PREDICTING ACADEMIC SUCCESS IN A JAPANESE INTERNATIONAL UNIVERSITY

A Dissertation
Submitted
to the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree of
Doctor of Education

By
Kristy King Takagi
May, 2011

Examining Committee Members

Steven Ross, Advisory Chair, University of Maryland
David Beglar, CITE/TESOL
Paul Nation, External Member, University of Wellington
James Sick, External Member, J. F. Oberlin University
Edward Schaefer, External Member, Ochanomizu University

ABSTRACT

The purpose of this study was to determine which types of student application information, as well as demographic information obtained through a questionnaire after matriculation, best predicted later academic performance in an international English-medium university in Japan, and to examine the "big picture" of how cognitive and non-cognitive variables interact over time in accounting for student success in an English for Academic Purposes (EAP) program and in the regular university program. The study was divided into three parts that separately examined student application information, university entrance examinations, and the larger picture of student success.

In the first part of the study, a hierarchical multiple regression was employed to determine the extent to which a variety of variables derived from application information predicted grade point average (GPA) in the EAP program, as well as first-year GPA and final GPA in the regular university program. The independent variables examined in the main regression analysis were: high school grade point average (HSGPA); ITP TOEFL scores obtained in April of the students' first year at the university; and *hensachi* rankings of the students' high schools. Results indicated that HSGPA was a consistently significant predictor of all levels of university GPA. ITP TOEFL scores significantly predicted EAP GPA, and *hensachi* rankings were significant predictors of first-year GPA and final GPA. In pullout studies, additional variables were examined, including high school grade factor scores, ITP TOEFL section scores, and non-cognitive variables derived from student responses to a questionnaire (e.g., gender and parents' education). Results indicated that high school grade factor scores and ITP TOEFL section scores varied in the degree to which they predicted GPA, female gender predicted HSGPA and EAP GPA, and the father's education predicted HSGPA (though the correlation was negative).

In the second part of the study, university entrance examinations were examined to determine the extent to which they predicted EAP GPA and first-year GPA. The Center Examination section scores, four types of university entrance examinations, and numerical scores from the examinations (e.g., English test, Japanese test, English essay, and Japanese and English interview) were examined in further hierarchical multiple regression analyses to determine how valuable each type and measure was for predicting university success. Results indicated that the Center Examination English test scores and the university entrance examination English essay scores significantly predicted EAP GPA, and that the Center Examination Math test scores and the university entrance examination English test scores were significant predictors of first-year GPA.

In the third part of the study, the larger picture of student success was examined. Logistic regression was first employed in order to determine to what degree HSGPA, high school grade factor scores, ITP TOEFL scores, *hensachi* rankings, gender, and parents' education predicted timely and exemplary completion of program requirements (e.g., finishing the EAP program on time (FOT), graduating on time (GOT), and graduating with honors (GWH)). Results indicated that HSGPA, the Numerical Ability factor scores, ITP TOEFL scores, and the father's education were significant predictors of FOT. Although only the ITP TOEFL scores were significant predictors of GOT, HSGPA, Language Ability factor scores, *hensachi* rankings, and ITP TOEFL scores significantly predicted GWH. Path analysis was then used to examine the path of success, from before matriculation, through the first year of university, and on to graduation. Well-fitting models were produced for both first-year GPA and final GPA in which all levels of GPA were stable predictors of later GPAs, and *hensachi* rankings and ITP TOEFL scores contributed most to the model up to the first-year GPA.

ACKNOWLEDGMENTS

I would like to thank Dr. Steven Ross and Dr. David Beglar for their support in overcoming the challenges encountered in writing this paper. I also thank the members of my committee for their careful reading of my paper and valuable advice, Dr. J. D. Brown, for his support early in the project, and Dr. Christine Pearson Casanave, for her kind and generous encouragement.

This project was made possible by the progressive leadership of Dr. Mineo Nakajima, as well as the moral support and assistance of Dr. Al Lehner; I am greatly indebted to both of them. I also thank the participants and all my colleagues, particularly those in the Admissions Office, Office of Student Records, and EAP Department.

Finally, I would like to thank precious friends and family for their unending help and patience. I especially thank my mother, Bonnie Reynolds King, for inspiring me always with her joy in learning and teaching, my father, Dr. Jules Kent King, and my husband, Roo Takagi, for all they have done for me through the years of work on this project, and throughout my life.

DEDICATED TO ROO TAKAGI,

WITH LOVE AND APPRECIATION

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The Background of the Issue

Selecting students who are most likely to succeed in a university program is a complicated and difficult task, but one of great importance to administrators, teachers, and students; certainly, it is to the benefit of all concerned when students who are accepted succeed academically during their years at university and then graduate in a timely fashion.

In Japan, students applying to universities submit application forms with a variety of information, but are still typically selected primarily on the basis of their performance on that university's entrance examination. As Brown (1995) explained, both high schools and universities make their own entrance examinations, and students must perform well on the examinations in order to be allowed entry to the schools. In order to prepare for these high-stakes examinations, students traditionally have had to endure a stressful period of study that is known as "examination hell" or *shiken jugoku* (p. 21). This intense period of study has been a part of university preparation since the beginning of the 20th century, according to Mori (2002). However, she also stated that this situation is changing due to the declining number of university applicants, particularly for students applying to enter smaller, newer, or less prestigious universities. While such schools have had to lower standards to maintain enrollment levels, the same is not true for more prestigious universities. Entering highly ranked schools is still competitive (p. 30) because, as Ono (2001) pointed out, the prestige of the college that a student enters in large part determines future job possibilities. Even if the number of students applying to universities is shrinking in Japan, it is likely that competition to enter high ranking universities will continue to be intense.

Despite the fact that competition to enter the most prestigious universities continues unchanged, some researchers believe that entrance examinations may undergo further change in response to changing conditions in Japan. Brown, in an interview with Newfields (2003), said that entrance examinations could become irrelevant due to the declining numbers of applicants and to the fact that there are now alternative ways to enter universities (e.g., recommendation admissions or special admissions for students returning from abroad). However, others, such as Poole (2003), contend that the Japanese university education system is by nature unchanging, and has been so since the Meiji era. According to Poole, passing an entrance examination is still usually required for university entrance.

Statement of the Problem

The first problem concerns a general lack of awareness of the potential usefulness of application information typically submitted to universities by high school students. Such information often includes individual high school grades and a high school grade point average (HSGPA); *hensachi* rankings of students' high schools; personal histories, including information about study abroad; letters of recommendation; and non-cognitive demographic information (though typically more limited in scope than that which American universities request). While these kinds of application information are given little attention in Japan by university administrators and researchers, they have been extensively studied in other countries in order to determine how well they predict later academic performance and student retention (e.g., Astin, 1993; Ramist, Lewis, & McCamley-Jenkins, 1994; Ransdell, 2001; Zwick, 2007). Studies are needed to establish a starting place in determining to what extent different types of application information can predict students' success in Japanese universities.

The second problem concerns the lack of information regarding the predictive validity of examinations used by Japanese universities to assess applicants. Because little has been written in English about the Center Examination, there are many unanswered questions about the degree to which it predicts university success. In addition, the pervasive university entrance examinations appear to have many weaknesses, and have therefore been frequently criticized. One weakness that Brown and Yamashita (1995) pointed out regards the type of test item and test method. They reported that items included on English entrance examinations tend to vary greatly in terms of the type of item used (p. 91); that is, universities differ in the test method used (e.g., multiple-choice, translation, and essay), and the language skills tested (e.g., listening, reading, writing, or speaking). Because the test items require different levels of English ability (e.g., multiple choice items, as compared to essay writing tasks), and the type of skill being tested reveals different information about the degree to which students have acquired receptive and productive English skills, these tests are clearly not measuring the same abilities. As a result, test results are neither comparable, nor uniformly useful in assessing the English abilities of applicants. Studies are needed to determine whether the English entrance examinations (and entrance examinations in general) and the Center Examination predict later academic success, which, as Zwick (2002) pointed out, is their most important role.

The third problem concerns the lack of any theory about student success in Japan; practically nothing is known about the "big picture" of how variables interact over time to account for student success in university. For example, no theory has been proposed to explain which variables (e.g., individual high school grades, test scores, *hensachi* rankings, and non-cognitive variables) are most important in predicting student completion of academic program components on time or graduating with honors. The larger picture of how these variables relate to each other over time, from before the first year in college to eventual graduation, has also not been examined in Japan. Study is needed in this area, so that researchers, teachers, and administrators can understand academic success in Japan from a wider perspective and become better equipped to assist students in their efforts to proceed smoothly through each phase of their university careers.

Purposes of the Study

The first purpose of the study is to investigate to what extent student application information can predict later university success, as has been examined in studies overseas. The variables examined in hierarchical multiple regression analyses, aimed at predicting success in an EAP and regular university program are: HSGPA, high school grade factor scores, total and section scores of the TOEFL, high school *hensachi* rankings, and non-cognitive data derived from student questionnaires and from information obtained through the Student Records office. It is hoped that the results of the study will help to inform the admissions process in the university under study, and in Japanese universities in general.

The second purpose of this study is to investigate to what extent Center Examination test scores and university entrance examination results predict later university success. Center Examination section scores are first examined through hierarchical multiple regression analysis to determine whether any of the scores predict academic performance in the EAP and first year of regular university study. As already noted, the entrance examinations are a primary source of information that universities make use of in selecting students. In order to evaluate how well entrance examinations predict university success, type of examination passed is used as a dummy variable in a hierarchical multiple regression analysis in order to determine to what degree the four types of university examinations given at the university under study are capable of predicting later success in the EAP and regular university program. In addition, four numerical test scores from the examinations (i.e., English test, Japanese test, English essay, and interview scores) are used in further hierarchical multiple regression analyses to determine the relative value of these measures for predicting later EAP and regular university success.

The third purpose of the study is to examine the “big picture” of how cognitive and non-cognitive variables interact over time in accounting for student success in the EAP and regular university program. One aspect of this larger picture of success includes completion of program requirements in a timely and exemplary manner; therefore, completing the EAP program on time, graduating on time, and graduating with honors are examined through logistic regression, in order to determine the variables that account most for this kind of success over time. A second aspect of this larger picture of success is how variables interact over time, from before the first year of university through graduation. This interaction of variables is examined through path analysis.

