly Harvard University, the history of college admissions testing in the U.S. took shape around the formulation of new, quasi-independent institutions that specialized in the creation, promotion, and dissemination of large-scale admissions tests.

Historically, perhaps the most important and influential organization in admissions testing in the U.S. is Educational Testing Service (ETS). From its founding in 1947, ETS has developed and produced some of the most impactful and recognizable tests in American history. At its inception, ETS took over the development and administration of the College Board's twenty-year old admissions exam, the SAT, and the Carnegie Foundation's Graduate Record Exam (GRE). Later, ETS would develop its own landmark testing products, including the Praxis test series for teacher certification programs, the Test of English for International Communications (TOEIC) for assessing English proficiency of non-native speakers in professional settings, and the Test of English as a Foreign Language (TOEFL) for evaluating the English proficiency of non-native speakers for college and graduate studies in the U.S. Coinciding with rapidly increasing numbers of college applicants from the U.S. and abroad, the establishment of ETS played an important role in the expansion of American colleges and universities after the 1940s.

This chapter describes the foundation of ETS in the wake of two world wars, examining the remarkable network of experts who crossed back-and-forth between lines of military service and academe. The leading figures in psychometric testing, both researchers and university administrators, played a crucial role in shifting the weight of university admissions toward large-scale, multiple-choice assessments. Their work to create ETS as an independent testing agency helped ignite standardized testing as a form of business or industry. The founding of ETS is a

significant moment in the history of American higher education. As the first independent testing organization in the U.S., ETS was well positioned to tackle the problem of English language assessment for university admissions.

The College Board and the Creation of the SAT

Entrance requirements to America's higher education institutions had existed in one form or another since the founding of Harvard in 1636. Demonstrating proficiency in Latin and Greek was the benchmark admissions standard for the better part of two-hundred years (Schudson, 1972, p. 39). But the 19th century brought a number of curricular changes and additions at many U.S. colleges and universities, which complicated the notion of college readiness at many institutions. By the 20th century, America's colleges often expected a wider breadth of knowledge from their prospective students than just Latin and Greek, including preparation in geography, English grammar, English composition, algebra, and history (Schudson, 1972). The "emergence" of the American university at the end of the 19th century, with the deemphasis of ancient languages (i.e., the *discipline and piety* approach) and the prioritization of useful subjects (i.e., *utility*), expanded the range of subjects that freshman students were required to know upon matriculation. The burden for evaluating prospective students fell squarely on the university faculty, who would individually examine students on campus prior to admission (Schudson, 1972).

For most of the history of American higher education, there was little articulation between universities and secondary schools (Schudson, 1972). In fact, throughout much of the history of American higher education, formalized secondary schooling did not exist or only nominally existed. Unlike the German model, which became so important for the founding of Johns Hopkins University, the American system of higher education did not have an established

system of public education that prepared children and teenagers for college (Herbst, 1989). The National Education Association (NEA) was founded in 1857 but would only begin to promote a national high school curriculum in the 1880s with its National Council on Education's reports on secondary schools and the nation's colleges (Schudson, 1972, p. 42). America's expansive geography and diverse array of higher education institutions made coordination between college curriculum and high school curriculum difficult at best. Somehow America's formative secondary school administrators would have to devise curricula that could not only meet the varied admissions demands of the nation's colleges and universities, but also provide a relevant course of education for those students not seeking higher education (Schudson, 1972).

Attempts to rein in this curricular chaos began primarily at the local level. Regional education associations linking colleges to secondary schools were formed between 1885 and 1898, and soon after these organizations began coordinating curriculum expectations and entrance examinations (Schudson, 1972, p. 41). The NEA eventually tried to address the problem of coordination between colleges and secondary schools in 1892 when it established the Committee of Ten, a team of "important and prestigious educators of the day" led by the president of Harvard University, Charles Eliot (Schudson, 1972, pp. 42–43). Published in 1894, the Committee's report introduced some important innovations, like establishing quantifiable units of instruction-time so that students' preparation could be compared across high schools. More germane to the Committee's purpose, however, was the report's recommendation to create a national organization of experts to lead the coordination efforts between America's universities and secondary schools. Acknowledging the absence of both federal control and significant state control of education, the Committee in essence followed Eliot's advice to establish an organizing body of experts to compensate for the country's longstanding lack of centralization (Schudson,

1972, p. 43). The Committee of Ten's report would help set the groundwork for the creation of the College Board.

The College Board was founded in 1900, and one year later the College Entrance Examination Board (CEEB) was created. Spurred in part by an 1899 report from another NEA-appointed committee, the Committee on College Entrance Requirements, the CEEB differed in regional scope and representation from the organizing committees that preceded it. In contrast to the Committee of Ten and the Committee on College Entrance Requirements, which were assembled by a nationally representative body, the College Board and its CEEB had limited regional representation. The College Board and the CEEB originated with distinct regional priorities, principally representing the attitudes and values of the Eastern colleges. Colleges and universities in the Midwest, which at the time preferred a system of high school accreditation over entrance examinations, were initially wary of the CEEB and its fledgling examination system (Schudson, 1972, pp. 44–45). But even Harvard, whose President Eliot supported the creation of the College Board and CEEB, was not quick to adopt the CEEB's entrance examinations. Along with Harvard, a number of schools in the East retained their own proprietary exams in lieu of the CEEB test. Ten years after the first administration of the College Board's entrance exams, the combined number of school-specific exams in the East continued to exceed the total examinations administered by the College Board (Schudson, 1972, p. 45). Despite a growing desire to attract students from a broader geographic region and more diverse secondary school backgrounds, the Eastern colleges were slow to join the College Board, let alone allow its entrance exam to shape their admissions decisions (Schudson, 1972, p. 46).

The College Board's original battery of entrance exams in 1901 were essay-based and constructed around a consensus of necessary subjects and content for college preparation. The

content was agreed upon by the College Board, which drew upon an 1899 report from NEA's Committee of Twelve, as well as recommendations from relevant professional associations (Schudson, 1972, p. 45). The subject-based essay format continued until around the end of World War I. In 1915, Harvard, along with Yale and Princeton, finally eliminated its own proprietary exams and requested that the College Board develop a more streamlined collection of comprehensive examinations to be used in tandem with other admissions indicators (e.g., high school academic records). Nevertheless, these new comprehensive exams would still follow an essay format (Schudson, 1972, p. 48).

Th close of World War I and the subsequent advancements of large-scale psychological tests dramatically altered the admissions testing landscape in 1918. Despite the U.S. Army's dwindling enthusiasm for psychological tests immediately after the war, the technological innovations of the Army Alpha exam offered new possibilities for assessment instruments in civilian spaces. The U.S. government, businesses, and education all gravitated to the promise and potential of efficient testing instruments that could differentiate and sort people. As early as 1919, Columbia University became the first higher education institution in the U.S. to employ psychological tests as part of its admissions procedure (Schudson, 1972, pp. 48–49). By 1920, as many as 200 U.S. colleges had dabbled with psychological tests for a variety of purposes, including admissions. With correlational evidence suggesting that these tests could better predict college achievement than other types of tests and high school performance records, the College Board began pursuing avenues to study and develop its own psychological test (Schudson, 1972, p. 49).

The College Board's new entrance examination, the Scholastic Aptitude Test (SAT), was first administered in 1926. More than 75% of the 8,000 prospective students who took the first

SAT applied to schools in the Northeast, including the big Ivy League schools and smaller liberal arts colleges (Schudson, 1972). The design of the original SAT followed closely with the psychological tests that preceded it, but the design committee continued to develop, evaluate, and improve the test. Within several years of its inception, the SAT began to take on some of its most enduring characteristic features, including distinct sections for mathematics and verbal aptitude (Schudson, 1972, p. 50).

The massive expansion of public high school education during the first four decades of the 20th century helped facilitate a growing interest and demand for higher education. As the numbers of prospective college students grew, the nation's colleges and universities became increasingly reliant on more efficient and economical techniques for evaluating applications and making admissions decisions. The notion of selection and the prospect of actually having to limit enrollment coincided with a rising interest at Eastern schools to broaden the demographics of their entering cohorts. Technology and enterprise were quickly reducing vast expanses of geography in the U.S., uniting the states with faster modes of transportation, commutations, and commerce. The long-established universities in the Northeast, which guided the founding and trajectory of the College Board and the SAT, used the new admissions test to help select and attract talented students from beyond the usual network of private prep-schools in the East. Through the SAT, schools like Harvard, Yale, and Princeton could cast their net far and wide for a truly national pool of applicants.

In 1934, James Conant, president at Harvard, created the National Scholarships at Harvard to attract a more geographically diverse group of applicants and put Harvard on a trajectory to becoming a truly national university. Conant appointed assistant deans Henry Chauncey and Wilbur Bender to ascertain the best approach for selecting Harvard's national

scholars. Chauncey and Bender concluded that the SAT, along with Harvard's subject exams, personal interviews, and recommendations, would provide the best package for student evaluation (Lemann, 1999, pp. 38–39; Schudson, 1972, p. 53–54). In 1937, coordinating with scholarship committees at Yale, Princeton, and Columbia, Harvard University enlisted the College Board to administer a series of SAT exams at 150 centers throughout the U.S. (Lemann, 1999, p. 39). These one-day SAT exams were an instrumental part of the new national scholarship competitions, but the scope and scale of the 1937 administration would augur a much greater proliferation of the SAT to come.

With the ostensible backing of the Ivy League schools, the SAT gained greater legitimacy and recognition nationally. Regarded as a valid predictor of college aptitude and success, the test gradually took on greater weight in the admissions process. The fact that the SAT could be administered and scored relatively quickly—and without imposing on the university faculty's time—made the test immensely useable and attractive. Steadily, the efficiencies and perceived validity of the SAT proved too persuasive to resist. As the U.S. approached the prospect of a second world war, America's colleges and universities surrendered more of the admissions decision to the SAT results. Within days after Japan's attack on Pearl Harbor, Yale, Princeton, and Harvard reduced their admissions exam requirements to a single test, the SAT (Schudson, 1972, p.54). The College Board soon followed suit, doing away with the essay exams (Schudson, 1972, p. 55). Between the National Scholarships at Harvard, first awarded in 1934, and the United States' declaration of war at the end of 1941, the SAT would become the single most important test for college and university admissions in the U.S.

The Foundation of ETS

The College Board was the first agency in the U.S. with a singular focus on testing, and its origin story coincides with not only the "emergence" of the American university, but also the development of intelligence testing in the U.S. The creation, development, and ultimate success of the College Board's SAT were rooted in a rapidly expanding system of public secondary education in the U.S. and a desire from the Eastern schools to attract more geographically diverse cohorts of students. Harvard University, through its leadership in Charles Eliot, was instrumental in the formulation of the College Board in 1901, and Harvard continued to play an outsized role in the expansion of the SAT in the 1930s and 1940s thanks to its then-president, James Conant.

Conant devised the National Scholarships at Harvard in 1934, which placed greater weight onto the SAT results, and then he coordinated with other Ivy League schools in 1937 to expand the administration and geographic reach of the SAT. Shortly after the successful SAT administration in 1937, Conant pushed further by proposing the creation of a new national testing organization that would manage and operate all major standardized educational examinations (Lemann, 1999, p. 39). Curiously, Conant faced major pushback on the idea from Princeton professor Carl Brigham, the creator of the original 1926 SAT, who warned of potential conflicts of interest. Brigham recognized that a testing agency, like the one Conant proposed, might become so invested in promoting its own tests that it could lose sight of whether the tests were, in fact, effective (Lemann, 1999, p. 40). In addition, Brigham anticipated negative washback from standardized testing, warning of a degradation in how mathematics and English would be taught in light of the way in which those subjects are instrumentalized on multiple-choice tests (Lemann, 1999, p. 40–41). But in January 1943, roughly a year after several Ivy

League schools had dropped all other admissions exams except for the SAT, Carl Brigham died, clearing the way for James Conant's ambitious plan to establish a national testing agency (Lemann, 1999, p. 41).

The second World War, much like the first, had important implications for large-scale testing. Although the U.S. military's interest in intelligence testing had cooled off by the end of World War I, the return of open hostilities in 1941 reignited the urgency to sort and select military personnel effectively and efficiently. The U.S. Army had utilized an adapted IQ test, called the Army General Classification Test, which would be administered to ten million men during the course of World War II (Lemann, 1999, p. 53). The U.S. Navy, for its part, devised a program that would identify servicemen who could demonstrate aptitude for performing highly technical work. Rather than sending these selected recruits directly into battle, the Navy would first provide them with special training at a college or university. The program was called V-12, and the Navy needed a team of experts to develop a secure, expedient, large-scale assessment instrument to sort the young men (Lemann, 1999, p. 55).

Although the College Board was based in New York City, its mental testing unit had long been ensconced at Princeton University, alongside the inventor of the SAT, Carl Brigham, who led the operations. When Brigham died in 1943, he was replaced by his deputy at Princeton, John Stalnaker (Lemann, 1999, p. 54). The network of psychometricians in the 1940s was still relatively small, and Stalnaker was well connected with the most prominent figures in the field of psychological testing. Key among them was Harvard Assistant Dean, Henry Chauncey, who had been instrumental in prioritizing the SAT in James Conant's assessment scheme for the National Scholarships at Harvard in 1934. Like many notable Americans in academe at the time, Chauncey, Stalnaker, and Conant became indelibly involved with the U.S. military during World

War II. Together, the three men would play an important role in securing the military contract for the Navy's V-12 program for the College Board.

Similar to the first World War, the military buildup in the 1940s provided a unique opportunity for psychometricians to make advancements in mental testing. The V-12 program, with its focus on placing candidates into colleges for training, aligned remarkably well with the expertise and experience of the College Board. However, the College Board was not without competitors. The American Council on Education (ACE) was also vying for the chance to develop the Navy's placement exam (Lemann, 1999, p. 55). The Navy contract was not just lucrative; it also offered Stalnaker and Chauncey the chance to push the scope of large-scale testing to new limits. With very little time and very much at stake, Chauncey enlisted James Conant to wield his influence in Washington. In short order, the College Board was awarded the Navy contract, but Chauncey soon found himself racing to develop and securely administer the new test (Lemann, 1999, p. 56).

Less than two months after being awarded the Navy contract, Chauncey had the test for the V-12 program ready to go. But by this time the stakes had increased. The Army had developed a program similar to the Navy's V-12, and now it too wanted to use the College Board's newly developed test. The new test, known as the Army-Navy College Qualification Test, was in many ways a variation on an old one. First administered on April 2, 1943 to a nationally diverse audience of 316,000 high school seniors, the Army-Navy College Qualification Test was mostly composed of adapted items from the existing SAT (Lemann, 1999). By all appearances, the test administration was a huge success, and Chauncey would continue his work with the Navy until June 1944. "By demonstrating that an SAT-like test could be successfully administered at one sitting to a group larger (by a factor of forty!) than had ever

taken it before, Chauncey had made a nationwide multiple-choice personnel scan of the entire American population into a real possibility" (Lemann, 1999, p. 57). The sheer magnitude of the initial administration of the Army-Navy College Qualification Test revealed powerful new possibilities for large-scale assessments.

The implications of the Army-Navy College Qualification Test on the SAT and the future of standardized testing would prove immediately important after the war. As the war drew to a close, American higher education was on the brink of a major disruption. In 1944, Congress passed the Servicemen's Readjustment Act, which attempted to mitigate the economic impact of millions of Americans returning from war. Commonly known as the GI Bill, the Servicemen's Readjustment Act provided financial incentives for returning soldiers to enroll in college. The intent of the GI Bill was to relieve pressure on the postwar labor market by providing an outlet for prospective jobseekers outside the labor force (Thelin, 2011, p. 262). From Congress's point of view, higher education could serve as a kind of way station for the surplus workforce coming home, allowing the nation's industries the necessary time to regear and adjust to the new postwar economic landscape. Many leaders at America's most prestigious higher education institutions were adamantly opposed to the GI Bill, including James Conant at Harvard, fearing a saturation of spaces and a diminishing of standards (Leman, 1999, p. 59). Even the bill's supporters underestimated its ultimate impact on college enrollment, anticipating that a mere 8–10% of veterans would make use of the federal program. But by 1950, roughly 16% of eligible war veterans had enrolled in higher education institutions under the auspices of the GI Bill (Thelin, 2011, p. 263).

The colleges themselves played an important role in the GI Bill's success and appeal. Many institutions, including elite schools like Harvard, explicitly promoted their programs to

returning veterans. In addition, admissions standards were somewhat different for GI Bill recipients, with schools acknowledging the unique qualities of wartime experiences that most veterans had recently acquired. The flexible requirements and compelling advertising from the nation's colleges were important in drawing large numbers of veterans to higher education after the war. But the GI Bill also had important features that played into its success. As an entitlement, the GI Bill had no limits to the number of eligible veterans who could participate. All eligible veterans could take advantage of the education benefit without having to compete for a finite number of awards. The GI Bill was immediately accessible to any eligible veteran. Furthermore, the GI Bill was portable, which allowed veterans to apply the education benefit to any federally approved school of their choice (Thelin, 2011, pp. 263–265). The ease of access and freedom to choose made the GI Bill a popular option for many returning from the war.

The GI Bill was a demographic turning point for American high education. Although enrollments had been steadily increasing during the first half of the 20th century, the advent of the GI Bill accelerated the pace and scope of enrollment numbers like never before. By 1946, many of the nation's colleges experienced a doubling of enrollment (Thelin, 2011, p. 264). The shift in numbers also meant a shift in public attitudes about college attendance. Thanks in no small part to the GI Bill, going to college gradually started to look like a viable pathway for many Americans. Higher education was no longer an experience strictly reserved for a select few. With enrollments growing at a phenomenal rate, America's institutions of higher education recognized a greater need for large-scale admissions tests to sort through the deluge of applicants.

At the end of World War II, James Conant still had not given up on his dream of creating an independent national testing agency. The Harvard president saw the College Board as being

too regional, with most of its membership and influence coalesced in the schools of the Northeast. The rapid rise of college enrollments from the GI Bill and wartime advancements in testing technologies created a perfect storm of need and capability for more standardized testing. However, Conant's eventual success in pushing for a new national testing organization was more the product of power politics and connections than a campaign of carefully crafted arguments about the relevance or urgency for such an agency.

Conant's influence in education was at its height at the end of the war. He had years of experience as Harvard's president and had worked intimately with the U.S. military's science division during World War II (Lemann, 1999, pp. 60–61). At the end of the war, Conant also had a longtime colleague and ally running the mental testing unit for the College Board at Princeton. In 1944, Johns Stalnaker had left his post as the head of the College Board's mental testing unit, moving to Stanford University to become dean of students (Lemann, 1999, p. 57). Replacing Stalnaker at the College Board was none other than Henry Chauncey, who was just coming off his work developing the Army-Navy Collegiate Qualification Test for the U.S. military. Conant and Chauncey would soon find themselves working together, and with mutual ambition, to coordinate the foundation of a new independent testing organization.

At this time, the power of the purse in educational testing fell squarely on the Carnegie Foundation for the Advancement of Teaching, which led the country in promoting standardized tests (Lemann, 1999, pp. 60–61). The Carnegie Foundation established the Graduate Record Examination (GRE) in 1936 to support graduate programs in their admissions decisions. Carnegie had also provided funding and support for other testing initiatives around the country. In 1945, the president of the Carnegie Foundation, Devereaux Josephs, reached out to James Conant to discuss the viability of having the College Board take over the GRE. In actuality, the

meeting served as a pretext for suggesting the establishment of a testing agency that could control all the principal educational tests for the nation (Lemann, 1999).

With the Carnegie Foundation on board, James Conant was certain to get his new testing agency. However, when Devereaux Josephs sought approval from the American Council on Education (ACE) to put the GRE under the auspices of the College Board, he faced considerable backlash. The leadership at ACE, which saw itself as the preeminent institution for education in the country, balked at allowing the seemingly elite and regional College Board to take control of the GRE. Conant, for his part, realized that he needed to do something to placate the powers-that-be at ACE. After being appointed by the Carnegie Foundation to lead a special committee to research the proposed maneuvering, Conant recommended the creation of an entirely new agency that would be controlled by ACE. Subsequently, and to Henry Chauncey's horror, the committee recommended that the College Board and its tests be placed under ACE's control as well (Lemann, 1999, pp. 61–63).

Privately, Conant was able to pacify Chauncey by explaining the subtleties of the committee's recommendation. For appearances sake, the new agency and the College Board would fall under ACE. But in practice, the control of the tests—and ACE—would be the College Board's. The College Board, however, failing to fulling accept Conant's assurances, proposed that the new testing agency be completely independent, both from the College Board and ACE. This result was a nonstarter for the leadership at ACE, but Conant still had Devereaux Josephs at the Carnegie Foundation to help turn the screws. In the end, the Carnegie Foundation threatened to withhold funds from ACE, leaving ACE with little choice but to accept the resolution and opening the way for a new independent testing organization (Lemann, 1999).

The new testing agency, Educational Testing Service (ETS), would open on January 1, 1948, and it would administer the SAT for the College Board and take complete control of the GRE from the Carnegie Foundation. Henry Chauncey was appointed president of ETS, and James Conant served as its first board chairman. Based in Princeton, New Jersey, ETS would become the preeminent testing organization in the U.S., administering and developing tests with national and, eventually, international reach. The establishment of ETS was a personal victory for both Conant and Chauncey, who each held distinct views on what the agency could be for America. Conant saw testing as the key to discovering the nation's talent among the masses, while Chauncey envisioned a premier research space to examine and advance the technologies of testing (Lemann, 1999, p. 43). Both men were able to see their visions of standardized testing in the establishment of ETS. ETS and its testing products would not be without their controversies in the decades to come, but the dream of an independent national testing agency had finally taken shape.

ETS at the Ready

The conclusion of World War II in 1945 marked the beginning of American higher education's "golden age," an era that would run for roughly twenty-five years. The principal catalyst of this golden age in higher education was the GI Bill, which greatly expanded college enrollments after World War II by providing an educational entitlement to all eligible veterans. Naturally, not every war veteran chose to claim the educational benefit of the GI Bill, but those who did contributed to the swelling college enrollments throughout the country. This initial shift in college attendance, one that promoted enrollments from a much broader swath of America than ever before, marked the beginning of America's changing attitudes and expectations about higher education. For the first time, the public imagination started recognizing higher education

as something accessible, desirable, and, most importantly, attainable. Between 1940 and 1950, college enrollments in the U.S. nearly doubled, from less than 1.5 million to nearly 2.7 million. American higher education added almost another million students by 1960 and then doubled to 7.9 million in 1970 (Thelin, 2011, pp. 260–261). College enrollments ballooned by a factor of five in the twenty-five-year span of American higher education's golden age.

The nature and breadth of American higher education institutions also changed and expanded during this "golden age" period. Regional state and community colleges grew to accommodate the rising trend of mass education, research universities flourished and became more deeply entwined with government and industry, and eventually for-profit schools emerged to vie for a piece of the higher education market. Regardless of how disparate these institutions might have appeared—from academically selective to open-admissions—most schools at this time shared a ubiquitous admissions feature: standardized testing (Thelin, 2011, p. 301). In most cases, the standardized test of choice was the Scholastic Aptitude Test (SAT), owned by the College Board and maintained and administered by Educational Testing Service (ETS).

ETS was off to a financially rocky start in 1948, but a major government contract in 1951 quickly brought the fledgling testing agency to solvency. The U.S. military, responding to the outbreak of war on the Korean peninsula, sought to administer IQ tests to college students for the purpose of military deferments. The scheme was devised to defer military service of America's brightest college students, ostensibly to preserve an intellectual elite whose talents could be tapped for impending weapons development. ETS's chairman of the board, James Conant, and the popular press cried foul, immediately recognizing the grisly, undemocratic implications of the military's stratagem. However, ETS's president, Henry Chauncey, capably massaged public opinion, emphasizing that the test was sorting people for deferment, not exemption, and he

cleverly rebranded the test as one of "scholastic aptitude," not a measurement of intelligence. In the end, Chauncey's hard-fought public relations effort paid handsomely for ETS, with a massive contract with the Selective Service System to produce and administer the test to college students (Lemann, 1999, pp. 73–77). The draft deferment test also helped to establish ETS's reputation as the premier testing company in the country, and the experience of promoting a test for the sake of testing, without quibbling over the merits of whether or not to test, set a precedent for the ethos of ETS to come. Above all else, as a self-supporting organization, ETS would need to adhere to a principle of more testing and more tests in order to ensure its survival and success (Lemann, 1999, pp. 77–78).

The Selective Service System's draft deferment exam helped put ETS on a path to becoming the leader in standardized testing in the U.S. In the decade that followed, ETS would continue to make advancements in standardized testing, researching elements of test security, administration, and issues of validity and reliability. Advancements in computing also expanded the capabilities of large-scale testing, allowing for more rapid and accurate scoring. IBM began developing a machine for scoring tests in 1928 and first used their machine to score exams for public schools in Providence, Rhode Island and the New York State Regents in 1936 (Lemann, 1999, p. 32–33). After continued advancements, IBM's machines were used to score large-scale educational tests for the first time in the 1950s (Lemann, 1999, p. 77).

ETS had few competitors during the 1950s, holding a virtual monopoly on college entrance exams (e.g., the SAT and GRE). In 1959, a competing college entrance exam joined the fray, American College Testing (ACT), but ETS would continue to be the leader in large-scale "scholastic aptitude" tests. ETS's preeminence in the educational testing world meant that it was able to attract and retain many of the nation's top talent and leading experts on testing and

psychometrics. The team of experts at ETS worked on test development and conducted correlational research studies, which supported claims that the SAT predicted students' college performance. For any organization hoping to develop a new test, ETS was obviously the first place to turn for consultation.

Considering the outsized sociocultural impact and sheer ubiquity of products like the SAT and GRE, it might be hard to imagine a more recognizable and prodigiously administered college admissions test. However, the rapidly changing demographics of American higher education between 1945 and 1970 would hasten a new kind of standardized test for undergraduate and graduate admissions in the U.S., one that few Americans would ever know about or ever be required to take. By the 1960s, America's college and universities were in search of a more efficient and reliable way to assess the English language abilities of prospective international students whose first language was not English. Ultimately, ETS would be charged with developing and administering the test. In 1964, the very first Test of English as a Foreign Language (TOEFL) was administered to 920 examinees at 57 test centers around the world (Taylor & Angelis, 2011, p. 28), setting ETS on a course to be a global leader in educational testing.