Significance of the Study

There is now practically no research on factors that predict academic success of university students in Japan. The current study provides a starting point for this kind of research and sheds light on the question of student success in a number of ways. First, the findings can help admissions officials sift through application information that they routinely receive, and attend to the most significant information. If they emphasize these important variables in the process of selecting students, they will be more likely to choose candidates with higher chances of future success in their institutions. With such knowledge, university officials will be in a better position to make fairer and more informed decisions.

A second contribution of this study is to provide admissions officials and university administrators with more information about the predictive value of the Center Examination and a variety of university entrance examination types and test scores. Because entrance examinations are an important part of the admissions landscape in Japan, it is important that officials and administrators understand how well these tests stand up to scrutiny, especially in terms of how well each accomplishes its primary purpose of predicting student success in university.

The third contribution of this study is to establish a starting point for a theoretical understanding of the larger picture of university success in Japan. This larger picture of success, being related to exemplary program completion, and to the overall path of success, might also lead to other worthwhile investigations, such as further study of the many possible predictors of academic success for Japanese university students in Japan, as well as for those attending institutions abroad. It is also hoped that the findings also provide an impetus to improving the Japanese university admissions process.

The Audience for the Study

The primary audience for this study is university staff and faculty members working in English-medium Japanese universities who are charged with the important task of deciding who is to be granted and denied admission. The information provided by the study should help these people identify candidates with the greatest potential for success, as well as consider how to support students who might be more likely to fail or drop out. In addition, information from the second part of the study should be of use to staff and faculty members who are given the important task of constructing entrance examinations.

Another audience is educational researchers studying the academic success of Japanese students in Japan, and abroad. As noted, little is currently known about factors that predict the academic success of Japanese students. The information derived from the current study establishes a starting place and suggests other avenues for study. In general, the results should be of use to those interested in factors related to student university success and graduation.

Delimitations

Because the data for this study were gathered from students enrolled in an international, English-medium university in Japan, the results should only be generalized to similar students in a comparable setting. That is, the results are most pertinent to comparatively motivated Japanese students who have chosen to enter a university that requires them to use English and to do considerable work in class and for homework. Because Japanese university students enrolled abroad have also chosen a similar setting and challenge, though, the results of the study might also be generalized to them.

The results of this study probably do not generalize well to Japanese students studying at Japanese-medium universities, particularly those majoring in disciplines other than English. Although it is possible that results regarding HSGPA and *hensachi* rankings might have relevance for these students, such claims cannot be made with confidence because of probable differences in curricula and grading standards.

Organization of the Study

Chapter 2, Review of the Literature, is divided into Cognitive Variables, Non-Cognitive Variables, Gaps in the Literature, and Purposes of the Study. Under Cognitive Variables, I discuss Theories of Intelligence and the *g* Factor, the *g* Factor in SLA, Criticisms of the *g* Factor Theory, Multiple Intelligences—An Alternative to the *g* Factor Theory, Emotional Intelligence, Sternberg's Theory of Successful Intelligence, Using Cognitive Measures to Predict University Success, and Studies of English Proficiency Measures as Predictors of Academic Success. Under Non-Cognitive Variables, I discuss Gender and Parents' Education, and under Purposes of the Study, I present Research Questions. In Chapter 3, Methods, I describe the participants, materials, data collection, procedures, and analyses. In Chapter 4, Results, I present results to the nine research questions, and in Chapter 5, Discussion, I explain and interpret the results to the research questions. Chapter 6, Conclusion, is divided into Summary of Study Findings, Limitations of the Study, Suggestions for Future Research, and Final Conclusions.

CHAPTER 2

REVIEW OF THE LITERATURE

Because most research related to predicting the academic success of students comes from countries outside of Japan, and particularly from the United States, I primarily discuss the variables that have been found to best predict the later academic success (and retention) of American university students. As Ransdell (2001) said, two types of measures have been used to predict academic success: cognitive/ability measures and non-cognitive measures.

Cognitive measures have a long history of being used to predict academic success. For example, as Ardila (2001) and many others have pointed out, the intelligence test devised by Binet and Simon was originally intended to be used to predict performance in school (p. 412). Cognitive measures that have been frequently studied in the United States include standardized tests, such as the SAT, ACT, and achievement tests, as well as HSGPA, student class rank, and essay ratings. These measures are also often used by universities to assess students' potential for future success.

Non-cognitive variables were described by Ransdell (2001) as factors such as interest in school, desire to study, academic support from parents, motivation, time spent studying, strategy use, choice of study group/individual study, and time spent working or doing extracurricular activities (p. 359). Reason (2003) cited a number of researchers who also studied the predictive value of conscientiousness, commitment, persistence, gender, race and ethnicity, and socioeconomic status. In fact, many variables have been classified as non-cognitive. In Astin's (1993) study of over 20,000 students and 200 colleges, he looked at a wide variety of factors in addition to those already noted, including drive to achieve, commitment to affect social values, commitment to being very successful financially, religious orientation, orientation to hedonism, time spent watching television, time spent commuting, and peer socioeconomic status (pp. 188-189). Some of these variables are lifelong characteristics, while others relate primarily to the manner in which students manage their university lives; however, all have a degree of potential for predicting students' chances of succeeding and remaining in college.

In my review of cognitive and non-cognitive measures, I focus most attention on the cognitive variables. I first present a history and discussion of relevant theories of intelligence; such a history offers insight into currently used cognitive measures and the theoretical grounding necessary for improving these measures. With this background in mind, I go on to discuss pertinent cognitive measures and research findings regarding how well each predicts university success. I begin with measures studied extensively in other countries, HSGPA and standardized test scores, and also discuss the importance of determining the validity, reliability, and unidimensionality of the GPA. Afterwards, I discuss findings regarding measures that might also have power to predict academic success for Japanese students—the Center Examination, university entrance examinations, *hensachi* rankings, and English proficiency test scores. Although there is little literature available in English regarding the first three measures, I present an extensive and critical discussion of findings from studies that have examined the possible predictive power of English proficiency test scores. In the review of non-cognitive variables, I focus on only two variables, gender and parental education (an aspect of socioeconomic status). I present a short discussion of findings regarding the relationship of each variable to academic success.

Cognitive Variables

Cognitive measures used as criteria for university admission typically have included tests, such as the SAT, ACT, GRE, and the GMAT. As Sternberg (2004) noted, these and other traditional IQ or aptitude tests have been found to predict success in many contexts (p. 188). He said that these tests were created with the practical goal of measuring cognitive and academic skills necessary for university success. Though all these tests are linked to intelligence, there is no underlying theoretical base for these and other such intelligence tests, according to Sternberg (p. 185). That is, instead of first defining the construct of intelligence and writing test items in terms of that definition, test writers have typically tended to create tests illogically, by defining intelligence by the ways it is operationalized (p. 186). This improper method of constructing these tests leads to serious problems; without a theory, test writers have no clear way of writing items for a test, or of validating a test. Therefore, according to Sternberg, cognitive measures have serious deficiencies, even though they have been shown to be linked to different kinds of success and intelligence.

Theories of Intelligence and the g Factor

Because cognitive measures are closely tied to the construct of intelligence, a review of pertinent theories of intelligence is in order. I first summarize a short history taken from Dörnyei (2005). One theory that has held considerable sway attributes intelligence to a *g* factor, or one general underlying intelligence or ability factor. Dörnyei pointed out that *g* was originally derived by computing a single higher-order factor from intercorrelating ability sub-tests; therefore, *g* was thought to represent the common variance that is part of various abilities (p. 32). Dörnyei explained that other researchers followed and expanded this idea. For example, in the 1920s, Spearman took intelligence to be a combination of *g* "which is available to an individual to the same degree for all intellectual acts" and "several specific factors which vary in strength from one act to another" (p. 32). In the 1930s, Thurstone postulated that intelligence was made up of seven main mental abilities: verbal comprehension, word fluency, number facility, spatial visualization, associative memory, perceptual speed, and reasoning (p. 32). Later, in the 1960s, Guilford hypothesized a structure-of-intellect model that was composed of more than 100 factors (p. 32). Cattell, also writing during approximately the same time, returned to the idea of general intelligence, dividing it into fluid intelligence, which relates to an ability to adapt to new situations, and crystallized intelligence, which is more akin to knowledge and acquired academic skills (p. 32).

Clearly, the *g* factor has been an influential concept in the intelligence field. As Sternberg and Hedlund (2002) pointed out, *g* is the “most widely studied and validated predictor of performance in employment and educational settings” (p. 143).

The *g* Factor in SLA

As Ellis (1985) pointed out, the concept of the *g* factor has also been appropriated by SLA researchers. He noted that Oller and Perkins claimed that “the ‘g’ factor of language proficiency is identical with the ‘g’ factor of intelligence” (p. 110). However, Ellis noted that this view was problematic in the sense that the *g* factor does not appear to be necessary for L1 acquisition (p. 111); that is, children of all intelligence levels (except for the extremely retarded) can acquire their native language, a fact that leads one to question whether intelligence is also unrelated to L2 acquisition. Ellis went on to resolve this point by showing how Cummins reconciled both sides of the issue (p. 111).

Cummins (1980) proposed that there are two kinds of language ability: cognitive/academic language proficiency (CALP) (p. 175) and basic interpersonal communicative skills (BICS) (p. 177). He proposed that CALP includes areas of proficiency in language that enhance the development of academic reading skills in both the L1 and L2 (p. 177), while BICS is similar to speaking ability needed for interpersonal communication (1982, p. 2). According to Cummins, BICS develops rather quickly, in 18 months to 2 years, but CALP can take from 5 to 7 years (p. 6). In fact, he said that this cognitive/academic component is the same in L1 and L2 (1980, p. 175), and that it is the "major determinant of educational progress" (p. 178).