Chapter 4:

The Invention of the Test of English as a Foreign Language

The TOEFL would eventually become ETS's largest and most successful product, but its initial inception and development was only made possible by the rapidly rising number of international students wishing to study in the U.S. after World War II. Prior to the TOEFL, American institutions of higher education had no reliable or efficient way to evaluate the English language proficiency of non-native speakers of English. In order to study in the U.S., international students would need to have high levels of English reading, writing, listening, and speaking skills to successfully navigate the American college curriculum. The TOEFL eventually would serve as an effective tool for gauging the language readiness of international students, giving college admissions councilors clearly defined cut-scores for making undergraduate and graduate admissions decisions. However, like most high-stakes tests, the TOEFL's influence would resonate beyond its intended purpose. The TOEFL would become a key factor in international student admissions, but its importance would also precipitate a worldwide demand for test preparation materials and services, influencing second language pedagogy and elevating the status of the English language.

A Fertile Landscape for English Language Testing

The first 250 years of American higher education provided scant incentive or opportunity for students to study in the U.S. from abroad. U.S. colleges and universities were largely seen as inferior to their European counterparts, and the institutions themselves had few, if any, formal support mechanisms for international students. The enrollment of Francisco de Miranda, a student from Venezuela, at Yale in 1784 marked the first documented international student in

America, but the flow of international students to follow would remain uneven and small throughout the 19th century (Bevis, 2019). The "emergence" of the American university, as Laurence Veysey (1965) identifies it, helped shift the fortunes of American higher education by gradually prioritizing science, *research*, and *utility* in the curriculum, moving away from a strict ethos of *discipline and piety*. The improved quality and purpose of American higher education, coinciding with advancements and successes in American industry and agriculture, recast the global spotlight onto the U.S. By the 20th century, the U.S. was beginning to look more attractive to international students pursuing higher education.

As U.S. colleges and universities "emerged" at the end of the 19th century, taking on many of the characteristics that would define them in their modern-day form, important changes were also taking place in U.S. immigration policy. Prior to the Chinese Exclusion Act of 1882, the U.S. federal government had very few legal restrictions on immigration. Up to this point in time, the small number of international students who traveled to the U.S. to study at a college or university faced virtually no bureaucratic challenges to their entry into the country. For the most part, international students only needed to coordinate with their receiving institutions in the U.S. to attend school. But the creation of federal immigration legislation quickly changed the free flow of immigrants into the country, whether they be emigrants or visiting students and scholars. The odious and overtly racist Chinese Exclusion Act of 1882 specifically targeted immigrants of Chinese descent, but it also necessitated new administrative capabilities in the federal government to oversee and enforce the nascent immigration laws. With each new immigration law to follow, the federal government's immigration enforcement system grew larger, more complex, and increasingly sophisticated. By the 20th century, foreign students wishing to study in

the U.S. would have to interact closely with the U.S. Immigration offices to properly present their credentials for entry into the U.S.

In 1903, the first examples of formalized institutional support for international students began to appear on U.S. college campuses. Cosmopolitan Clubs, initiated by international students in response to their shared challenges and circumstances, served as forebears to the institutional support services that would emerge on many campuses by the 1930s (Bevis, 2019, pp. 110–112). The Cosmopolitan Clubs helped raise institutional awareness of international students and the unique set of experiences and needs they faced while attending school in the U.S.

Interest in international students grew considerably in the first decades of the 20th century, and a number of organizations formed to study and address international students and topics pertaining to international exchange. By 1925, there were more than 115 privately funded organizations in the U.S. that were focused to some extent on international educational exchange (Bevis, 2019, p. 104). In response to the senseless devastation of the World War I, the Institute of International Education (IIE) was created in 1919 to elevate educational exchange as a panacea against global misunderstanding and conflict. The IIE would later contribute to the creation of the National Association of Foreign Student Advisors (NAFSA) in 1947, which helped formalize the professionalization of student services personnel working with international students (Bevis, 2019, pp. 149–150). As U.S. higher education was entering its "golden age" in 1945, international students in the U.S. had acquired a new level of attention from American universities, the U.S. government, and a growing number of organizations dedicated to international exchange and education.

Estimates of international student enrollments in the U.S. are generally considered imprecise prior to 1949. Although the U.S. Department of Education had a longstanding interest in comparative education, it had little direct involvement with monitoring international student flows to the U.S. before the end of World War II (Bevis, 2019, p. 103). Most of the data that inform international student enrollments before 1949 represent potentially uneven or incomplete survey results administered to a portion of U.S. colleges and universities by private organizations. The U.S. Department of Education had not attempted to maintain a census on international students. But in 1947, the IIE, in coordination with the Committee on Friendly Relations Among Foreign Students, began a project to conduct a more accurate census of foreign students attending college in the U.S. The conclusion of World War II had reinvigorated international exchange, and U.S. colleges and universities were seeing more international students than ever before (Bevis, 2019, pp. 135–136). Despite pre-1949 estimates lacking precision, there is a reasonable pattern of international student enrollment growth. Likewise, data on international student enrollment leading up to and throughout the war years reveal an understandable period of attrition. But immediately following World War II in 1946–1947, estimated international student enrollment skyrockets to nearly 50% above its pre-war peak in 1930–1931. At the time of the first reliable census of international students in 1948–1949, conducted by IIE, the number of international students had doubled from the estimated figures in 1946–1947 (Bevis, 2019, p. 136). Clearly, international students were becoming a more prominent demographic in American higher education after World War II.

World War II had introduced the world to new excesses in human suffering, mass murder, and technological terror. But the war also left the United States in a position of geopolitical and economic strength. The U.S. had engendered deep admiration and gratitude

from its European allies, it had pulled itself out of the Great Depression, and—for the time being—it was the sole custodian of the most horrific military weapons the world had ever seen.

America's colleges and universities, too, were in excellent standing after the war. Institutions of higher education in the U.S. had contributed significantly to the war effort, providing an invaluable well of expertise and innovative brilliance from which the U.S. military could draw. From developing standardized tests (e.g., Henry Chauncey and John Stalnaker) to harnessing the atom (e.g., Richard Feynman and Robert Oppenheimer), America's academic talent served the U.S. military to great, and lethal, effect. American higher education also benefited from the migration of Jewish intellectuals fleeing Europe during World War II. Circumstances for Jews in Germany became untenable well before Hitler's invasion of Poland in 1939, and some European Jews were able to survive the brutality and persecution of the Third Reich by emigrating to the U.S. Europe's brain drain during the war years added to an already distinguished group of intellectuals and scientists working in American higher education, helping to further bolster the reputations of U.S. colleges and universities immediately after the war.

Within a generation, the world had experienced two unprecedented and catastrophic global conflicts. As with the aftermath of World War I, the second world war prompted extensive reflection, debate, and discussion on how the world could maintain a lasting and durable peace. Similar to the intentions behind the creation of the Institute of International Education (IIE) in 1919, many believed that education and cultural exchange could serve the interests of global goodwill and understanding in the wake of World War II. Senator James William Fulbright, having been himself a Rhodes Scholar at Oxford University, recognized the power of international exchange to promote mutual understanding and friendship across nations. With the war coming to a close in 1945, Fulbright introduced legislation to establish a

scholarship program for international educational exchange. Signed into law in 1946 by President Truman, the Fulbright Program became the preeminent U.S. scholarship for cultural, educational, and scientific exchange.

The Fulbright Program offered new opportunities and incentives for international students and scholars to study and research in the United States, and it introduced a model for many other international exchange scholarship programs to follow. Initially, funding for the Fulbright Program only accounted for American students and scholars studying overseas, but by 1948, with the passing of the Smith-Mundt Act, the program had the funds to support a truly reciprocal flow of international exchange between the U.S. and other countries. Rather than managing the Fulbright Program itself, the U.S. Department of State's Bureau of Education and Cultural Affairs turned to the IIE and its sister organization, the Council for International Exchange and Scholars (CIES), to maintain and operate the program. Like the Fulbright Program itself, the IIE was created as an instrument for furthering international understanding and friendship, a mission that would include new and expanded responsibilities with the Fulbright under its charge (Bevis, 2019, pp. 141–142). By the 1950s, with new avenues for scholarship and a rapidly increasing global reputation, America's higher education landscape had become especially attractive and inviting for international students.

The 1948–1949 academic year marked the beginning of the IIE's annual census of international students in the U.S., and the data accumulated by the IIE in the decades that followed reveal a steady increase of foreign students on American college campuses. In the first IIE foreign student census (1948–1949), there were 25,464 international students studying in the U.S. By the 1960–1961 census, twelve years later, that figure had more than doubled to 53,107 (Institute of International Education, 2021a). Throughout the 1950s, America's colleges and

universities enjoyed a steady increase in international student enrollments, prompting some college campuses to develop or expand support systems for international students. Administrators were also becoming increasingly frustrated with the challenges of assessing international students for admission. Each country posed its own system of educational standards and expectations, often intertwined with distinct cultural norms and values. In order to admit an international student, U.S. colleges and universities were tasked with determining whether a student's educational records adequately matched the admission requirements of their institution.

Fortunately, at the time, American colleges and universities could get assistance from the U.S. Department of Education's division of International Educational Relations. Established in 1918, the International Educational Relations Division was equipped to help higher education institutions determine the educational equivalencies of other countries (Bevis, 2019, p. 152). Thanks in no small part to America's evolving immigration laws, the U.S. government had taken an outsized role in the administration of international students. Other U.S. departments (e.g., Education) developed their own bureaucratic elements that attended to international students, often in coordination with U.S. Immigration. In addition to the Department of Education, American colleges and universities could also solicit help from organizations like the IIE, the American Association of Collegiate Registrars and Admissions Officers, and NAFSA to evaluate international students' credential (Bevis, 2019). Remarkably, throughout most of the 20th century, the U.S. government provided substantive assistance to America's institutions of higher education in evaluating the documentation of international students.

As the numbers of international students increased in the years and decades following World War II, the U.S. experienced a gradual shift in the proportion of international students whose first language was not English (Institute of International Education, 2021b). In 1945, there

was no efficient or reliable way for college admissions councilors to determine the English proficiency of the international students they were evaluating. Students applying from overseas were often interviewed by personnel at U.S. embassies and consulates to determine their English proficiency before departing for the U.S. By 1961, the U.S. government had become increasingly concerned with its evaluative responsibilities of international students, especially those students who were unsponsored or privately sponsored. Without "some standardized and reliable way of assessing candidates' English proficiency," the U.S. government recognized that the increases in international student applications were leading to "personal hardship, institutional waste, and bureaucratic frustration" at U.S. posts abroad (Spolsky, 1990, p. 101). The colleges, for their part, were equally concerned with the problems of English language evaluations. Language assessment results were often difficult to interpret or compare, and many evaluations of English proficiency relied too heavily on subjective evaluations from non-experts. The sheer variety of testing conditions also raised perennial questions about test security (Spolsky, 1990.). At the start of the 1960s, U.S. higher education and the U.S. government were in agreement about the need for some kind of standardized language assessment for international students whose first language was not English.

Clearly, the time for large-scale, standardized English language tests had arrived. American higher education had experienced a profound improvement in its reputational stature since the "emergence" of the university at the end of the 19th century. The fortunes of the university were closely linked with those of the U.S. after World War II, and both the country and its system of higher education had attracted admirers from around the world. With improved government coordination, the U.S. was able to establish organizations that focused on international educational exchange and eventually introduce government-sponsored foreign

exchange scholarships. As American college campuses became more inviting for international students, enrollments continued to rise. American higher education, having already experienced swelling enrollments in the wake of the GI Bill, would also enjoy a steady increase in international students throughout its "golden age" and beyond. In order to evaluate the English language proficiency of this growing international student population, a variety of stakeholders would turn to America's leading company for developing educational tests. With its wealth of expertise and experience, Educational Testing Service (ETS) was well positioned to address the new testing challenges posed by America's universities at the beginning of the 1960s.

The First TOEFL

The aspirations of Senator Fulbright for enhanced global understanding and peace were swiftly tempered after World War II. On March 12, 1947, roughly eight months after the Fulbright Program was signed into law, President Truman announced the Truman Doctrine to Congress, marking the beginning of the Cold War with the Soviet Union and precipitating the formation of the North Atlantic Treaty Organization (NATO). By the end of 1949, the Soviets had successfully detonated an atomic bomb for the first time, signifying an escalation in military buildup. By the time President Truman left office in 1953, the U.S. had already been embroiled in another foreign war, this time on the Korean Peninsula.

President Dwight D. Eisenhower was inaugurated on January 20, 1953, and, in less than two months, Joseph Stalin was dead. The leadership of the two Cold War superpowers abruptly changed, but the military arms race persisted. The Truman Doctrine's policy of communist containment meant a heightened level of global engagement for the United States, which included the promotion of American-style democracy around the world as the gold-standard for representative government. Nevertheless, the 1950s exposed the world to numerous

imperfections in the American system, including the abuse of institutional powers in the hands of Senator Joseph McCarthy and the legacy of inequality, injustice, and violence against racial and ethnic minorities.

On October 4, 1957, the race for military supremacy, and the battle for the hearts and minds of the world, were briefly eclipsed by a remarkable achievement in Soviet technology. The Soviet's had successfully launched the first artificial satellite into Earth's orbit. Dubbed, *Sputnik*, the satellite had no immediate military application, but the accomplishment was deeply unsettling to the U.S. Within a month, the Soviet's launched a second satellite, *Sputnik II*, which was six time larger and orbited even higher than the first *Sputnik*. Despite a range of anxious, if not hysterical, public reactions, Eisenhower remained calm and managed to put the nation at ease, insisting that the Soviet's had not surpassed America's military capabilities (Smith, 2012).

Two years prior, Eisenhower had insisted that Congress pass legislation to provide federal funding to the states for school construction. The bill had stalled in Congress, but *Sputnik* provided the impetus to revisit the objectives of the legislation. On September 2, 1958, Eisenhower signed the National Defense Education Act into Law. The bill provided substantial federal funds to support and improve American education, including funding for graduate fellowships and public school instruction in science, technology, engineering, mathematics, and foreign languages (Smith, 2012). Most importantly for American higher education, the National Defense Education Act provided new funding opportunities for research and institutional development throughout the 1960s (Thelin, 2011, pp. 279–280).

The National Defense Education Act of 1958 included foreign language study alongside a variety of STEM fields, elevating second language acquisition as a national security priority. Naturally, some languages were of greater importance to national defense than others in 1958,

particularly Eastern European languages. However, the U.S. understood that the country would require capable speakers of a variety of languages to maintain foreign relations and provide strategic defense capabilities. Likewise, the study of how languages are learned would also amplify in importance. Linguistics and applied linguistics would be serviceable to the U.S. government, helping to guide more efficient and effective foreign language pedagogy and proficiency assessments.

The U.S. government had already gained some useful experience in foreign language instruction and testing before 1958. Like other advancements in standardized testing, wartime circumstances hastened the development and use of foreign language assessments. In 1942, the U.S. Army established the Army Specialized Training Program (ASTP) to prepare military personnel for communicating in foreign languages within specific fields or areas, such as engineering or medicine. Unique to language instruction at the time, the ASTP focused more on language-in-use rather than on rote memorization and grammar-translation. Prioritizing spoken language, trainees were directed to learn practical colloquial forms and attain a strong understanding of the relevant contexts (e.g., medicine). The ASTP relied on raters and rubrics, not pen and paper tests, to assess the spoken language proficiency of trainees. Approximately 140,000 soldiers underwent ASTP instruction in 1943, but the program was suspended in April 1944 by the U.S. government (Fulcher, 2003, pp. 6–7).

In 1952, the U.S. Foreign Service Institute (FSI) began developing foreign language assessments for its personnel. Responding to new documentation requirements by the Civil Service Commission, the FSI started work on what would become the first publicly available and widely used rating scale for speaking. Red tape and politics held up the test development project until 1956, when survey evidence suggested that many of the FSI-trained personnel lacked the

language skills to adequately perform in their roles. The FSI's Oral Proficiency Interview (OPI) became the first published test of spoken language, and, by the 1960s, confidence in the OPI was so high that the techniques were eventually adopted by the Peace Corps, the Defense Language Institute, and the Central Intelligence Agency (CIA). Even at the inception of the OPI in 1952, test developers acknowledged the important role that linguistic and communicative competence plays in evaluating oral language proficiency (Fulcher, 2003, pp. 8–10).

Lessons learned in the development of foreign language proficiency tests would later inform design considerations for the TOEFL, but the U.S. also had a brief history with English language test development. The Immigration Act of 1924 placed new quotas on the numbers of immigrants entering the U.S. One exception to the quotas, however, pertained to foreign persons entering the U.S. temporarily to study at a school, college, or university (Bevis, 2019, p. 103). In 1926, the U.S. government added a language requirement to prospective foreign students, requiring receiving institutions to document and report students' knowledge of the English language before being admitted (Fulcher, 2003, p. 2). The 1926 English language requirement was a response to the exception for foreign students in the 1924 legislation, which some saw as a loophole (Spolsky, 1993, p. 3). Evaluating language proficiency was meant to be restrictive, preserving the intention of the 1924 Immigration Act to keep foreigners out (Spolsky, 1993, p. 10). In response to the 1926 immigration rule, the Association of College Registrars reached out to the College Board for assistance in designing an English language proficiency test (Fulcher, 2003, p. 2). Developed by the College Entrance Examination Board (CEEB) between 1927 and 1930, the English Competence Examination became perhaps the first institutional test of English as a second language in the U.S. (Spolsky, 1993, p. 2). Like many testing instruments, the English Competence Examination was relatively expensive to develop and administer. As the

numbers of candidates began to diminish, the cost of administering the test became unfeasible. By 1934, the English Competence exams had been suspended for lack of funds (Spolsky, 1993, p. 10). Nevertheless, the test was a milestone in both English language testing and America's anti-immigration sentiment in the 1920s.

By the 1960s, U.S. foreign interests were deeply intertwined with economic prosperity and national security. The world, too, looked to the U.S. for expanding opportunities, and foreign students came to America's colleges and universities in increasing numbers. Burdened by the rising interest in American higher education from abroad, colleges and universities would look to more efficient, objective, and expedient ways to assess the English language proficiency of prospective foreign students. Given that the language proficiency of pre-departure students was often assessed at U.S. embassies and consulates abroad, the U.S. government also had a strong interest in improving the language tests, as well as shifting the onus of testing solely onto the universities.

Seeking to address the rising need for standardized English proficiency testing at colleges and universities, the Center for Applied Linguistics (CAL), in coordination with IIE and NAFSA, sponsored a testing conference in Washington, DC, in 1961 (Spolsky, 1990, p. 99). The conference brought together experts and stakeholders from a variety of universities, institutions, government offices, and organizations, including Georgetown, Harvard, Columbia, University of Michigan, UCLA, University of Texas, the Language School of Lackland Air Force Base, the Washington Institute of Psychiatry, the Modern Languages Association (MLA), CAL, ETS, CEEB, IIE, NAFSA, the Ford Foundation, the U.S. Department of Health Education and Welfare, U.S. Department of State, and the British Embassy (Spolsky, 1990, p. 115). The plans for the original Test of English as a Foreign Language (TOEFL) would germinate from the

presentations and discussions at the CAL-sponsored conference, which reflected on extant language assessment programs and theoretical issues pertaining to English language tests. By the end of the conference, the participants largely agreed that test development should prioritize assessing the English language proficiency of pre-college, undergraduate, and graduate-level students (Taylor & Angelis, 2003, pp. 27–28). Unlike the SAT or GRE, the TOEFL would not be specific to any one category of student.

One year after the CAL conference in Washington, the National Council on the Testing of English as a Foreign Language was established. Comprised of representatives from more than thirty private organizations and public agencies, the council continued to refine the plans for a language proficiency test for international student admissions. The TOEFL was finally initiated in 1963 with funding from the Ford Foundation and the Danforth Foundation. MLA was tasked with managing the TOEFL development program. Finally, in 1964, the first TOEFL was administered. 920 prospective students took the first TOEFL across 57 test centers around the world. By the following year, the administrative and operational responsibilities would be handed off to ETS and the College Board. As demand for the TOEFL grew, ETS would make changes to the TOEFL to accommodate larger pools of test-takers. In 1973, ETS assumed full responsibility for the development, finances, and administration of the TOEFL. However, the original stakeholders were able to maintain some measure of control on testing policy by establishing a TOEFL Policy Council, which included members who represented the College Board, the GRE Board, and other agencies and organizations with vested interest in international student admissions (Taylor & Angelis, 2011, p. 28).

The original TOEFL was a five-section exam, focusing on (1) Listening Comprehension, (2) English Structure, (3), Writing Ability, (4) Reading Comprehension, and (5) Vocabulary. For

the sake of objectivity, the test did not solicit writing samples or collect oral responses. The absence of productive language responses all helped keep costs down and expedited scoring. Naturally, questions about validity of the test emerged early on, and ETS established a TOEFL Research Committee to address concerns over the new test and make improvements for future iterations (Taylor & Angelis, 2011, p. 32). Much of the early research focused on correlational studies to understand how well the TOEFL aligned with language production. The correlational data largely supported a strong relationship between the TOEFL Listening Comprehension section and oral interviews, as well as a strong relationship between the multiple-choice Writing Ability section and essay samples. Thus, ETS felt justified in continuing the TOEFL as a strictly multiple-choice exam. However, ETS also used the research to inform their decision in 1975 to consolidate the TOEFL into a three-section test, basically combining the English Structure section with the Writing Ability section and the Reading Comprehension section with the Vocabulary section (Taylor & Angelis, 2011, p. 32; Chapelle, Enright, & Jamieson, 2011, p. 360).

By the mid-1970s, ETS had full control over the TOEFL administration. As international student enrollments increased each year, the TOEFL grew in demand and perceived importance. The TOEFL was quickly becoming internationally famous for its powerful role in satisfying the gatekeepers at U.S. college and universities. In the nine years between the first TOEFL administration in 1964 and ETS's full control of the test in 1973, the number of international students attending American colleges and universities had nearly doubled, from 74,814 students (1963–1964) to 146,097 students (1972–1973). By the 1980s, the figure would nearly double again, to 286,343 (1979–1980) (Institute of International Education, 2021a). With each passing year, the proportion of international students whose first language was not English would also

increase (Institute of International Education, 2021b). The TOEFL was becoming an important admissions instrument across both undergraduate and graduate education. However, questions would continue to persist about its validity and reliability for assessing students' language skills for university work. Interest in the TOEFL would extend well beyond the research division at ETS, and eventually include college administrators, admissions officers, and the professors and English language instructors working closely with international students.

Revising and Renewing the TOEFL

Standardized tests have long faced criticisms, complaints, and challenges. Not long after ETS's inception, an underemployed man from Brooklyn, NY named, Stanley H. Kaplan, decided to get a little extra work by tutoring the neighborhood kids for the SAT. The results were excellent, and Kaplan continued to attract clients and was subsequently able to establish a small business in his name. That small business, Kaplan, Inc., has since become a behemoth in the world of test-preparation, but initially ETS refused to believe that their test was coachable. But Kaplan's results didn't lie, and ETS would eventually have to concede that test results could be improved through dedicated study and practice (Lemann, 1999, pp. 110–112). ETS would also have to recognized that the magnitude and success of their test development enterprise would spur other types of peripheral industries, like businesses and publishers dedicated to test preparation.

By the time Henry Chauncey retired from ETS in 1970, nearly two million high school students were taking the SAT every year. The SAT had long become a household name, but its continued fame drew more critics and critiques. Probably the most famous critic of ETS in the 1970s was Ralph Nader, whose investigation of ETS and the SAT led to legislation in New York that forced greater transparency at ETS (Lemann, 1999, pp. 218–221). The SAT aroused

concerns from students, parents, teachers, journalists, former ETS employees, and consumer advocates. The sheer ubiquity of the SAT made stories of its potential shortcomings and possible abuses resonate with a wide American audience.

The TOEFL, on the other hand, did not lend itself as easily to coordinated consensus amongst test-takers. Prospective international students taking the TOEFL were separated by expanses of land and sea, national boarders, and languages. For test-takers, the TOEFL was a necessary part of applying to a U.S. college or university, and not something to challenge or question. Nevertheless, other stakeholders in the 1970s began questioning the merits of the TOEFL, pressing for ways to improve the test to make better admissions decisions.

The earliest investigations into the TOEFL came from ETS itself, under the TOEFL Research Committee. The objective of the Committee was to establish a research agenda around the TOEFL, which would review proposals and reports, as well as provide funding recommendations for the TOEFL Policy Council. A program budget at ETS provided the bulk of research support for investigating the TOEFL. The focus of the TOEFL Research Committee during the late 1970s centered primarily on making incremental improvements to the test and revisiting validity issues in assessing language production (i.e., speaking, writing) (Chapelle et al., 2011, p. 360).

By the 1980s, a modern conceptualization of language proficiency was beginning to emerge in the fields of applied linguistics and second language acquisition (e.g., Canale & Swain, 1980). The fields were moving toward a focus on "communicative competence," which would influence new and innovative communicative approaches to second language pedagogy. Of course, as a multiple-choice exam, the TOEFL did not prompt examinees with communicative tasks. Many outside ETS, in fields like applied linguistics, language assessment,

and Teachers of English to Speakers of Other Languages (TESOL), began raising issues about the TOEFL's content and validity, and the constructs that the TOEFL purported to measure.

Similar to the SAT, the TOEFL's success attracted an entire industry around test preparation. However, given the global reach of the TOEFL, the market for test preparation courses, tutoring, and publications for the TOEFL was far more extensive and international in scope. The preparation materials largely taught to the test, with an emphasis on memorization of grammar rules, vocabulary, and general test-taking strategies for multiple-choice assessments. The design of the TOEFL itself, meant to be efficiently administered and scored, did not reflect the communicative approach to language learning that was being championed by applied linguists and TESOL professionals in the early 1980s.