According to Ellis (1985), BICS develops naturally, but CALP is a different aspect of proficiency because it is closely tied to skills needed for academic success and appears to be equivalent to the *g* factor proposed by Oller and Perkins (p. 111). Ellis concluded that intelligence does appear to impact SLA, at least in terms of CALP; he said that CALP probably predicts SLA performance in classes in which teachers use traditional and formal teaching methods (p. 111).

Criticisms of the *g* Factor Theory

Many researchers believe that the *g* factor is far from sufficient in being able to account for all that intelligence encompasses. For example, Howard Gardner, well known for his theory of multiple intelligences, has criticized the attempt to measure intelligence unidimensionally through IQ tests. One criticism pointed out by Gardner and Moran (2006) is that IQ tests, though correlated with similar tests and with academic success, have a weaker relationship with success in the real world (p. 227). In addition, despite a general correlation between IQ test scores and some kinds of success, there is no consensus about what the tests measure or predict. Finally, they criticized IQ tests for producing only scores that lead to either good or bad feelings, but not to opportunities for teaching and learning. Despite these problems, Gardner and Moran lamented the fact that IQ tests continue to be routinely used and are “well entrenched in the educational psyche” (p. 228).

Another researcher who has criticized the notion of using a single *g* factor in intelligence testing is Robert Sternberg. He and Hedlund (2002) have noted that the notion of a *g* factor is limited and controversial (p. 144). Such problems include the following: the *g* factor of intelligence actually accounts for only about 20% to 25% of the variance in performance, and therefore does not explain up to 80%; IQ tests based on the *g* factor tend to produce results that separate racial and ethnic groups; IQ test items are often unrelated to “real life” problems; conceptions of a *g* factor assume that intelligence is unchanging and that *g* is able to reliably predict performance in a variety of situations, when in fact people can score differently in different situations; in addition, researchers do not agree on what *g* represents on a psychological level or on what IQ tests measure (p. 144). Like Gardner and others who view intelligence as a multi-dimensional construct, Sternberg and Hedlund made the point that “Many people—researchers and laypersons alike—agree that there is more to intelligent performance than what is measured by a standard IQ test” (p. 144).

Multiple Intelligences—An Alternative to the *g* Factor Theory

Gardner's theory, as originally expounded in his 1983 text, *Frames of Mind*, proposed that eight intelligences exist: linguistic, logical-mathematical, musical, spatial, bodily kinesthetic, naturalistic, interpersonal, and intrapersonal; in addition, he has been examining the possibility of an existential intelligence, but has not yet determined whether it meets all the relevant criteria. Gardner and Moran (2006) explained that his theory has come out of an amalgamation of research from many disciplines and that the focus of the theory is how these intelligences interact; such interaction is of central importance because it illustrates the way that minds function (p. 228). Gardner has said that though the theory of multiple intelligences "does not lend itself easily to testing through paper-and-pencil assessments or a one-shot experiment," developing tests that measure the primary parts of each of the intelligences is possible (p. 230). To date, however, such tests have not been developed, and may never be, given Gardner's preference to help learners understand and develop their own abilities, rather than to create tests that attach rankings and labels to people (p. 230). In conclusion, then, the theory of multiple intelligences stands in opposition to a unitary *g* factor theory of intelligence, but at present does not offer measures that might be useful for the large-scale prediction of academic success.

Emotional Intelligence

The theory of emotional intelligence (EI) also does not characterize intelligence as a single factor. As Mayer, Caruso, and Salovey (1999) pointed out, this theory was developed in the 1990s and defines EI as an ability to recognize emotions, draw conclusions about them, and take appropriate action (p. 267). In addition, EI is characterized as a “hot intelligence;” that is, it is intimately tied to how the self deals with emotion (p. 268). Emotional intelligence was conceptualized as including four areas: the abilities to recognize, take in, understand, and handle emotions (p. 273). Mayer, Caruso, and Salovey also constructed a measure of EI, called the Multifactor Emotional Intelligence Scale (MEIS) (p. 273) and tested it. As a result, they concluded that EI was able to qualify as an intelligence because it met the necessary criteria of indicating mental ability, being like, but not the same as, other mental abilities, and growing with age and experience (pp. 269-270). Though these researchers said that we are only just beginning to learn about emotional intelligence, EI is probably another important component of general intelligence; accordingly, they suggested that defining intelligence in a way that includes EI will enable researchers to predict important achievements in life with greater accuracy (p. 295).

Sternberg’s Theory of Successful Intelligence

In response to the limitations of the *g* factor and to the lack of an adequate theory of intelligence, Sternberg has researched and written extensively regarding his own theory of intelligence—the theory of successful intelligence (2004, p. 185). Successful intelligence is defined “in terms of the ability to achieve success in life in terms of one’s personal standards within one’s sociocultural context” (p. 186). The three aspects of successful intelligence, according to Sternberg, include analytical, creative, and practical skills (p. 187). These skills, he said, are central to universal processes underlying intelligence. (In recent years, Sternberg (2007) has extended his model of intelligence to also include the reasoning and moral judgment that are essential to wisdom.)

Sternberg has tested his theory of successful intelligence in terms of how well it can be used to predict performance in university and found that many of his measures of analytical, creative, and practical intelligence have added significantly to the prediction of later college GPA. In one study, for example, Sternberg (2004) collected scores from 777 students in their first year of college that were “the standard measures used today to predict college performance”—the verbal and math scores on the SAT-I (formerly known as Scholastic Aptitude Test) and HSGPA (p. 191), as well as scores from his Sternberg Triarchic Abilities Test (STAT), which measures analytical, creative, and practical skills. He used these scores to examine how well each predicted students’ college GPA. Through a series of hierarchical regression analyses, he derived a variety of findings, all pointing to the value of the STAT measures. For example, by adding the STAT scores to the SAT and HSGPA measures, 24.4% of the variance in students’ college GPA could be predicted (with the STAT scores accounting for about 8% of the total) (p. 191).

He also found that the three STAT measures and HSGPA were statistically significant predictors throughout the analyses (p. 191). Though he did not include specific statistics pertaining to tests of multicolinearity, Sternberg said that the purpose of running three final multiple regression analyses (with HSGPA, SAT scores, and one of the three STAT measures in each) was to examine whether construct overlap was present. Because each analysis showed that the beta coefficient of each STAT measure was statistically significant, Sternberg concluded that the STAT measures were valuable in adding to prediction of university GPA, and thereby implied that construct overlap was not problematic. Sternberg's final conclusion was that these measures could be used consistently to predict college GPA. Besides adding predictive power, the STAT measures were also found to decrease differences due to ethnic and racial group more than traditional measures such as the SAT did (p. 192).

Practical intelligence alone appears to be of great potential use for admissions testing. As Sternberg noted, tests of practical intelligence have been examined and appear to measure abilities that differ from those that traditional intelligence or ability tests measure, and can also explain more about performance (p. 188). In the second part of his 2004 study, Sternberg studied University of Michigan business school graduate students. He examined how well practical intelligence measures supplemented the GMAT (Graduate Management Admissions Test) in predicting business school GPA, as well as success outside of the classroom (success in business school was measured by GPA at the end of the first year, GPA at the end of the program, and score on a team project in consulting; success outside of school was measured by number of extracurricular activities, number of interviews and offers of internships and jobs, and starting salaries). He found that the GMAT and undergraduate GPA significantly predicted business school performance, but that the practical intelligence measures appeared to: (a) predict success in and outside of class; (b) measure abilities that the GMAT and undergraduate GPA did not measure; (c) contribute significantly (over and above GMAT and undergraduate GPA) to the prediction of business school success; and (d) reduce differences in scores of males and females, as well as racial groups (p. 197). Sternberg has concluded that practical intelligence tests offer a promising way to improve prediction of success, and that the theory of successful intelligence also offers a theoretical framework for including a wider range of measures that are useful for assessing the skills that students need to be successful in college (p. 192).

In conclusion, intelligence can be viewed from a number of theoretical viewpoints and is certainly related in important ways to academic success. As Williams and Burden (1997) and many others have pointed out, IQ tests can predict academic success (p. 19). However, current admissions tests used to evaluate whether or not a student will be granted access to university do not spring forth from a clear theoretical conception of intelligence. In addition, high school grades are problematic: The link between high school grades and intelligence is not clear, and how valid and reliable grades are is also an issue of concern that is rarely addressed. Therefore, the results from studies of admissions tests, typically used in relation to HSGPA, will always be less than satisfactory in explaining the precise kinds of intelligence that lead to academic success.

Until the day that admissions tests are clearly based on a coherent theory of intelligence, researchers wishing to predict student performance in college must work with the disparate, but best available tests and other measures of cognitive ability, such as high school grade point average, class ranking of students, and scores on application essays, all of which seem to predict college performance to varying degrees. In the following paragraphs, I discuss cognitive measures that have been used to predict university success, as well as measures used as criteria for entrance to universities in Japan.

Using Cognitive Measures to Predict University Success

Cognitive measures, as Ransdell (2001) noted, are “necessary but not sufficient markers of progress and prognosis” (p. 357). They are not enough in and of themselves because many factors are related to whether students succeed or not, and because no test or measure is ever perfect. In addition, Ransdell pointed out two other factors worth remembering: How well ability measures predict success varies from one college to another, and depends on the number of variables being studied (p. 359). Therefore, when attempting to understand how well cognitive measures predict university success, one must remember to consider other possible predictors, as well as to look at trends in the research as a whole.

According to Pike and Saupe (2002), researchers have been attempting to predict college grades for nearly a century in the United States (p. 188). Grades, as reflected in the grade point average (GPA) are a typical criterion measure because they are accepted by most researchers as an index of academic achievement, are easily obtained, and have practical importance in obtaining admission to graduate school or employment (Young, 1993). Completing degree requirements is another important aspect of academic success (Alderman, 1981; Nora, Barlow & Crisp, 2005; Tinto, 1975). Cognitive variables that have been used by researchers and hundreds of universities to predict university success include high school grades and standardized tests such as the SAT and ACT (and less frequently, achievement tests) (Young, 1990b).