The TOEFL's success and subsequent ubiquity in higher education elevated its impact on everything from language teaching to language test design. Known as a washback effect, the TOEFL's perceived legitimacy as an assessment instrument and its obvious importance and value for university admissions influenced how English was being taught and how other language tests were being developed (Brown, 2005). American universities also took notice, as professors and administers began acknowledging that some admitted international students—even those with strong TOEFL scores—were struggling in the classroom. In response, many colleges and universities with larger intakes of international students established English Language Institutes (ELIs), Intensive English Programs (IEPs), or some similarly named department, to provide English language support and coursework for admitted students whose English abilities fell below undergraduate- or graduate-level expectations. Despite often having a good knowledge of English grammar and vocabulary, many of the students assigned to ELI or

IEP courses needed help improving their productive language skills (e.g., oral communication, academic writing).

In 1979, ETS introduced a new English language testing product alongside the TOEFL, the Test of Spoken English (TSE) (Chapelle et al., 2011, p. 360). The TSE was meant to complement the TOEFL and provide another data point for universities to consider. The TSE was an important first step to move language assessment away from strictly multiple-choice exams, introducing interview protocols and rubrics for evaluation. Subsequent research on the TSE in the early 1980s focused on its utility for its validity in a variety of contexts, including instructional settings, and for other populations, such as international teaching assistants and even health care professionals (Taylor & Angelis, 2011, p. 36.). By 1981, discussions had already begun on the development of an exam that would directly measure English writing proficiency. And between 1981 and 1983, the TOEFL Research Committee and TOEFL Committee of Examiners (COE) met annually to consider the implications of communicative competence on the TOEFL (Taylor & Angelis, 2011, p. 35). In 1984, funding was finally approved by the TOEFL Policy Council to develop a new writing assessment for the TOEFL line of products. As with the TSE, and for lack of a multiple-choice format, the new writing exam would need to develop interpretable scales to determine levels of proficiency. In 1986, the Test of Written English (TWE) was launched as a separate language assessment product alongside the TOEFL (Chapelle, 2011, p. 360). America's colleges and universities now had assessment options for evaluating the writing and speaking proficiencies of their international student applicants.

By 1990, the TOEFL was ETS's fastest growing test, and particularly prolific and profitable in the Asian market. For students wishing to attend university or graduate school in the

U.S., the stakes for doing well on the TOEFL could not have been higher. The combination of high stakes and high numbers of test-takers helped turn the TOEFL into the most cheated-on test in ETS's assessment lineup (Lemann, 1999, pp. 241–242). In addition to the longstanding questions over the TOEFL's effectiveness in evaluating communicative competence, American universities were becoming increasingly concerned about the veracity of international applicants' scores.

The 1990s would prove to be an important decade for the future of the TOEFL and the delivery capabilities of large-scale testing. Between 1990 and 1996, the TOEFL COE started allocating part of each meeting to discuss ideas for a new TOEFL test, granting the TOEFL Policy Committee oversight on the eventual project. In 1992, the TOEFL Policy Council approved development and funding for what would become known as *TOEFL 2000*. By 1993, the *TOEFL 2000* project team had already established a set of goals that would shape their agenda for the continuing phase of the project. Key to the new test would be a focus on communicative competence that solicited performance-based tasks. The project team also drew on focus groups and surveys to determine the needs of stakeholders, specifically institutions that needed to interpret and apply TOEFL scores (Chappelle, 2011, p. 360). In 1995, the project team proposed substantive changes to the TOEFL listening portion, including the removal of "single-statement" listening comprehension questions, which would be replaced with sections with samples of academic lectures and longer dialogues. In addition, vocabulary testing would no longer be testing with discrete items, but instead be prompted within the context of reading comprehension passages to emphasize vocabulary-in-context (Taylor & Angelis, 2011, p. 46).

Perhaps the most remarkable design shift in 1995 was the focus on making *TOEFL 2000* computer-mediated (Taylor & Angelis, 2011, p. 40). Ideas for using computers to deliver the

TOEFL emerged as early as 1986, but the technology was still too expensive and unreliable (Taylor & Angelis, 2011, p. 37). In 1995, the use of computers seemed especially promising, and test developers considered innovative ways for delivering test prompts and items, including audio and video clips. Likewise, productive-language responses could be collected in writing through the keyboard or even spoken and recorded into a microphone. The introduction of computers into large-scale testing created new opportunities for soliciting language production and improving the TOEFL's measurement of communicative competence. Computers also raised new questions about test security, but ultimately offered the promise of innovative solutions to security issues, both old and new.

Like many large undertakings, the *TOEFL 2000* project was running behind schedule. Revamping an established and popular test was complex and expensive, involving teams of experts on a multitude of development fronts. In the interim, TOEFL would implement a computer-based version of the paper-based test, which would include a new section for a writing sample. The first TOEFL Computer-Based Test (CBT) was introduced in 1998, and examinees were given the option of providing a handwritten or computer-typed response to the writing prompt (Wang, Eignor, & Enright, 2011, p. 276). The TOEFL Paper-Based Test (PBT) would be slowly phased out as the TOEFL CBT gradually expanded to more testing centers and institutions. The PBT would also live on as the Institutional TOEFL, administered by institutions across the world for localized (and lower stakes) purposes. ETS's Online Scoring Network (OSN), which provided a secure internet connection, went live in 1999 and allowed TOEFL raters to work from home and improve the convenience and opportunity for rating large numbers of exams quickly (Pearlman, 2011, p. 257).

The *TOEFL 2000* project was still in development in 2002, but by then the project team had shown that their computer-based test could be successfully delivered to a relatively small-scale audience (approximately 3,000 test takers) (Pearlman, 2011, p. 227). The new TOEFL was clearly coming along, and study materials were produced and disseminated by ETS to help acquaint stakeholders and prospective students with the test revisions and new integrated tasks (Wang et al., 2011, p. 311). Piloting continued for the next two years, with additional test preparation materials and manuals being released for the TOEFL's wide audience of teachers, curriculum coordinators, academic directors, admissions officers, and prospective students (Wang et al., 2011, pp. 304 & 311). By the time the new TOEFL was first administered in September 2005, *TOEFL 2000* had assumed a new name. ETS's new communicative academic English proficiency exam would be known as TOEFL iBT, Internet-Based Test (Wang et al., 2011).

In anticipation of the TOEFL iBT, ETS orchestrated a conference for American and European designers and publishers of English language test preparation materials. The goal for ETS was to emphasize the importance of study materials that "reflected the communicative constructs and task characteristics of the TOEFL iBT" (Wang et al., 2011, p. 311). Unlike the early days of ETS and the SAT, ETS no longer was at odds with the test preparation industry. Rather, ETS made every effort to get on top of the washback effect of its new test. Washback from tests is not necessarily a bad thing. The driving force behind the massive *TOEFL 2000* undertaking was, ultimately, to design a test that could produce positive washback in the English language classroom. In other words, the power and influence of the TOEFL could potentially be harnessed to shape language learning and pedagogy for the better, particularly with regards to communicative practices in the classroom.

The TOEFL iBT remains ETS's most prominent testing product today. Currently, more than 35 million people have taken the TOEFL, and TOEFL scores are accepted by over 11,500 universities and institutions in more than 160 countries (ETS, n.d.-a; ETS, n.d.-b). The TOEFL has influenced how English is taught around the world, and its importance in admissions decisions lends credence to the dominance of English as a global *lingua franca*. Although the TOEFL does not prominently appear in the historiography of American higher education, the test has become massively important to American colleges and universities, as increasing numbers of international student choose to study in the U.S. The TOEFL is an artifact of American higher education's growth and success in the wake of the "emergence" of the university at the end of the 19th century. The broad appeal of American higher education has attracted students from around the world with diverse linguistic backgrounds. Meeting the English language requirements of America's colleges and universities is both a matter of admissions and eventual college success.

Furthermore, the TOEFL represents a high point in the evolution of psychometric testing technologies in the 20th century. Aided through computers and the internet, the TOEFL is able to deliver a large-scale, standardized assessment to hundreds of thousands of test-takers each year. But unlike the U.S. Army Alpha, the SAT, or even the original paper-based TOEFL, the TOEFL iBT elicits far more information than multiple-choice responses. Rather, thanks to computer-mediated testing, the TOEFL resembles elements of the early clinical intelligence tests in France. But unlike the 19th century French clinicians and psychologists, today's highly trained TOEFL iBT raters can evaluate complex, open-ended responses over the internet (perhaps from the comfort of their own homes), and at an enormous scale. For American colleges and universities,

the TOEFL iBT provides a clearer picture of the English language skills and communicative competence of prospective international students than ever before.

Chapter 5:

English Language Hegemony and American Higher Education

The decades-long project to revise the TOEFL produced a sophisticated product that is able to measure complex constructs like communicative competence. This chapter describes the linguistic context today that has positioned the English language far and above most others. The success and influence of American higher education bears some responsibility in the hegemonic status of English. The TOEFL, too, as a high-stakes instrument of university gatekeeping, elevates the value of the English language and quantifies degrees of attainment. This chapter explains the variety of costs of English language hegemony today and situates American higher education and the TOEFL within this complex dynamic.

American higher education has responded to globalization with both the rhetoric and practice of internationalization, which functions as a kind of coping mechanism for dealing with the context of globalization (Altbach & Knight, 2007). Internationalization can also be understood as a strategy for configuring academic capital to suit the competitive practices of U.S. higher education institutions within a global postsecondary education market. Critics of the commodification of U.S. higher education often draw attention to neoliberal forces that have been driving the global economy and espousing the virtues of free markets, including educational markets (e.g., Rhoads & Szelényi, 2011; Torres, 2009). The overwhelming dominance of U.S. higher education in the global education market is often seen as a cause for alarm, potentially threatening indigenous cultures, values, and ways of knowing (Torres, 2009).

Often excluded, or treated as an aside, in the higher education literature is the role of the English language in shaping the global higher education market. Likewise, applied linguists and

sociolinguists may address the implications of English language policy in the neoliberal economy, but often fail to fully contend with the position higher education plays in perpetuating English language hegemony. English language teaching and learning is not simply an educational goal, but a bought-and-sold commodity, holding a position as both academic and linguistic capital (Phillipson, 2008). Language policies can exist both implicitly and explicitly, but the effects of those policies can have an enormous impact on the higher education market, international student flows, and the landscape of local cultures around the globe. Furthermore, the costs attributed to language learning are not merely economic. Learning a language with *lingua franca* status raises the stakes, and subsequently the stress, for many English language learners around the world. The gatekeeping function of the English language in higher education creates a genuine human cost to individuals and families in many countries.

Higher education institutions, especially those in the U.S., are not merely passive participants in what might be interpreted as the natural adoption of English as a global language. Rather, U.S. higher education has agency in shaping the primacy of English in other countries. Recognizing the importance of that influence and examining the consistency between institutional behavior and institutional mission will help clarify what it means for U.S. colleges and universities to internationalize. The Test of English as a Foreign Language (TOEFL) also plays a role in reifying the prestige and importance of the English language across the world. The TOEFL establishes a quantifiable goal for English attainment, with the potential high-stakes payoff of a world class education at a U.S. college or university. In essence, the TOEFL's broad power to classify English language ability enables it to deliver *bona fide* language credentials, which carry real exchange value in the global market of education, ideas, and commerce.

English Language Hegemony and the Higher Education Literature

Much of the higher education literature addressing globalization and internationalization pays scant attention to the importance or impact of English language education or English language policies beyond the fact that they simply exist. The primacy of English and English language learning is not altogether ignored, but rather serves as a part of a larger collection of trends or patterns signifying internationalization at higher education institutions. Altbach (2014), in his overview of the field of higher education, makes several references to the connections between research production and the English language. He notes that "the main centers in terms of knowledge production and research are in the major English-speaking developed countries" and goes on to enumerate the various areas in the academic world where English dominates (e.g., publishing, international organizations, websites) (Altbach, 2014, pp. 5 & 13). Marginson and van der Wende (2007), in their description of globalization and higher education, repeatedly acknowledge the influence of English-speaking countries in higher education, and even refer to the role of English as the principal language of academic communication to exemplify global transformations that depend on isomorphic change (Marginson & van der Wende, 2007, pp. 13–14). Like Altbach, Marginson and van der Wende offer a wide-ranging list of examples of how English functions in global higher education, both illustrating its importance as a *lingua franca* and normalizing its logical dominance (Marginson & van der Wende, 2007, p. 21). Descriptions of these themes and patterns also appear in Marginson and Rhoades (2002), but the authors also draw attention to the importance of English as a second language (ESL) within the U.S. itself. American colleges and universities are often providers of English education to immigrant populations that serve in the U.S. labor force. Marginson (2006) further explicates the global advantage English holds in the "global hierarchy of nations," noting the particular importance

English commands in scientific research. Similarly, Cantwell and Maldonado-Maldonado (2009, pp. 5 & 11) highlight the fact that "English has become the global language of science" and that "more and more universities go English" as an "ever-increasing imperative" for success. Publication counts in English-language scientific journals are becoming particularly important indicators in determining global university rankings, thus further raising the stakes for having a command of the English language (Cantwell & Maldonado-Maldonado, 2009).