In a paper commissioned by the National Association for College Admission Counseling, Zwick (2007) explained the history of admissions testing, and the related development of the SAT and the ACT. She noted that there is some disagreement about where and when admissions examinations began, but most people agree that they began in Germany and England in the mid-1800s. In the United States, admissions testing first began in the early 1900s. At that time, there was a "bewildering array" of university entrance examinations, and they varied significantly from one another (p. 4). In order to make examinations more uniform and systematic, 12 universities from the northeastern United States established the College Entrance Examination Board.

The College Entrance Examination Board created a standard examination with nine subject areas for the university members of the Board. In the early years, the examination was hand scored by officials at the Board. Later, in 1926, the revised examination, which was made up of primarily multiple-choice items, was administered to 8,000 university applicants. Called the Scholastic Aptitude Test, it was very similar to the World War I Army Alpha tests, which had been developed by psychologists and were much like IQ tests. The test became even more widely used after machine scoring of the test began in 1939; as a result, admissions testing became a "bona fide industry" (Zwick, 2007, p. 5). In 1941, the test was scaled so that the scores ranged from 200 to 800, with an average score of 500 and a standard deviation set at 100 (p. 7). In 1947, the Educational Testing Service (ETS) was established in Princeton, New Jersey, to administer testing. One other admissions test was developed in 1959, when the American College Testing Program in Iowa developed the ACT. The ACT was designed to be tied to the high school curriculum and materials (with four sections in English, mathematics, social studies reading, and natural science reading), while the SAT (with only verbal and mathematics sections) was regarded as more of a test of reasoning and aptitude for success in college (p. 8). Despite differences in design, the two tests have been shown to correlate highly, at about .92 (p. 12).

Zwick (2007) noted that the SAT has undergone a number of revisions over the years. In the 1990s, there were changes in content and in some procedures; in addition, the test was recentered in 1995 so that the average remained at 500. In

2005, another change came about, much because of a controversy caused by the 2001 comments of Richard C. Atkinson, who was president of the University of California at the time. Atkinson claimed that the SAT was not a fair test because it was more like an IQ test, and led to unproductive use of classroom time, such as drilling in analogies. He wanted the University of California to eliminate use of the SAT and to stop using quantitative formulas for admission; instead, he said, the university should use tests tied to high school college preparation courses, and focus more on admitting minority and low-income students (p. 12).

Beginning in 2002, the College Board trustees and representatives from the University of California met regarding criticisms of the SAT, and in March, 2005, a revised SAT was launched. According to Kobrin, Patterson, Shaw, Mattern, and Barbuti (2008), the goal of the 2005 revision was to further emphasize high school curricula and skills needed for success in university. The revised test has three sections: critical reading, mathematics, and a new writing section (which has multiple choice questions and a section requiring students to write an essay). Scored like the previous test versions, each section ranges from 200 to 800 points; in addition, the College Board has announced that score scales are equivalent for the math and verbal sections of earlier tests and math and critical reading sections of the 2005 revised SAT (Zwick, 2007, p. 8). (As was the case with earlier versions of the test, there are also additional SAT Subject Tests, used for evaluating knowledge in a variety of subjects. At present, 20 such tests are available.) In their validity study of this 2005 version of the SAT, Kobrin et al. found that the revised test was able to predict first-year university GPA to the same degree as past versions of the test, and that the new writing section was the best predictor of the three sections of the test (p. 19).

Researchers have studied standardized tests (particularly the SAT) and high school grades as separate variables, as well as in combination with each other. As Zwick (2002) explained, effectiveness of predictors is measured by how closely the predicted college GPA approximates the actual earned GPA (p. 85). When using more than one predictor, researchers typically report the multiple correlation coefficient, though they also refer to it as a correlation, or validity coefficient (Zwick, 2007, p. 14). As Astin (1993) pointed out, "Hundreds of studies using various measurements and methodologies have yielded strikingly similar results: college grade point average (GPA) can be predicted with modest accuracy (multiple correlation around .55) from admissions information" (pp. 186-187). Pike and Saupe (2002) also found nearly the same results, that models based on high school performance and test scores explain 25 to 33% of the variance in the first-year college GPA (p. 188). As specified above, most researchers predict first-year GPA, and less frequently, four-year cumulative GPA, graduating on time (sometimes referred to as retention), or graduating with honors. As Burton and Ramist (2001) explained, first-year GPA is an especially good measure because it is available relatively early in a student's career, and that GPA is based on a set of courses and grading standards that are more similar for all students than those from the latter part of students' college careers, particularly in a large university where there are many diverse fields of study.

HSGPA and standardized test scores.

According to Astin (1993), the most powerful predictors of university grade point average have been HSGPA and scores of admissions tests, with HSGPA almost always a better predictor than test scores (p. 187). In a study of more than 20,000 students and 200 schools, Astin found that HSGPA and the SAT Verbal test score were the two most important predictors of college grades. In addition, HSGPA was approximately twice as important as the SAT Verbal score, with both scores producing a multiple R of .50 (p. 188). Ramist, Lewis, and McCamley-Jenkins (1994), in a study involving more than 46,000 students from 45 universities, found similar results. Referred to by Zwick (2002) as "the most detailed and painstaking study ever conducted of the utility of the SAT as a predictor of college grades," this study has been especially well regarded (p. 85). As Zwick explained, the findings from this study were like those of other such studies in two ways. First, HSGPA predicted later university grades slightly better than admission test scores, and second, combining HSGPA with test scores resulted in better prediction of college grades (p. 86). Certainly, HSGPA and standardized test scores should be considered when predicting students' college performance.

Table 1 presents the overall results of the often-cited 1994 study by Ramist, Lewis, and McCamley-Jenkins, as well as results of a recent study of the revised

SAT by Kobrin, Patterson, Shaw, Mattern, and Barbuti (2008). Both of these studies were large in scale. As noted, Ramist, Lewis, and McCamley-Jenkins studied over 40,000 students from 45 universities. Kobrin, Patterson, Shaw, Mattern, and Barbuti studied more than 150,000 students from 110 universities. Multiple correlations of the predictors, HSGPA, SAT total, and the combined predictors, HSGPA and SAT total, are shown. It should be noted that the SAT total included mathematical and verbal sections in 1994, and mathematical and critical reading sections in 2008. However, as previously noted, the College Board has announced that the score scales for math and critical reading on the 2005 revised SAT are equivalent to those of the math and verbal sections of earlier tests. (Correlations for the new writing section are not included in Table 1.) Uncorrected and corrected correlations are shown, although the method of correction differed somewhat, as explained in the table note.

Though Table 1 presents statistics for the sample as a whole, it should be noted that the SAT has been shown to systematically overpredict university GPA for some groups, and to underpredict for others. This problem in prediction, referred to by Zwick (2002), as a “subtle and complex aspect of validity analysis” affects the prediction of first-year GPAs of particular ethnic groups, for example. In the case of some ethnic groups, regression equations tend to overpredict their university GPAs; in other words, students earn lower grades than predicted in their first year in university. In the current study, ethnic group will not be a focus of examination, but it is important to make the point that these measures do not predict equally well for everyone, and the differences in the efficiency of the predictors can lead to mistakes in how university admissions officials assess the predicted university GPAs of applicants, and use such information in their admissions decisions.

Table 1. *Unadjusted and Adjusted Correlations for HSGPA, SAT Total, and HSGPA and SAT Total with University First-Year GPA, as Reported by Ramist et al. (1994) and Kobrin et al. (2008)*

Predictor	Ramist et al. (1994)	Kobrin et al. (2008)
HSGPA	.39 (.61)	.36 (.54)
SAT total	.36 (.57)	.32 (.51)
HSGPA, SAT total	.48 (.68)	.44 (.61)

Note. SAT total scores included verbal and mathematics scores in 1994, and mathematics and critical reading scores in 2008 (scores from the new writing section were not included here). Adjusted multiple correlations are in parentheses above. Ramist et al. adjusted for both restriction of range and criterion unreliability, but Kobrin et al. adjusted only for range restriction.

Even though researchers consistently find evidence of overprediction and underprediction from admissions tests like the SAT, test experts do not view this problem as a fault of the test. As Zwick (2002) explained, testing professionals refer to this problem not as “test bias,” but as “prediction bias” (p. 20). Mattern, Patterson, Shaw, Kobrin, and Barbuti (2008) explain that the problem is that the relationship between test scores and first-year GPA “varies by subgroup” (p. 2).

Zwick (2007) reported that there are systematic under- and overprediction patterns connected to gender. As for general test performance, women usually score better on English and reading tests (specifically, on the ACT). In the past women also tended to outscore men on the SAT verbal test, but that pattern changed in

1972 (p. 24). Women also frequently outscore men on essay writing (p. 24). Men, on the other hand, tend to score better on the ACT math, science, and composite scores, as well as the SAT math and verbal sections. The overall result is that women's grades in university are underpredicted by regression equations based on the SAT and HSGPA, and men's grades are overpredicted (p. 25); in other words, women's grades in the first year of university are higher than predicted, and men's grades are lower. Young (2004) said that the average underprediction for females was .06 on a four-point scale, and noted that this underprediction was an important effect, especially because women constitute the majority of university students. Mattern, Patterson, Shaw, Kobrin, and Barbuti (2008) reported that they found that first-year university grades of females were .09 higher than predicted, and those of males, .10 lower than predicted (p. 8).

Zwick (2002) noted that researchers have worked to understand and explain overprediction and underprediction by gender since about 1973 (p. 147). Reasons have been suggested for this differential prediction, but Zwick (2007) said that such reasons are "murky at best" (p. 26). She listed a number of reasons that have been suggested: (a) greater numbers of female students who take the SAT come from families with lower levels of income and parental education; (b) because the number of male students in university math and science courses is larger, university grades of males tend to be lower as a result of stringent grading; and (c) female students tend to attend class, study harder, and be better prepared for university work (p. 24). In order to increase awareness of this "gender gap" Zwick (2002) recommended that researchers continue to study how the content of tests affects men and women, study tests and prediction of achievement at the same time, and investigate grading patterns in college. She also recommended that testing officials consider adding a writing test to the SAT and ACT (which was done in 2005, in the case of the SAT). Finally, she recommended that test information more clearly indicate that women's college grades will be underestimated by their test scores (pp. 156-157).

Researchers have also studied prediction of GPA beyond the first year of university, although such studies are far fewer in number, as compared to studies of first-year GPA. Burton and Ramist (2001) presented results from Wilson (1983), of studies of cumulative university grades from 1930 to 1980, and their own review of studies after 1980, and said that both reviews were based on "relatively few studies from scattered institutions" (p. 16). In summary, they found that their own review of studies generally confirmed Wilson's earlier study, in that SAT scores contributed significantly to prediction of final cumulative GPA and general college success, and that the combination of SAT scores with HSGPA was a better predictor than either HSGPA or SAT scores alone (p. 16). Besides predicting final GPA, this combination also predicted other measures of academic success, such as honors, and acceptance into graduate and professional schools (p. 25). Table 2 presents the overall results from the two reviews presented by Burton and Ramist. The review of studies from 1930 to 1980 was based on eleven studies and more than 9,000 students, while the review of studies after 1980 was based on twelve studies and more than 25,000 students. Multiple correlations of the predictors, HSGPA, SAT total, and the combined predictors, HSGPA and SAT total, are shown. It should be noted that these are uncorrected correlations.

Table 2. *Average Unadjusted Correlations for HSGPA, SAT Total, and HSGPA and SAT Total with Final University GPA, from Studies of Two Time Periods, as Reported by Burton and Ramist (2001)*

Predictor	1930 to 1980	1980 to mid-1990s
HSGPA	.49	.42
SAT total	.42	.36
HSGPA, SAT total	.47	.52

Note. Correlations were based on varying numbers of institutions and students. The correlations are weighted averages, or averages of correlations reported by all studies that have been weighted by the number of students included in each study.

In conclusion, Burton and Ramist (2001) said that their review of studies showed that high school GPA and SAT scores were “good” predictors of final university GPA, graduating with honors, being accepted into graduate school, and earning a graduate degree (p. 26). HSGPA and SAT scores also predicted graduation itself, but not as well as they predicted the other outcomes. Zwick (2002) reported similar overall findings. Burton and Ramist also found that predictions using HSGPA and SAT scores were similar for different ethnic groups, students with disabilities, and women. However, they encouraged further study because relatively few studies on prediction of final GPA have been conducted, and even fewer have studied whether prediction varies for minorities, students with disabilities, or women. More study appears indicated for these groups, as well as for students who speak English as a second language. According to Burton and

Ramist, this “very important and growing” group has been studied little, and not at all in terms of long-term success (p. 26).

Tables 1 and 2 both show that HSGPA and standardized test scores predict short- and long-term college performance in the United States. Therefore, it is plausible that both might also predict college performance in Japan to some degree (depending on the quality of each measure), and could be useful information for officials making admissions decisions. Although Japanese universities typically rely on their own entrance examinations and make little or no use of HSGPA, and only limited use of the one standardized national entrance examination (the Center Examination), more study of predictors of university success in Japan might help to inform, and eventually improve, the Japanese university admissions process.

Validity and reliability of the grade point average.

Messick (1975) explained that construct validity encompasses the entirety of validity, at least from a scientific perspective, and is more general than predictive validity. He explained that the two main requirements of construct validity are convergent evidence, or evidence that shows that the measure is related to other measures of the same construct, as well as any other variables that are theoretically linked with it, and discriminant evidence, or evidence that a measure is not related to measures of other, different constructs (p. 956). Bachman (2004) also described validity as a unitary concept that is closely linked to how a test score is interpreted and used, and to the consequences of using that score. In order to justify such interpretations and use of the test, and to make a validation argument, researchers must collect a variety of validity evidence; they typically analyze the content of the test, and correlate the measure with other measures of the same ability, as well as with future performance (p. 260). In the current study, the validity of GPA and test scores is examined, when possible, by correlating them with other measures of the same ability, as well as with future performance.

When confirming that measures used are valid, one necessary step is to see that they are reliable measures (Bacon & Bean, 2006). Reliability is an important issue in measurement, and can be defined as the extent to which a scale consistently reflects the construct it is purported to measure (Field, 2005). Another way of understanding reliability was explained by Clark (1964), who said that the reliability coefficient reveals how error-free a particular measure is. Zwick (2002) also referred to unreliability as "imprecision," and said that unreliability of class grades can come from a variety of sources, such as restriction in the range of ability of students, differences in the grading stringency of teachers, and other problems associated with grading, such as inconsistency and bias (p. 88). In explaining how grading practices relate to error, Clark said that high reliability coefficients indicate that grading is done carefully, according to clear standards, while low coefficients indicate that grades were probably based on short tests and subjective impressions of teachers. Because there are many sources of measurement error, Bedeian, Day, and Kelloway (1997) noted that it is not uncommon for variables in the social sciences to have reliabilities of less than .70, with R^2 values below .50 (p. 785).

Speaking specifically about the reliability of independent variables, Pedhazur (1997) said that values are usually moderate at best (.50 to .80) in non-experimental research (p. 34). Reliability of the GPA should be addressed in predictive validity studies because, as Schmitt (1996) explained, unreliability in the criterion measure produces a considerable underestimate of the relationship between it and any predictor variable. Pedhazur (1997) explained the impact of unreliability in terms of the type of variable; unreliability in the independent variables can lead to underestimating the regression coefficient while unreliability in the dependent variable can lead to a larger standard error of estimate, which in turn weakens significance tests (p. 34).

Even though reliability is an important issue to address in predictive validity studies, many researchers who use GPA do not discuss the reliability of it or other measures used in their studies. As Zwick (2002) noted, determining the degree of reliability (and the degree of error and bias) in GPA is quite difficult; therefore, researchers avoid discussing the subject (p. 75). Bacon and Bean (2006) concurred; they contended that GPA was not understood well in their field of marketing research. In addition, they said that none of the studies they reviewed had discussed the reliability of GPA.

Though perhaps not in the majority, a number of researchers have discussed the importance of estimating and improving the reliability of the GPA. As Armor (1974) pointed out, improving the precision and reliability of measures means that it is possible to establish clearer relationships with other variables, and to more accurately determine the predictive validity of measures.

There are a number of ways to assess reliability, including test-retest, equivalent forms, and internal consistency reliabilities (Brown, 1988). The last, internal consistency, includes the Spearman-Brown prophecy formula and Cronbach’s alpha, which have been used to assess the reliability of the GPA (Bacon & Bean, 2006). Ferketich (1990) and Armor (1974) explained two additional coefficients that can be used to assess GPA reliability, theta and omega. Though there are a variety of formulas to estimate internal consistency reliability, Li and Wainer (1998) made the point that (as long as assumptions are met) equations that appear to be very different can produce very similar estimates of reliability. However, in estimating the reliability of the GPA, theta might be one of the best estimates because it was proposed as a way to deal with data that potentially include more than one dimension or factor (a possibility for college grades). The formula is based on principal-components factor analysis, and uses λ, or the largest eigenvalue obtained, and *p*, or the number of items in the data set. The formula that Armor (1974) gave for theta in the case of a one-factor solution was:

$$\theta = p / p - 1 \, (1 - 1 / \lambda)$$

Finally, Werts, Linn, and Jöreskog (1978) suggested a method of assessing GPA reliability through examining correlations in longitudinal data. They examined eight semesters of undergraduate GPA and determined that the data had a simplex form, which they defined as meaning that the correlations between GPAs of two semesters decrease as the time between them increases. Bean and Bacon (2006) also observed this simplex form in their data, though they did not use this term. Instead, they said that the GPA “drifts over time,” so that correlations between consecutive semesters averaged .76, while the correlation between the GPAs of the first and fourth year was .60 (p. 36). Werts et al. explained a method of calculating reliability with data that is in this simplex form: the reliability of any observed measurement, r_{ii}, can be calculated with the following formula, where r_{in} is the correlation between that observed measurement and another previous measurement; r_{im} is the correlation between that observed measurement and another subsequent measurement; and r_{nm} is the correlation between the other two measurements (p. 90).

$$r_{ii} = \frac{r_{in} r_{im}}{r_{nm}}$$

Werts et al. said that if there is more than one previous or subsequent measurement, it is possible to have more than one reliability estimate, but if the model is simplex, all of the reliability estimates will be equal within the bounds of sampling error (p. 90). This method of estimating reliability is useful when information about specific courses is not available.

Other researchers, not content with simply estimating GPA reliability, have used different methods to adjust it (Elliott & Strenta, 1988; Young, 1990a; Young, 1990b); however, as Stricker, Rock, Burton, Muraki, & Jirele (1994) showed, adjustment methods vary and have a variety of problems. Although adjusted grades in their study correlated highly with the original GPA and appeared to be more reliable and predictable from admissions information, the results did not support the assumption that adjusted GPAs were in fact more valid indicators of success in college (p. 180). As Bedeian, Day, and Kelloway (1997) noted about the related issue of adjusting for measurement error, which is always present in real-world measurement, correcting for attenuation (and by implication, correcting GPA) can lead to overestimated correlations, possibly taking researchers into a “fantasy world” (p. 787).

Stricker, Rock, Burton, Muraki, and Jirele (1994) asserted that any problems of the actual, unadjusted GPA should not be overemphasized. They maintained that the unadjusted GPA is actually reliable, predictable from admissions information, and shown to have a simple factor structure (typically with one large general factor, perhaps along with one or two smaller factors). Similarly, Bean and Bacon (2006) found the GPA scale to be reliable and suggested that it was likely that student grades measure the same latent construct. Burton and Ramist (2001) also suggested that grades will continue to be important as predictors and criteria in university admission, along with standardized test scores. They said that using scores of tests such as the SAT help compensate for any unreliability and subjectivity in grades while grades help compensate for the limited scope of test scores.

Unidimensionality and the grade point average.

As noted above, Stricker, Rock, Burton, Muraki, and Jirele (1994) pointed out that the factor structure of the GPA should be evaluated. Specifically, multidimensionality in measures can be problematic. As Young (1990b) stated, the GPA is problematic not only because departmental grading standards differ (resulting in lower construct validity of the GPA scale), but also because the GPA is a composite of grades in a wide range of courses. As such, he said, it will be a multidimensional, and therefore unreliable, measure (p. 320). However, Young (1993) also pointed out that a composite usually becomes more reliable as more items are included, as long as the items (in this case, college courses) are the same. Therefore, if the GPA for a particular group of students is a composite of comparable course grades, the chances of it having unidimensional and reliable qualities improve.