The articles mentioned above provide a small sample of how the topic English as a global language enters the higher education literature, particularly within discussions about internationalization and globalization. The articles address English language as an important part of higher education and recognize the imperative for researchers and institutions across the globe to function in English. However, these articles never extend the discussion about English language hegemony in higher education to include critical analyses of the sociocultural, political, and power implications for global English in academia. This, of course, is no shortcoming of the authors, who never claim that such analyses are to be included within the scope of their respective articles. The point here is to merely highlight the relative absence of critical discussion of English language hegemony within the higher education discourse. Considering that other normative activities emerging from higher education institutions have attracted thoughtful, critical analyses from within the field of higher education, it seems that the significance of English in global higher education has been overlooked, or at least overshadowed. Trends in international higher education, like the prioritization of science publications, the influence of global rankings and league tables, and the increased pursuit of internationalization initiatives (e.g., international branch campuses, international student recruitment), have elicited extensive critical responses, particularly in connection with

neoliberalism and the values of global competition. Yet English as a global language, which is instrumental in all of these activities for much of the postsecondary market, remains peripheral to the organizational agency of higher education institutions.

One of the most revealing critiques ascribed to the role of English in the global economy is the fact that it is treated as a given, as the natural state of affairs (Phillipson, 2008). Like so many other characteristics of higher education today, the dominance of English is treated simply as part of the natural order within the free market (Ricento, 2012). English is blithely regarded as beneficial to efficient exchanges of information and communication, completely neutral to its political and cultural heritage. The neoliberal environment of the present day welcomes a dominant global language to conduct business and expedite the creation and flow of knowledge. The fact that English so happens to be the de facto global language is routinely accepted as a natural consequence of market forces, and therefore a benign and unremarkable circumstance.

The fields of sociolinguistics and applied linguistics, while relatively new in the academic landscape, have offered recent critiques of English language hegemony. Historically much of the focus has been on the protection and preservation of indigenous languages facing extinction or irrelevance in the face of English as a *lingua franca* (Mesthrie & Leap, 2009). Sociolinguists and applied linguists have also examined the concept of competing world Englishes. World Englishes are simply varieties of English that span the globe (e.g., the English spoken in India as a first language), but their status is often dependent on perceptions of the local population and can lead to prejudices tied to language standardization and any perceived shortcomings to those standards (Kachru & Nelson, 1996; Lowenberg, 2000). A final example of the breadth of sociolinguistic research is the study of language planning and policy. This traditionally refers to planning for bilingual education, such as English language learner (ELL)

curriculum in the U.S. K–12 system or bilingual French-English education in the Canadian public school system (Wiley, 1996). These areas of inquiry continue to be rich sources of sociolinguistic research today, but rarely touch upon the role of higher education in perpetuating English language dominance around the world.

However, over the past twenty-five years, a number of sociolinguists and applied linguists have begun examining English language hegemony in light of the global political economy. Inevitably, their research has been intersecting with the organizational and institutional dynamics of higher education. Given that English is the primary focus, specific concerns about higher education in Anglophone countries—especially the U.S.—become paramount. Nevertheless, much in the same way higher education researchers limit the scope of their discussion of English language dominance, sociolinguists and applied linguists rarely extend their critiques of English as a global language into an extended discussion about higher education. Both fields of study, higher education and sociolinguistics, have examined their subjects in relationship to the political economy and neoliberal ideology, but neither have fully resolved the relationship between English language hegemony and international higher education, or at least made that relationship central to their arguments.

The Colonial Legacy of English Language Hegemony

One of the most fascinating questions about English language hegemony and its power implications revolves around the notion of agency in implementing language policy. The apparent lack of overt language policies for perpetuating English language dominance around the world provides some measure of cover for neoliberal apologists who contend that global English is merely a response to free markets, and therefore harbors no malicious intent or ill effect. While this certainly seems to be the case when examining the present place of English in world affairs,

the perspective of global English as a happy accident of market forces is willfully ahistorical. The legacy of English language hegemony is closely tied with the U.S.'s historical strategies for global dominance (Phillipson, 2008). Therefore, an historical understanding of the explicit attitudes toward the English language from British and U.S. leaders and policy makers helps set the stage for the present position English holds in the world.

"There have been blueprints for U.S. dominance of the two American continents since the Monroe Doctrine of 1823 and for global domination for more than a century" (Phillipson, 2008, p. 2). The values and beliefs of American and British leadership have been exemplified in the continued framing of common geopolitical interests expressed through the concept of "English-speaking peoples" (the most famous example being Winston Churchill's four volume *A History of the English-Speaking Peoples*, 1954–56) (cited in Phillipson, 2008). A product of colonial history and ideas of Manifest Destiny, English has been implemented as a common or chosen language for economizing communications among the colonized. However, the language itself needed to be transmitted and taught. This, of course, led to a concerted effort to develop English language teaching as a singular profession. Much of the early training for English language teaching emerged in Great Britain to service the colonies in Asia and Africa. American teachers soon followed suit, many of whom were involved with the Berlitz corporation and the Berlitz language learning approach (Howatt, 2004). The most remarkable aspect of this new profession was that it positioned native-speakers of English as world-travelers, disseminating their language and culture to distant lands (Phillipson, 2008). It may only be a coincidence that this nomadic job description, in part, resembles some of the characteristics of evangelism or missionary work, but as it turns out the two were not always unrelated. Certainly, British and American missionaries brought with them not only the Good News, but also the English

language. Furthermore, British and American officials would occasion to align the spread of English with something resembling a religious mission. “The ultimate aim of many ‘anglicists’… was not enlightenment… but conversion, and a knowledge of English was an essential pre-requisite for Christianization” (Howatt, 2004, p. 146). The burgeoning purpose and need for English language teachers portended the development of an entire branch of educational training, as well as a number of new educational industries (e.g., English language schools, teacher education, English language assessment).

As with many other professions, English language teaching eventually became organized. This meant the development of professional organizations, publications, and conferences. The first international conference aimed at English teaching professionals was The Carnegie Conference in October 1934 (Phillipson, 2008). The agenda at the Carnegie Conference expressed “an explicit intention of spreading English as a world language on a basis of U.K.-U.S. collaboration” (Phillipson, 2008, p. 12). The following year another conference was held in London at the British Colonial Office. Also generously supported by the Carnegie Foundation, the follow-up conference established a yearlong course for training teachers of English as a foreign language. The language pedagogy being developed and discussed in these formative years of the English teaching profession is germane to the topic of global English insofar as it introduces standards for learning English that alienate the mother-tongue and devalue non-native speakers. The Basic English movement, for example, proposed a simplified version of English that would help fast-track foreign language speakers to the rudiments of communication, but ultimately served as a foundation for learning “proper” English (Phillipson, 2008). Ultimately, the formalization of an English teaching profession would produce a native-speaker population of teachers to spread English around the globe.

The aftermath of World War II revealed a completely changed world order. The British Empire was in shambles and hegemonic world power was divided between just two countries. The combination of the Cold War and decolonization had interesting repercussions for English language teaching and the promotion of English language hegemony. In the face of a perceived Soviet menace, the U.S. government directed large sums of private and government funding, including support from the Ford Foundation, to promote English around the world (Phillipson, 2008, p. 17). On the decolonization side of the coin, Great Britain was allowing indigenous languages to attain official-language status in its former colonies, but often in tandem with English as an official language. This often left former colonies with at least two official languages (when English was not singled out as the sole official language) and maintained English language education as a requirement in the public school systems (Howatt, 2004).

In order to help facilitate the decolonization processes pertinent to English language education, an international conference was held in Uganda in 1961. Although many of the leading British experts in English language teaching were unable to attend, the conference set certain precedents and agendas that would have lasting effects for English teaching and English language hegemony (Howatt, 2004). Five tenets emerged from the conference, but three are of particular importance to the direction of English as a global language. The first tenet recommended monolingual instruction (Howatt, 2004, p. 312). In other words, when English is taught it should be taught using *only* English and not the students' first language. While recommended on pedagogical grounds, this first tenet also implies that English supersedes other languages, specifically the native tongue. The second tenet was, "the ideal teacher of English is a native speaker" (Howatt, 2004, p. 314). The primacy of the native-speaker suggests that indigenous teachers of English are somehow inferior, perpetuating the notion that there is only a

single, correct English. Finally, the third tenet of note (in actuality, the fifth of five) suggests that students' English language ability will automatically decline if they use other languages besides English (Howatt, 2004, p. 315). This had implications for English as a medium of instruction. According to this last tenet, not only should the English language be used to teach English, but it should also permeate other avenues of life, including the instructional medium for other subjects.

The *othering* effect of English language pedagogy emerges, in part, from the concepts of native-speaker status, monolingual instruction, and the implied superiority of English as the medium of instruction. Further, the most common acronym to describe the English teaching profession (and the name of its largest professional organization), TESOL, literally distinguishes English language speakers from *other* language speakers: Teachers of English to Speakers of Other Languages (Phillipson, 2008). Any language besides English is *other*, as are those who speak other languages.

The legacy of English language policies and teaching practices from the days of empire continues today. Although the overt paternalism expressed in the days of colonialism have subsided, the practices and policies that promoted English in the colonies have succeeded in securing English as a global language *par excellence* (Ricento, 2015b). English, as a *lingua franca*, is perpetuated today not as part of U.S. or British foreign policy, but by its linguistic capital (Phillipson, 2008). This capital is manifested in what Bourdieu (1992) describes as the symbolic power of language. Those who are subjugated by the language, or the powers which the language represent, are ultimately complicit in their own weakened position. The importance of English in commerce, research publications, postsecondary education, and pop culture and entertainment reifies the belief that English is essential and beneficial (Phillipson, 2008). This positioning of linguistic capital is not simply restricted to developing or postcolonial countries,

but also includes major economies in Europe and Asia. The E.U., Japan, and South Korea are no less complicit in the symbolic empowering of English than India, China, or Ghana. Autonomous nation-states set their own policies on foreign language education, not the U.S. or the U.K. Nevertheless, the consensus overwhelmingly favors compulsory English education as part of national policy in many countries, and individuals and families within those countries purchase English education products and services to position themselves or their children for competitive advantage (Ricento, 2015a).

The competition associated with English language acquisition fits neatly within neoliberal ideology. The neoliberal economy relies on so-called free markets to ensure a *survival of the fittest* competition of goods, services, organizations, and even languages. Therefore, the dominance of English is seen as a natural and necessary state of affairs for a healthy, efficient economy. A counterpoint to this neoliberal outlook similarly ameliorates English as a *lingua franca*, but as a means of communication necessary for strengthening global networks and institutions that could serve to advance global interests of economic justice, environmental sustainability, and the reduction of conflict (Van Parijs, 2000). Nevertheless, Bale (2015, p. 90) recognizes that "neoliberalism has coopted the language of social justice to frame its agenda," creating doubt about claims that simultaneously support global English on the basis of both neoliberal ideology and left-leaning social values. Thus, the resulting debate on the good or ill of global English often pits progressive academics against each other without a clear resolution in sight.

Regardless of the merits or weaknesses of English as a *lingua franca*, there is little doubt about the linguistic capital English carries in the world today. Attempting to disentangle the wrongs and injustices associated with the historical legacy of global English may ultimately

prove fruitless. Presently, the linguistic capital and symbolic power of English appear to be uncontested, resulting in an English hegemony that is more a matter of fact than a matter of debate. Understanding the role English plays in the political economy is therefore of particular importance.

The Costs of English Language Learning

English language hegemony may be uncontested, but it has real implications in the global economy. Many of the present problems and questions associated with English as a global language stem from both the human and financial costs to learning English. English language education is a huge industry. Textbook publishing, assessment tools, and language schools penetrate an international market that is hungry to consume English. Native-speakers from Anglophone countries are hired to teach English all over the world. Depending on the hiring source and country of destination, the required credentials for English teachers vary greatly. In many cases the hired teacher has virtually no teaching experience or training whatsoever. Many consumers of English language education approach their purchases with uncritical eyes, attracted to flash over substance and promises of quick and easy results, with little appreciation for the actual work involved to acquire a foreign language. The financial costs to individuals and governments can be enormous, but there are also human costs arising from the way English language education affects individuals' wellbeing, time, and personal goals.

Countries in East Asia, like Japan, South Korea, and China, spend billions of dollars every year on English education. In addition, these countries incur substantial human costs to studying English. In order to compete in an international market, major Japanese companies like Rakuten and Uniqlo are enforcing English-only operation policies (Botting, 2010). Employees at these businesses are expected to know English or learn English for a number of key

communication activities. The assumption in these cases is that internationalization and international business share an equivalency with English. Benchmarks for English ability are measured with costly foreign-made assessment tools, like the TOEFL or the Test of English for International Communication (TOEIC) from ETS, and English language courses for employees are encouraged and often subsidized by their companies. However, the expectations and accountability for using English creates additional human costs in the form of work-related stress and impositions on one's personal time.