Bond and Fox (2007) in discussing test dimensionality, explained that unidimensionality is a “focus on one attribute or dimension at a time” (p. 32), though this focus on one attribute does not mean that the performance is a result of only one mental process. They asserted that more than one psychological process can be involved in a performance, but the measure is unidimensional as long as the processes all work together in a similar way (p. 140). Therefore, they said that though educational achievement can be conceptualized as multidimensional (so that achievement in math, language, science and other courses is each measured separately, resulting in a number of different dimensions), educational achievement can also be thought of as unidimensional (so that achievement is conceptualized in a wider and looser manner). This logic is applicable to GPA as well.

Researchers who have investigated the dimensionality of GPA have used a variety of methods. As already noted, Armor (1974) determined dimensionality with a principal-component factor analysis before going on to estimate reliability. Schoenfeldt and Brush (1975) also used factor analysis (with Varimax rotation) and found three factors in their study, but it should be noted that the variables included HSGPA and test scores, together with a set of specific university grades. Though there were three factors, they said that there was one primary factor, General Academic Achievement, which accounted for 23.4% of the variance (p. 318). They concluded that GPA was not perfectly unidimensional, but it could be generally considered as such for predictive validity research. Stricker, Rock, Burton, Muraki, and Jirele (1994) agreed that the factor structure of the GPA is simple, and typically contains one large general factor, and one or two smaller factors.

Schmitt (1996) said that though some researchers have used Cronbach's alpha as a measure of unidimensionality, it is not an appropriate index to gauge dimensionality because one can find a high level of alpha in a multidimensional set of data. He proposed another increasingly popular method for assessing unidimensionality: testing if a correlation matrix fits a unidimensional model, by assessing fit statistics produced by LISREL or other structural modeling software. Schmitt also made the point that researchers usually cannot clearly determine whether unidimensionality is unambiguously present or absent in a set of data.

Rather, he suggested that researchers ask how closely the data come to being unidimensional. According to Schmitt, if a measure has “reasonable unidimensionality” and covers appropriate content, it can be a useful measure, even if it has low reliability (p. 352).

Yu, Popp, DiGangi, and Jannasch-Pennell (2007) discussed a number of methods that have been used to evaluate dimensionality. One widely used method is confirmatory factor analysis using EQS or LISREL software. However, they noted that factor analysis can overestimate the number of latent factors because it confounds item difficulty variance with dimensionality. In their study, they chose to evaluate three methods for assessing dimensionality (the Rasch model, Parallel Analysis, and TETRAD). They found that the results from the different methods did not agree, based on usual standards of interpretation. Therefore, they stressed the idea that researchers must often interpret and make judgments as they deal with ambiguity in assessing unidimensionality. For example, their results showed two factors, but because the two were highly correlated, they concluded that the set was unidimensional within the bounds of Rasch analysis. They justified their conclusion by pointing out that researchers must consider the practical use of a scale or instrument. If one score is used as the basis for judgments and inferences, then researchers must establish that the scale or instrument is unidimensional.

Burg (2008) investigated the dimensionality of math achievement tests used for American children in the third to eighth grades. She found evidence of multidimensionality and ambiguity, but made the interpretation that math skills being tested in five content areas overlapped to such a degree that she considered the test to be fundamentally unidimensional. Again, the researcher's judgment and reasoning were needed to make final conclusions about unidimensionality of the data set.

In conclusion, assessing the unidimensionality of GPA is challenging. As Burton and Ramist (2001) noted, grades represent a variety of academic and nonacademic skills, and researchers know that grades are often not comparable or reliable. Particularly if the school has a diverse assortment of majors, it is probably not reasonable to claim that grades from all courses are related to a single latent construct (Young, 1993). On the other hand, it is possible to use the looser conceptualization of unidimensionality that other researchers have used, and to argue that all grades measure a more general dimension of academic achievement; in addition, this argument is even more substantial if the university in question has a limited number of fields of study, and more courses in common (Young, 1993). If analysis of the data that is available in this study reveals one large general factor after extraction, then it can be argued, as Young (1990a) and other researchers have done, that the GPA is sufficiently unidimensional.

The Center Examination.

The only standardized nationwide entrance examination in Japan is the *Daigaku Nyuushi Senta*, or the Center Examination. The Center Examination is sometimes referred to as Japan's SAT, probably because, as Mori (2002) pointed out, early forms of the exam were modeled on the American Scholastic Aptitude Test (p. 32). It covers five areas (with subtests), including Japanese language, social studies (including geography, history, civics), mathematics (including mathematics IA, mathematics II B, and bookkeeping and accounting), science (including biology, chemistry, physics, and geology), and foreign language. All national and public universities require applicants to take the Center Examination, and many private universities and junior colleges also do so. When applying for the examination, students select test subjects, depending on the requirements of the universities they hope to enter (National Center for University Entrance Examinations, 2003, p. 9). The Examination is administered in January each year, and the cost has been gradually increasing. According to the NCUEE current website, students who take the Examination in 2011 will pay 18,800 yen for three or more subject tests, and 12,800 yen for one or two subject tests.

The Center Examination is administered to more than 500,000 students every year (Guest, 2008), and scores enable students to identify which university entrance exams (called *Niji*) they are most likely to pass (p. 86). Guest explained that many universities establish a minimum Center Examination score for applicants. If students score below this minimum, they will not be eligible to take the *Niji* examination of that university. In addition, some universities use applicants' Center Examination scores as part of a total score needed for admission to the university.

Mori (2002) explained that there have been several attempts to create this kind of common university examination in Japan, with first efforts going back to the early 1900s. It was hoped that a common examination would help alleviate "examination hell," the intense pressure students felt as a result of preparing for individual university examinations (pp. 31- 32). That is, the plan to enable students to enter universities by means in addition to the one university entrance exam was designed to reduce the intense competition and stress that students felt when applying to enter universities. This planned use of a common examination was unlike the initial plan for the SAT. As noted earlier, the College Entrance Examination Board created a standard and uniform examination in order to replace the array of university entrance examinations that were used at the time.

The early attempts at creating a common examination in Japan failed because the intense competition caused by individual university entrance examinations remained strong. After World War II, the Ministry of Education again tried to institute a common examination. In 1948, it was proposed that the common test be modeled on the Scholastic Aptitude Test, the well-known American test of academic aptitude. From 1963 to 1968, the Ministry attempted another common test. All of the earlier attempts lasted a very short time, due in large part, according to Mori, because those in academia did not trust policies created by the Ministry of Education (p. 32). Finally, in 1971, the Ministry of Education proposed the *Kyotsu Ichiji*, a common examination that eventually led to the Center Examination. This test was designed to measure high school academic achievement, and to be used, along with the individual university examinations, to choose students who had the "basic academic ability demanded in higher education" (p. 32). In 1977, *Daigaku Nyushi Senta*, or the National Center for University Entrance Examinations, was created, and first administered the test in 1979 (p. 33). As Otsu (2004) pointed out, the NCUEE changed from a national to an independent organization in 2001; according to Otsu, this change transferred most organizational power to the President of the NCUEE (p. 32).

Guest (2008) discussed revisions of both the *Kyoutsuu Ichiji* and the Center Examination. Initially, the *Kyoutsuu Ichiji* was criticized for not testing analytical skills, creativity, and critical thinking, and for being inappropriately difficult. As a result, the Interim Education Council (*Rinji Kyouiku Shingikai*) worked towards revising the test. In 1991, a new system using the *Daigaku Nyuushi Senta Shiken,* or Center Examination, began. Since that time, the Ministry of Education, Culture, and Technology (MEXT) has made a number of revisions and placed more emphasis on comprehension and analytical skills and less on memorization. The latest revision—an inclusion of a listening section for the English test section, in 2006—was the result of a MEXT goal of improving the communicative skills of Japanese people (p. 87). If students choose to take the English test, they also sit for this listening test.

The Center Examination is clearly an increasingly important examination in Japan, and the NCUEE, a growing organization. Indeed, as can be seen on the organization's current website (http://www.dnc.ac.jp), there are many committees and many branches of the organization. Besides an Advisory Board, President, Inspector, Vice President (Administration), Vice President (Academic), Deputy Vice President (Director of Research Division), Deputy Vice President (Planning and Coordinate Officer of NATLaS), Test Items Adjustment Officer, Test Items Researcher, and Listening Test Equipment Adjustment Officer, there are three large Divisions (Operations Planning, Test Administration, and Research), as well as at least 15 committees.

Otsu (2004), in a slide presentation for NCUEE in China, explained more specifics about the organization. He said that NCUEE employed 109 full-time staff members in 2004, and that the budget that year was 11,214 million yen, with 83% of that amount going to test administration (pp. 38-41). He stated that NCUEE did not give a pre-test for “difficulty equating,” that the Examination was “not moderated with modern test theory,” and that score equating had not been done, except on two occasions (1989 and 1998), to compare mean differences between selected subjects within the same discipline (pp. 24-37). He also pointed out that those in charge of developing test items were university professors, not full-time employees of NCUEE (p. 38). Hayashi (2005), in another slide presentation for NCUEE in China, noted that there are approximately 450 of these university professors involved in writing test items (p. 20). Both Otsu and Hayashi said that the process of developing test items involved three committees. Otsu explained that the responsibility of the first committee was to write test items, the second, to review the items and look for problems, and the third, to “moderate the problems between subtests or areas” (p. 39).

Views of the Center Examination vary. Some researchers and administrators praise the Examination. On the current NCUEE website, the President said that the test was playing an “important role” because more students and universities were making use of it. In addition, he stated that test items had been “highly evaluated” due to the teamwork of those on test planning and assessment committees (p. 1). (However, the President did not specify exactly who had evaluated the test or how it had been evaluated.) He added that their organization continues to “analyze, evaluate, and actively improve” the test, to “openly share ideas with universities,” and to work towards the “goal of carrying out a better test” (p. 1). Similarly, Hashimoto (1998) said in a conference presentation in Australia that the NCUEE examination system operates effectively and that those who work at high schools and universities have a high regard for it. It is also clear that teachers in Japan regard the Examination as important and spend class time to help students prepare for it. Kobayashi and Rinnert (2002) said that students in their study reported that a “large part” of their last year in high school was used to prepare for the Center Examination (p. 100).