Piller and Cho (2013) present an excellent example of the human costs of English language learning in South Korea's higher education system. Like other parts of East Asia, South Korea spends an exceptional amount of money on English education. In 2009, the private market alone for English language learning in South Korean was KRW 1.5 trillion (approximately USD 1.1 billion), which accounted for about 40% of the public education budget (Piller & Cho, 2013, p. 29). Recent higher education reforms in South Korea have placed greater emphasis on English language acquisition, and some institutions like the Korea Advanced Institute of Technology (KAIST) transitioned their curriculum into English as the sole medium of instruction. However, during a four-month period in 2011, the South Korean public were captivated by the stories of four KAIST students and one KAIST professor who committed suicide. Although, the circumstances surrounding each suicide may have been particular and varied, the South Korean media framed the stories as a symptom of the undo pressures at KAIST from its English requirements and curriculum.

Pillar and Cho (2013) are careful not to claim that the strict English medium-of-instruction environment was directly to blame for the suicides, but instead merely recognize that the pressures to learn, teach, and publish in English are a substantial burden for those who speak

English as a second language. Like the example of Japanese businesses, the South Korean education system makes the connection between internationalization and English. English language curriculum and foreign student recruitment are tied to indicators of "internationalization" in international university rankings and league tables, and therefore quite easy and expedient to manipulate. Even the internal rankings in South Korean media sources favor English research outlets for measurement. Scraping for any modicum of positional advantage, many universities intensify their English curriculum to improve their international profile (Piller & Cho, 2013, p. 25). Ultimately, the authors argue that the competition inherent in neoliberal exchanges of academic capital functions as "a covert form of language policy" (Piller & Cho, 2013, p. 24). The resulting costs to a nation and its higher education system are not simply measured in currency, but in human costs.

The financial and human costs of English education around the globe are substantial, but it is also important consider what is not spent—or, at least, which countries are not spending. In a globally competitive higher education market that favors English, there are obvious advantages for the U.S., the U.K., and other Anglophone countries. While much of the world is spending large portions of their educational budgets on English language learning, the U.S. and the U.K. are not. François Grin (2005, cited in Phillipson, 2008) undertook a study to understand the implications of English education in Europe. His calculations include "quantifiable privileged market effects, communication savings effects, language learning savings effects… alternative human capital investment effects… and legitimacy and rhetorical effects" (Phillipson, 2008, p. 28). Grin concluded that continental European countries transfer at least €10 billion per year (but more likely about €16–17 billion) to the U.K. and Ireland. This figure well exceeds the annual €5 billion British E.U. budget rebate, revealing a large disparity in market fairness. In addition, Grin

found that the U.S. “saves $19 billion [per year] by not needing to spend time and effort in formal schooling on learning foreign languages” (Phillipson, 2008, p. 28). English as a global language provides advantages to the U.S. and U.K. that extend beyond its simple utility. There are real savings in both monetary and human costs by not having to tie resources to large-scale English language education.

High stakes English language assessments, like the TOEFL, also elevate the market value of English. ETS has worked diligently, and at great expense, to develop the modern-day TOEFL iBT to measure English language proficiency and communicative competence. Test scores provide a quantifiable credential of English language proficiency, which is recognized at thousands of institutions of higher education—making the scores both valuable and portable. However, the advancements in the testing technologies that brought the TOEFL iBT into being, including the use of computer software, artificial intelligence, and the internet, have also escalated the fees associated with the test. Like other instruments of English language hegemony, the TOEFL impresses the importance and value of the English language while exacting real-world costs on English language learners. Certainly, the cost come in the form of test preparation courses, tutoring, and published preparation materials, but the costs of taking the test have also increased. Presently, the fee to take the TOEFL in the U.S. is $185. Naturally, this fee does not include related costs to getting to a test center or the opportunity costs of spending an afternoon taking the examination. But in this way, too, the TOEFL solidifies the status and value of English while placing a tax on the English language learner.

The topics of globalization and internationalization are drawing a great deal of focus and attention from the fields of higher education, sociolinguistics, and applied linguistics. Much of the concern has revolved around the implications of neoliberal ideology and the movement of

both academic and linguistic capital. Higher education researchers have addressed the role of English in international postsecondary education, and sociolinguists have alluded to the position higher education holds in English language learning. The history of English language policies imposed by empires or superpowers reveals agency in the position English maintains today. However, the primacy of English as a global language endures, in part, because of so-called free market forces that advantage the language of America and Britain. Neoliberal ideology provides sanctuary for English as a natural and inevitable outcome of rational markets. The importance of English in global communication, scientific research, and the knowledge economy is virtually impossible to dispute. However, the inequities created or implied by English language hegemony are important to consider.

For many universities around the world, English stands as a synonym for internationalization. The fact that English is the language of the U.S. gives ammunition to those who see globalization as a euphemism for *Americanization* (Phillipson, 2008, p. 26). Concerns about English as a global language include financial and human costs for English education, but also include questions about the impact English has on local cultures, minority languages, and economic competition. While some may see English language hegemony as destructive and threatening, there are also those who recognize the status English holds as a positional good and express an interest in improving people's access to English education (Ricento, 2012).

Research in international higher education would benefit greatly from closer inspection of English language hegemony. English plays an integral role in the global higher education market, lending linguistic capital to the sciences and providing a defining characteristic of global research universities. Higher education researchers may not wish to embroil themselves in

questions about language equity and language extinction, but they would do well consider the implications of English and the political economy on postsecondary education.

In the global economy, English is privileged. Higher education in the U.S. is still the envy of the world, but some of that veneer is beginning to rub off (Bok, 2013). Much of the edge U.S. universities hold (e.g., in research and development, scientific publication output, international student flows) stems from the role of English as a *lingua franca*. The relationship between U.S. higher education and the political economy is complex and often controversial. English language hegemony is neither natural nor neutral, but instead privileges U.S. colleges and universities in the sphere of global competition. The dominance of English offers an important avenue for exploring U.S. higher education, forging linguistic and academic capital into instruments of geopolitical and economic power. Examining the intricacies of English language hegemony may reveal new insights into the role U.S. higher education plays in shaping language policies, managing linguistic capital, and competing with emerging international competitors.

Chapter 6:

Conclusion

The story of the English language within the historiography of American higher education is largely untold and unacknowledged. Quietly residing in peripheral passages and oblique asides, the story reveals the important role English has played in the evolution of the university curriculum, the "emergence" of the university at the end of the 19th century, and the international prominence of American research universities today. The outsized success of America's colleges and universities is also intricately tied to the economic and geopolitical fortunes of the United States. Situated between the histories of a prestige system of higher education and a nation recognized as a global "superpower" rests a language that has assumed hegemonic status around the world. The ubiquity of the English language extends well beyond the borders of the United States and other Anglophone countries, penetrating nearly every avenue of knowledge production and communication. In a globalized world, English routinely dominates the discourses of science, technology, industry, trade, and diplomacy.

However, higher education in the United States did not begin with a curriculum in English. At the time of Harvard College's founding in 1636, the languages of study, institutional rituals, and campus conversations were Latin and, to a lesser extent, Greek. Gradually, the utility and convenience of English would wear away at the institutional rules for using Latin and Greek, transforming the felicity of campus rituals, instruction, and conversation. By the time of the American Revolution, the vestiges of Latin and Greek remained pronounced in American higher education, but these ancient languages were now confined solely to the curriculum (Rudolph,

1977, pp. 36–37). As a matter of mental discipline and piety, Latin and Greek were to be studied, but not necessarily used.

The English language presented a slippery slope for the college curriculum. Naturally, the usefulness of the English language lent itself to the study of more useful things. And before long, students, professors, and upstart colleges and universities were gravitating toward curricula that favored *utility* over the traditional *discipline and piety* mindset. In the 1828 Yale Report, the guardians of tradition, discipline, and piety, squared their collective shoulders to press the old curriculum back into its rightful position atop the aforementioned slope. But the Yale Report's victory for *discipline and piety*—and Latin and Greek—would be short-lived. By the end of the 19th century, new universities (e.g., Chicago, Johns Hopkins, Stanford) and visionary leadership (e.g., Charles Eliot, Daniel Coit Gilman, William Rainey Harper) would help transform the curriculum and the idea of the American university itself. The American university, as we recognize it today, would "emerge" into the 20th century with an emphasis on *utility* and *research*, leaving the adherents of *discipline and piety* to play Sisyphus in their labors to restore curricular tradition.

In the wake of the "emergence," American higher education was well positioned for the rapid technological advancements and cataclysmic disruptions in world affairs that characterized the first half of the 20th century. Early developments in intelligence testing by American psychologists, coupled with the exigencies of World War I, transformed the possibilities of large-scale, standardized testing. America's college and universities capitalized on the techniques of large-scale testing to generate new admissions tools for efficiently evaluating a growing number of prospective high school students. America's nascent system of secondary schooling initially lacked curricular coordination with institutions of higher education. Thus, determining

college readiness and aptitude was not immediately clear from school reports and records alone. The College Board, formed principally by the established universities in the East, created the Scholastic Aptitude Test (SAT) in 1926 to provide a common assessment instrument available for all schools to use. Prior to the SAT, colleges and universities administered their own institution-specific battery of admissions tests.

The SAT gained traction in the 1930s when Harvard University president, James Conant, utilized the test for making award decisions for the new National Scholarships at Harvard. The success of the SAT administration to large numbers of prospective students, as well as the prestige factor of Harvard University, helped elevate the SAT nationally for college admissions. By the time the U.S. entered World War II in 1941, the SAT had supplanted most institutional admissions tests.

Those responsible for the SAT's success, including James Conant and Henry Chauncey (dean at Harvard), contributed to large-scale test development projects for the U.S. military during World War II and would go on to help found America's premier independent testing organization, Educational Testing Service (ETS). Eventually, ETS would be responsible for developing and administering the SAT and the Graduate Record Exam (GRE).

The creation of ETS coincided perfectly with a growing need for large-scale admissions testing after World War II. The GI Bill of 1944 created new and far-reaching opportunities for American veterans returning from the war, which contributed to a sustained period of enrollment growth at America's colleges and universities. The gradual change in student demographics, influenced by the surge of veteran students from all walks of life, helped alter the public's perceptions about college access. College not only appeared more attainable, but there were indications that it could be immensely useful, too.

The growth in student enrollments after World War II signaled a "golden age" for American higher education. Students from around the world were also taking note of the American success story and the increasing prominence and prestige of America's higher education system. As foreign student enrollment rapidly increased in the 1950s, the U.S. government, institutions of higher learning, and organizations dedicated to international exchange began contemplating avenues for efficiently and effectively gauging the English language proficiency of prospective international students. In 1961, interested parties and stakeholders met in Washington, DC to discuss the development of a large-scale, standardized English language assessment that could be administered to pre-college, undergraduate, and graduate students whose first language was not English. As a result, the Test of English as a Foreign Language (TOEFL) was born.

First administered in 1964, the TOEFL has since become the most important test in ETS's product line. The original TOEFL utilized many of the standardized testing techniques developed in wartime, including security features and multiple-choice responses. Over time, however, it was becoming clear that the TOEFL was not doing a particularly good job evaluating communicative competencies. By the 1990s, ETS had put TOEFL on a path to redevelopment. The new TOEFL would take into account the standards of communicative competence and introduce computers and the internet for administering the exam. Delayed by several years, the new TOEFL would debut in 2005 under the name, TOEFL iBT (internet-Based Test). International students taking the TOEFL iBT would now need to produce writing and speech samples, responding to lectures, conversations, and writings that simulated academic content and contexts.

The TOEFL has become a powerful gatekeeping tool for American higher education and a globally famous product for ETS. The TOEFL has been taken by more than 35 million people and is recognized by more than 11,500 institutions in 160 countries (ETS, n.d.-a; ETS, n.d.-b). It is also the most cheated-on test at ETS (Lemann, 1999, pp. 241–242). The stakes for doing well on the TOEFL are extremely high for international students who wish to study in the U.S.

In many ways, the TOEFL has helped reinforce the hegemonic status of the English language around the world. The linguistic capital of English is difficult to refute. English dominates the discourse of science, industry, and statesmanship, and English is a key marker of American higher education. Prospective students from abroad understand that not only do they need English to study in the U.S., but also their proficiency in English will immediately transfer to their future careers.

The linguistic capital of English may seem like an abstraction. After all, what is English really worth? And how much English does one need for it to be worth something? It's hard to put a figure on it. But the TOEFL adds a clearer sense of exchange value to the dynamic. TOEFL scores are actually worth something and, if high enough, they can facilitate admission into a world class university in the United States. TOEFL scores are quantifiable, and prospective students can interpret their scores to estimate their own level of English proficiency. For many international learners of English, the TOEFL provides an early opportunity to see, in numeric form, how much English they "have." Thus, the TOEFL helps to establish a more tangible, quantifiable notion of English's worth, reifying the linguistic capital of the English language itself.

The TOEFL also comes with a price tag. Today, the fee for taking the TOEFL iBT in the United States is $185. Tests are expensive to develop, and complex computer-based tests—with

a network of paid raters and evaluators—are expensive to administer and score. Advancements in testing technologies have come with steadily increasing costs to test-takers. For international students from wealthier countries, the increased costs might be perceived as nominal. But for other prospective students from developing parts of the world, the costs of taking the TOEFL might be prohibitive. The prestige quality of the English language, particularly when pursued for the purposes of American higher education, may prove too costly for many eager and capable students around the world.

English language hegemony has fascinating implications for the global exchange of ideas, goods, and services. Through its own successes and acclaim, America's system of higher education has contributed to the entrenchment of the English language in international communications. The evaluative processes for admitting international students to U.S. colleges and universities raise the stakes for admissions while communicating selectivity and scarcity to students. For students whose first language is not English, the TOEFL can serve as either a barrier to admission or a metric of linguistic accomplishment. Participation in America's higher education system, much like many aspects of today's global economy, requires the linguistic capital contained in the English language.

Looking Forward

The history of American higher education, like all great stories of human experience and endeavor, is replete with lessons, values, themes, and drama. As storytellers, historians are skilled at illuminating important and meaningful moments in time and drawing careful and reasoned conclusions about the course of events. Historians may not always agree with each other's interpretations about the past, but they generally agree not to take pains to predict the future. The study of history, as a discipline, is averse to soothsaying and speculation.

Nevertheless, the story of American higher education is filled with disruptions and change. David Labaree (2017) describes this culmination of wrong turns and happy accidents as a "perfect mess," a seemingly haphazard series of events and circumstances that somehow birthed a most remarkable system of higher learning. Scholars of American higher education, unlike historians, are too often piqued by the past to resist the urge to ascribe meaning, cast warnings, or give advice, usually with some imagined future in mind. In Labaree's estimation, for example, the future of American higher education would do best if left to its own devices, as history has shown. The most successful policies for higher education tend to "feed" the system, rather than try to "reform" it. For Labaree (p. 196), the future health of American higher education depends on policies that offer funding and resources without simultaneously taxing the system with reform-minded regulations, rules, and requirements.

More recently, Arthur Levine and Scott Van Pelt (2021), attempt to make predictions and recommendations about American higher education's "uncertain" future by examining its history alongside current trends, developments, and disruptions. Levine and Van Pelt take inspiration from recent technological and cultural changes in mass communication to envision a higher education future that hardly resembles its historical antecedents. They see a future where educational content is delivered through multiple modes of communication over the internet and is produced and disseminated by a much wider variety of higher education providers. Certainly, the old institutions will persist, but Levine and Van Pelt also imagine new enterprises emerging to compete with the traditional colleges and universities.

Salient to the future of higher education assessment, Levine and Van Pelt insist that testing and accountability will become increasingly important. As the "student-as-customer" metaphor takes deeper root in the business models of an evolving higher education industry,

Levine and Van Pelt predict a greater emphasis on proven signals of competency and demonstrable student outcomes. The higher education assessment industry is poised to flourish in a future where knowledge attainment needs to be quantifiable and objectively interpreted.

Perhaps assessment companies like ETS can take some measure of comfort in Levine and Van Pelt's prediction. Recent disruptions in higher education due to the COVID-19 pandemic have actually posed a possible existential threat to the SAT and ACT admissions exams. In light of lockdowns, quarantines, and other disruptions at America's high schools in 2020, U.S. colleges and universities waived the standardized test requirements for the Class of 2024. The test-optional approach has continued into 2021, with some colleges (e.g., Harvard) waiving admissions test until at least 2024. However, some schools (e.g., MIT) are already reversing course and requiring admissions exams for the next enrollment cycle (Anderson, 2022).

The most impressive lesson from this real-life experiment in college admissions, sans SAT and ACT, was that America's colleges and universities were perfectly capable of assembling the Class of 2024 without test scores. The question remains whether U.S. colleges and universities will continue on a course of test-optional admissions, acknowledging the limitations of the SAT and ACT, or whether they will revert back to standardized testing.

Unlike the SAT and ACT, the future of the TOEFL is probably much clearer. Regardless of academic readiness, America's undergraduate and graduate programs will continue to need some measure of international students' English language proficiency before they depart for the U.S. Furthermore, the hegemonic status of the English language will continue to infuse English with enormous linguistic capital. English is the language of science, both spoken at international conferences and written in journals. In elite journals of the natural sciences, regardless of country

of origin, over 98% of articles are written in English (Gordan, 2015, pp. 294 & 302). The importance of English as a global language will likely continue for the foreseeable future.

The world has seen a collision of seemingly disparate disruptions in recent years: the immediate and visceral effects of climate change, the vocal rise of white nationalism and anti-immigration sentiment in the United States, stirrings of and sympathies for authoritarianism around the world, a years-long global pandemic, and a return to lawless belligerence by a nation-state, as seen in Russia's senseless and brutal invasion of Ukraine. Understandably, the flows of international students to U.S. colleges and universities have been tempered as a result of some of these global calamities. But predicting the future of international student participation in American higher education, as with the higher education enterprise itself, is a nearly impossible task.

Perhaps the most likely course, the one David Labaree suggests, is that U.S. higher education will continue to persist. The messiness of American higher education is a virtue, lending colleges and universities a greater degree of flexibility and creativity than most other conventional enterprises. Of course, many of those qualities need to be preserved and protected. Labaree (2017) notes that American higher education's institutional autonomy, sensitivity to consumers, eclectic collection of constituencies, innate ambiguity, and organizational complexity are defining traits that allow it to adapt to change and disruption. These qualities may, in fact, help preserve the American system of higher education. If Levine and Van Pelt (2021) are correct, however, the future of American higher education may look very different than the one we recognize today, the one we have come to know since the time of the "emergence." But should American higher education eventually assume a transformed and unfamiliar guise at some future date, I would predict—or, more rightly, hope—that its underlying DNA will remain much

the same. American higher education will continue to be messy and aspire to something more perfect.

REFERENCES

REFERENCES

Altbach, P. G. (2014). The emergence of a field: Research and training in higher education. *Studies in Higher Education*, *39*(8), 1306–1320.

Altbach, P. G., & Knight, J. (2007). The internationalization of higher education: Motives and realities. *Journal of Studies in International Education*, *11*(3–4), 290–305.

Anderson, N. (2022, March 29). MIT resumes mandate for SAT or ACT scores. Many other colleges have not. *The Washington Post*.

Bale, J. (2015). Language policy and global political economy. In T. Ricento (Ed.), *Language policy and political economy: English in a global context* (pp. 72–96). Oxford University Press.

Bevis, T. B. (2016). *Higher education exchange between American and the Middle East through the twentieth century*. Pelgrave Macmillan.

Bevis, T. B. (2019). *A world history of higher education exchange: The legacy of American scholarship*. Pelgrave Macmillan.

Bevis, T. B. & Lucas, C. J. (2007). *International students in American colleges and universities: A history*. Pelgrave Macmillan.

Bok, D. (2013). *Higher education in America*. Princeton University Press.

Botting, G. (2010, August 22). Rakuten's English-only policy endures close media scrutiny. *The Japan Times*. http://www.japantimes.co.jp/news/2010/08/22/national/media-national/rakutens-english-only-policy-endures-close-media-scrutiny/#.VTVH5GRVhHw

Bourdieu, P. (1992). *Language and symbolic power*. Polity Press.

Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment*. McGraw-Hill ESL/ELT.

Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, *1*(1), 1–47.

Cantwell, B., & Maldonado-Maldonado, A. (2009). Four stories: Confronting contemporary ideas about globalisation and internationalisation in higher education. *Globalisation, Societies, and Education*, *7*(3), 289–306.

Carson, J. (2007). *The measure of merit: Talents, intelligence, and inequality in the French and American republics, 1750–1940*. Princeton University Press.

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (Eds.) (2011). *Building a validity argument of the Test of English as a Foreign Language*. Routledge.

ETS (n.d.-a). *About the TOEFL iBT*. Retrieved February 10, 2022, from https://www.ets.org/toefl/test-takers/ibt/about/

ETS. (n.d.-b). *Why choose the TOEFL iBT Test?*. Retrieved February 10, 2022, from https://www.ets.org/toefl/test-takers/ibt/why/

Geiger, R. L. (2015). *The history of American higher education: Learning and culture from the founding to World War II*. Princeton University Press.

Gordin, M. D. (2015). *Scientific Babel: How science was done before and after global English*. University of Chicago Press.

Herbst, J. (1989). *And sadly teach: Teacher education and the professionalization of American culture*. The University of Wisconsin Press.

Howatt, A. P. R. (2004). *A history of English language teaching* (2nd ed.). Oxford University Press.

Institute of International Education. (2021a). International student and U.S. higher education enrollment, 1948/49–2020/21 [Data set]. *Open Doors Report on International Educational Exchange.* Retrieved from https://opendoorsdata.org/

Institute of International Education. (2021b). "International students by place of origin, selected years, 1949/50–2020/21 [Data set]. *Open Doors Report on International Educational Exchange.* Retrieved from https://opendoorsdata.org/

Kachru, B. B., & Nelson, C. L. (1996). World Englishes. In S. L. McKay & N. H. Hornberger (Eds.), *Sociolinguistics and language teaching* (pp. 71–102). Cambridge University Press.

Labaree, D. F. (2017). *A perfect mess: The unlikely Ascendancy of American higher education*. University of Chicago Press.

Lemann, N. (1999). *The big test: The secret history of the American meritocracy*. Farrar, Straus and Giroux.

Levine, A., & Van Pelt, S. (2021). *The great upheaval: Higher education's past, present, and uncertain future*. Johns Hopkins University Press.

Lowenberg, P. H. (2000). Non-native varieties and the sociopolitics of English proficiency assessment. In J. K. Hall & W. G. Eggington (Eds.), *The sociopolitics of English language teaching*. Multilingual Matters.

Marginson, S. (2006). Dynamics of national and global competition in higher education. *Higher Education*, *52*, 1–39.

Marginson, S., & Rhoades, G. (2002). Beyond national states, markets, and systems of higher education: A glonacal agency heuristic. *Higher Education*, *43*, 281–309.

Marginson, S., & van der Wende, M. (2007). *Globalisation and higher education*. (OECD Education Working Papers, No. 8). OECD Publishing. https://www.oecd-ilibrary.org/docserver/173831738240.pdf?expires=1650316940&id=id&accname=guest&checksum=845D4B683DC8FEF37235D7CE8778DC6D

McCullough, D. (2011). *The greater journey: Americans in Paris*. Simon & Schuster.

Mesthrie, R., & Leap, W. L. (2009). Maintenance, shift, and death. In R. Mesthrie, J. Swann, A. Deumert, & W. L. Leap (Eds.), *Introducing sociolinguistics* (pp. 242–270). Edinburgh University Press.

Pearlman, M. (2011). Finalizing the test blueprint. In C.A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 227–258). Routledge.

Phillipson, R. (2008). The linguistic imperialism of neoliberal empire. *Critical Inquiry in Language Studies*, *5*(1), 1–43.

Piller, I., & Cho, J. (2013). Neoliberalism as language policy. *Language in Society*, *42*, 23–44.

Rhoads, R. A., & Szelényi, K. (2011). *Globalization, citizenship, and the university: Advancing social life and relations in an interdependent world*. Stanford University Press.

Ricento, T. (2012). Political economy and English as a 'global' language. *Critical Multilingualism Studies*, *1*(1), 31–56.

Ricento, T. (2015a). "English," the global lingua franca. In T. Ricento (Ed.), *Language policy and political economy: English in a global context* (pp. 276–304). Oxford University Press.

Ricento, T. (2015b). Introduction: Language policy and political economy. In T. Ricento (Ed.), *Language policy and political economy: English in a global context* (pp. 276–304). Oxford University Press.

Rudolph, F. (1977). *Curriculum: A history of the American undergraduate course of study since 1636*. Jossey-Bass Publishers.

Rudolph, F. (1990). *The American college and university: A history*. The University of Georgia Press. (Original work published 1962)

Schudson, M. S. (1972). Organizing the 'meritocracy': A history of the College Entrance Examination Board. *Harvard Educational Review*, *42*(1), 34–69.

Smith, J. E. (2012). *Eisenhower: In war and peace*. Random House.

Spolsky, B. (1990). The prehistory of the TOEFL. *Language Testing*, *7*(1), 98–118.

Spolsky, B. (1993, August 2–8). *Testing the English of foreign students in 1930* [Paper presentation]. The 15th Annual Language Testing Research Colloquium, Cambridge, England, UK.

Taylor, C. A. & Angelis, P. (2008). The evolution of the TOEFL. In C.A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 27–54). Routledge.

Thelin, J. R. (1990). Rudolph rediscovered. In F. Rudolph, *The American college and university: A history* (pp. ix–xxiii). The University of Georgia Press.

Thelin, J. R. (2011). *A history of American higher education* (2nd ed.). The Johns Hopkins University Press.

Torres, C. A. (2009). *Education and neoliberal globalization*. Routledge.

Van Parijs, P. (2000). The ground floor of the world: On the socio-economic consequences of linguistic globalization. *International Political Science Review*, *21*(2), 217–233.

Veysey, L. R. (1965). *The emergence of the American university*. The University of Chicago Press.

Wang, L, Eignor, & Enright, M. K. (2011). A final analysis. In C.A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 259–318). Routledge.

Wiley, T. G. (1996). Language planning and policy. In S. L. McKay & N. H. Hornberger (Eds.), *Sociolinguistics and language teaching* (pp. 71–102). Cambridge University Press.