The primary source of information about the Center Examination is a journal entitled *Forum,* which is written in Japanese and has been published since 1982 by The National Center for University Entrance Examinations (NCUEE). Because the *Forum* is unavailable at many universities, including my own, I requested that the NCUEE send me the volumes of the journal. Because an incomplete set was sent, ten missing volumes had to be located and reviewed. Most were found at either Tokyo University or Waseda University. At present, there is no computerized way to search for information within the journal; therefore, it was necessary to review the volumes, one by one, and page by page. With the assistance of a trilingual graduate of the Waseda University graduate school, I reviewed every volume of *Forum* from 1982 to the most current volume that was published in 2008, in order to find out what kind of research had been conducted by the developers of the Center Examination, especially that regarding test design, research goals, and statistical findings regarding the test, particularly those related to reliability and validity.

Regarding test design, only one article, in the 2006 issue (Volume 29, p. 5), appeared somewhat relevant. It concerned evaluation of the English listening test design. However, this evaluation did not include information about item analysis, or about the reliability or validity of the test. Instead, the main focus was to list four points about test design, including the following features: (a) the question and answer choices are to be on the student's script; (b) the question is to be repeated twice; (c) the speed of speaking is to be approximately 150 wpm; and (d) items are to be ordered from easy to difficult. In the volumes from 1982 to 2008, no other article about test design was found.

As for research goals, I expected to find specific goals stated in *Forum* volumes that were related to investigating and improving the test, especially because statements in the Preface section of the NCUEE website allude to general goals of improving the Examination. However, there were no such regularly stated, clear research goals in the *Forum*. Only two statements that were somewhat relevant were found. The first appeared in 1986 (Volume 8), when it was noted that high schools had been requesting a listening section for the English test, and that it was believed that a listening test would bring the "greatest educational effect" (my translation, p.79). Then, in 1994, acting President Ryohei Takahashi referred to NCUEE goals when he said that there was a need to make efforts to: (a) expand the use of the Center Examination; (b) increase the number of subjects and the curriculum studied in school; (c) expand the number of subjects and curriculum for the Examination; (d) prepare more items, with a wider range of difficulty, for each subject; (e) improve the timing of when the test is offered; (f) give the test many times in one year; and (g) inform students not only of their total test score, but also of their section scores. This was the only list of NCUEE goals found.

In terms of statistical findings about the test, the most typically reported statistics were descriptive, particularly numbers and percentages of students who answered section items correctly or were able to attain the Center Examination score required to enter preferred universities. No articles or reports that displayed results about reliability or validity of the Center Examination were located.

At times, research articles were listed in a few of the *Forum* volumes (as they were in Volume 17 of 1994). Though some of the titles in these lists of research pertained to statistical techniques such as canonical correlation analysis, regression, or path analysis, the focus of these articles was not the Center Examination; instead, the articles often came from other disciplines, such as mathematics and science. The only listed research article that discussed reliability and validity concerned two Japanese language personality inventories, not the Examination.

Only in 2008 (Volume 31) did a contributor, Yoshinori Watanabe, a professor currently teaching at Sophia University, discuss the validity of the Center Examination (pp. 41-43). Watanabe said that validity was the most important area to study and then went on to explain content, construct, and predictive validity. He said that issues of validity were important for any test, but especially so for the Center Examination because it is used for making important decisions. However, Watanabe pointed out that testing the validity of the Examination would be an “innovation,” and therefore required five conditions. Of these conditions, Watanabe said that the last two, “triability” and “observability” were problematic because the Center Examination, being a large-scale and highly confidential test, was not amenable to the testing required to meet these last two conditions. In other words, he said, it was “very difficult to test beforehand and to verify the reliability and validity” and that it was “not possible to try to conduct studies on a small scale and say anything meaningful about the large scale;” his final statement was that innovation of the Center Examination was difficult, and any changes in the Examination would be “extremely difficult and risky” (my translations, p. 43).

Such a statement in the most current issue of the *Forum* does not suggest that validity or reliability studies of the Center Examination will be in the offing.

Research about the Center Examination in English is scarce, as Guest (2008) pointed out (p. 87). He noted that the first English references to the test appeared in 1995, when Brown and Yamashita wrote their well-known critique of Japanese university entrance exams. Guest cited a number of criticisms of the test, including that it was “poorly designed, discrete-item-based measure of grammatical minutiae and ‘testwiseness’ made by ‘amateurs’” (p. 87). Ten years after Brown and Yamashita’s study, Ito (2005) noted that he had been able to find only two studies of the English section of the Center Examination (both written in Japanese), one of which showed that high school English grades and the English test section of the Center Examination were correlated from about .33 to .62, and another that reported the reliability of English test sections to be between .94 and .96 (p. 93). Another report of the reliability of the Center Examination was obtained from a researcher who had worked with the NCUEE; he reported that the reliability of the test as a whole had been studied, and the resulting Cronbach’s alpha statistic was .86 (S. Ross, personal communication, June 19, 2010). Ito attempted to investigate the validity of the Center Examination English test by correlating these scores with scores from a cloze test he had constructed. Ito found that the Center English test scores and the cloze test scores correlated at $r = .82$; this result, together with the intercorrelations of the five subtests of the English section test, led Ito to conclude that the Center Examination English test was a “somewhat valid test of English ability” (p. 108).

Guest (2008) pointed out that no one has ever “publicly proposed” a purpose for the Center Examination, despite the fact that it has been used and revised since 1991 (p. 89). Guest made the point that the Center Examination is “*not* an achievement test” (p. 89), and that the main purpose of the Center English section is to “determine student aptitude for academic study of English at the tertiary level in Japan” and to help examinees understand their own rankings, and concomitant chances of succeeding in the *Niji* examinations (p. 89).

Though what Guest said about the English portion of the Center Examination might be generally true, he is not completely correct in stating that the examination is not an achievement test, according to the 2003 report of the National Center for University Examinations, which stated that the Center Examination aims primarily to “measure the level of basic academic achievement of students who desire to enter university” so that universities could “use the test results, applying their own criteria to measure examinees’ abilities and aptitudes from diverse aspects” (p. 1). In other words, though NCUEE personnel designed the Center Examination to function as a kind of achievement test, they did not specify exactly what the results mean or how universities should use the results.

Guest (2008) claimed that the revisions of the Center Examination have been creating washback that affects high school instruction positively. In his view, the Center Examination has improved in that it now focuses more on communication and “comprehensive, holistic reading skills” (p. 96) and might be coming closer to “reflecting (or fostering) healthy pedagogical and educational practices” (p. 87). In addition, he said that the variety of topics and genres and task types has improved the validity of the test (pp. 96-97).

Despite such a rosy view, the facts that MEXT has not completely clarified the purpose of the Center Examination and that there is so little research on it are worrying. Ito (2005) noted the same problems about the Center English section test. He cited the following concerns: Research on the Center English section test was “extremely limited;” access to information about the test and test results was insufficient; no information was provided by the test makers regarding test reliability and validity; and the exact nature of the test was unclear (p. 92). In addition, he said that the test was “supposedly an achievement test based on the Japanese high school curriculum” (p. 92), but that many believe the test to be a “university entrance examination developed by Japanese university professors” because high school teachers (who know the high school curriculum) are not appointed to the test development committee (p. 93). In the end, Ito concluded that the Center English section test measured both achievement and proficiency. In response to Ito’s criticism that high school teachers are not appointed to the test development committee, one researcher who had worked with the NCUEE disagreed; he reported that high school teachers participate in the test review panel and post hoc reviews (S. Ross, personal communication, June 19, 2010). In any case, it is clear that there are misunderstandings, as well as legitimate concerns about the Center Examination. The sorts of concerns noted by Ito raise serious questions about the Center English section test, and perhaps about each of the other five sections of the Examination.

It is imperative that the developers of a high-stakes nationwide entrance test like the Center Examination should clearly state the purpose of the test, specify what construct it is purported to measure, and assess and report the validity and reliability of the Examination. Unfortunately, as noted by Ito, studying the Center Examination is difficult because of lack of access to test data. For example, Westrick (2005) noted in his study of placement testing that “all information regarding student scores on the Center test and the university’s entrance exam is considered private and was unavailable for the study” (p. 78). Such control of Center Examination data is standard practice; however, access to this data is necessary in order for researchers to evaluate the examination and suggest ways to improve it.

Unfortunately, my careful review of *Forum* volumes did not yield information that can be used in the current study. Very little is known about the reliability and validity of the Center Examination, but it is possible that scores on the English section might prove useful in predicting university success at an English-medium university. In addition, sections of the Center Examination assessing student achievement in Japanese also might be useful predictors, even for students enrolled in an English-medium university, because many researchers believe that skills and abilities transfer from a person’s L1 to L2. If that is the case, strong scores in Japanese and other Japanese-medium tests on the Center Examination might predict achievement in English. As Saville-Troike (1984) pointed out, L1 transfer does occur, and plays an important role in second language acquisition. For example, she hypothesized that reading achievement in English is more related to reading ability in the student's L1 than it is to oral English proficiency. Therefore, Japanese-medium tests on the Center Examination might prove useful in predicting academic achievement in English-medium universities.

Similarly, Ward (1997) found a significant relationship between the L1 proficiency (especially reading and writing) of Hispanic students in the sixth through twelfth grades and their scores on standard achievement tests in English (pp. 342-343). In other words, students who were able to read skillfully, think logically, and write well in their native language seemed able to transfer these abilities to the L2. Therefore, it is plausible that Center Examination scores from Japanese sections of the test (e.g., Japanese, civics, geography and history, mathematics, and science) could also be useful in predicting academic success for students attending an English-medium university.

University entrance examinations in Japan.

According to Brown (1995) the criterion for admittance to most Japanese universities is passing the school's entrance examination, and this is particularly true for the more prestigious universities (pp. 23-24). Though Mori (2002) noted that beginning in the late 1980s, universities started to consider additional sources of information about applicants, such as recommendation letters from principals, student essays, and interviews (p. 36), others, like Poole (2003) have contended that entering university continues to be equivalent to passing the entrance examination. He said, “Though admissions procedures are becoming more creative in recent years, the majority of colleges have resisted any change in a system that has been in place, arguably, since the Meiji Era in the late 1800s” (p. 5).

Mori (2002) explained that university entrance examinations have been criticized since the early 1900s for questions that were far too abstruse for students just out of high school, and for not being well matched to the high school curriculum (p. 31). However, she explained that Japanese university education was originally instituted in the late nineteenth century in order to train civil servants who would be able to help Japan catch up with the West. Therefore, the purpose of the early entrance examinations was to choose students who could outscore other students. In other words, the entrance examination served as a “gate to the elite system” (p. 31).

As Poole (2003) pointed out, those who develop these university admission examinations are not testing or admissions experts; rather, they are an assorted group of faculty and staff members who happen to be assigned this task in a particular year (p. 9). From a modern testing standpoint, it is not surprising that such examinations would have a number of flaws.

Though Brown and Yamashita (1995b) noted that entrance examination items from the English sections of tests they analyzed in 1993 and 1994 were generally written clearly, with few errors (p. 97), they also criticized the tests on a number of fronts. For example, they said that reliability, validity, and practicality of the tests appeared weak because the test items did not match pedagogical practice, the tests tended to be short, and many test items depended on knowledge of a particular topic and its vocabulary (p. 98). In addition, they noted that Japanese universities apparently did not analyze their tests for reliability or validity, a common and necessary practice for examination writers in the United States.

Brown and Yamashita advised Japanese universities to follow the *Standards for Educational and Psychological Testing* or to create testing standards appropriate in Japan (p. 98). Brown (1995) also recommended that universities consider a variety of test scores, HSGPA, recommendation letters, essays, interviews, and other information, in making admissions decisions (p. 24). In short, universities should use many measures in order to properly assess a student's potential for college success. In fact, more universities are allowing students to enter by more than one method, especially in this era of economic difficulties and declining student population; however, Amano and Poole (2005) made the point that the most highly ranked universities continue to make admissions decisions based on the traditional entrance examination that includes a wide variety of subjects. They pointed out that entrance examination wars will not end because students will continue to compete to enter these highly ranked universities; in short, the issue of reforming entrance examinations is a permanent one (p. 694).

Another reason that Japanese universities typically rely on their own entrance examinations and make little or no use of HSGPA, and only limited use of the one standardized national entrance examination (the Center Examination) is most likely financial. For example, alumni in Japan do not contribute to their alma maters, perhaps because there are no tax incentives given for donations, as they are in the United States, for example. Therefore, universities are dependent on other sources of income, such as that generated by their own entrance examinations. Such income can be quite large. For example, according to Kamiya (2009), applicants pay 17,000 yen for individual examinations in public universities, and between 30,000 to 35,000 yen for each department they apply to in private universities. Amano and Poole (2005) said that private universities charge each applicant the equivalent of hundreds of dollars to take the examination. Therefore, the entrance examination system is a “multi-million yen business” that universities are quite reluctant to give up (p. 706).

***Hensachi* rankings.**

According to McVeigh (2002), *hensachi*, or *z* scores that indicate ranking of students, were first produced with computers in 1965, as an index of students’ academic ability, and are used to guide students in deciding which high school or university they are likely to be able to enter (p. 88). These individual *hensachi* rankings are considered confidential and are not used for applying to schools.

LeTendre (1996) explaining more specifics of how *hensachi* are used in Japan, said that tables of *hensachi* rankings are produced on the basis of results of practice tests that an agency (private company, cram school, or other school) has given. This agency then creates a list that includes students, scores, and schools that the students are applying to enter. This information allows students and particular teachers who also serve as guidance counselors to know where each student falls in the distribution. In addition, students can determine the likelihood of their passing a school's entrance examination by comparing their own scores with the average *hensachi* of students who passed the examination in the previous year.

LeTendre (1996) noted that the guidance process in junior high and high school is extensive. He explained that one group of junior high students he studied had taken 13 tests at school and four tests given by companies outside of school for placement purposes before high school entrance examinations had even begun. In addition, they had been surveyed about the high schools they desired to enter and received 35 handouts related to placement. Students in junior high and high school take part in many explanatory sessions and are exhorted to study hard and set goals as early as possible (p. 200).

McVeigh (2002) and others have pointed out that the Japanese Ministry of Education has not condoned the use of *hensachi* rankings or cram schools, and in 1993 sent an official notice to junior high schools directing them to stop using *hensachi* rankings produced by private testing companies. However, schools resisted and continued using the rankings. The Ministry eventually had to "admit defeat to the powerful and embedded forces of the unofficial educatio-examination regime" (p. 89) and accept the use of *hensachi* rankings, as well as the use of cram schools, particularly for students who have higher or lower than average scores. McVeigh also suggested that the Ministry might have agreed to acknowledge the role of cram schools because these more than 50,000 schools are an important "industry" for the Ministry of International Trade and Industry (p. 89).

McVeigh (2002) explained that high school students often use publications such as *Zenkoku daigaku juken nenkan* (National university examination yearbook) in order to compare scores and judge their chances of passing the entrance examination of their preferred university. In addition, admissions officials also routinely use published *hensachi* ranking information in evaluating applicants. For example, officials at the university where I work use published lists of *hensachi* rankings of high schools within Japan, in order to keep careful track of these rankings. These lists are available to the public on the Internet, and sites include http://hennsati.seesaa.net/article/47027227.html; http://kintaro.boy.jp; and http://momotaro.boy.jp. These *hensachi* rankings are regarded as indicative of the academic level of high schools and are regarded as an important criterion by not only admissions officials, but also many others involved in interviews and other tests for entrance examinations. In general, students who have only fair HSGPAs, but who come from high schools with high *hensachi* rankings, are still considered eligible candidates for admission. However, students who come from high schools that have lower *hensachi* rankings are generally expected to have higher HSGPAs.

As noted, *hensachi* rankings are considered to indicate the academic level of high schools; in other words, it is believed that high schools vary in terms of academic level and grading standards. Researchers in the United States have also long known that there are differences among high schools. As Betts and Morell (1999) pointed out, some of the differences are due to socioeconomic factors of the school neighborhood and to the level of teachers' experience (p. 288). Differences among high schools are also due to grading practices. Camara (1998) discussed, for example, the fact that high school teachers have much flexibility in determining grades. He also pointed out that teachers tend to grade high and lower ability students differently; they typically grade students with greater ability on achievement, and students with less ability on effort (p. 1).

Linn (1966) explained that awareness of differences in grading practices among high schools has led researchers to suggest using high school class rank instead of GPA, or to adjust HSGPAs to make them more comparable, and therefore more useful in predictive validity studies. However, Linn reported that attempts to use adjusted high school grades to predict college grades have resulted in "discouragingly small" improvements in prediction over unadjusted grades (p. 326). Linn said that a standard means of countering high school differences has been to include a standardized test as another predictor of achievement in university (p. 313). The SAT and ACT have been used in the United States in this manner for years; in Japan, the Center Examination is also now used, although not yet in a parallel manner. Another method noted by Linn is that admissions officials "have undoubtedly made rough adjustments” in grades of students from different high schools by remembering the performance of students from particular high schools (p. 313).

In Japan, adjustment for differences among high schools has come by way of *hensachi* rankings, as noted previously. For example, for the sample of students at the university under study, the correlation between HSGPA and *hensachi* rankings was approximately -.37, revealing the pattern in how students were chosen: students with lower HSGPAs were from high schools that had relatively higher *hensachi* rankings, and vice versa. These rankings are currently and widely used, but their value has not yet been established empirically. Newfields (2006) suggested that the *hensachi* rankings are not an adequate measure of student performance and questioned whether they had any worth in predicting students’ academic success. However, if these rankings are accurate in representing the academic level of high schools, they might predict university achievement. The predictive validity of the *hensachi* rankings is examined in the current study.

If results of this study determine that high school *hensachi* rankings predict university achievement of students, these rankings could be useful in selecting among applicants. However, admissions officials might object to using these rankings as criteria for admittance on the basis of what Selvin (1958) first called the “ecological fallacy” (p. 615). He and others before him, including Thorndike (1939) and Robinson (1950), pointed out that statistics derived from studying a group do not necessarily apply to individuals. Although there have been proposals to remedy the difficulty of using group data to make inferences about individuals (e.g., Firebaugh (1978)), officials must consider how to make the best use of these rankings, as well as discuss both the advisability and feasibility of obtaining the individual *hensachi* rankings that are not currently used in the application process.

English proficiency.

A number of measures of English proficiency could be useful in predicting the eventual academic success of Japanese university students in English-medium universities in Japan or abroad. However, many researchers have questioned whether English proficiency can actually predict academic performance. Nevertheless, as Bayliss and Raymond (2004) have noted, because the number of international students at universities has been increasing, researchers are again looking at whether any aspect of proficiency in an L2 might predict academic success (p. 30).

In the following section, I discuss English proficiency in general, and then discuss studies of a variety of proficiency measures, all of which highlight the relationship of English proficiency measures to academic success. Because the question of whether English proficiency scores predict academic performance or not is controversial, but also central to this study, I examine these studies in some detail and critique aspects of the studies that may have led to questionable findings.

English language proficiency was defined by Ellis (1985) as knowledge of the L2, or linguistic and communicative competence that is measured in comparison to the competence of a native speaker (p. 302). The two dimensions of proficiency, linguistic competence and communicative competence, referred to by Ellis above, appear to be analogous to Cummins' (1980) concepts of cognitive/academic language proficiency (CALP) and basic interpersonal communicative skills (BICS), respectively, which were discussed above. English proficiency appears to be related to intelligence, particularly in terms of performance in academic skills, such as reading, grammar, and vocabulary (Ellis, 1983, p. 111). The CALP dimension of proficiency would seem to have the most relevance to the sort of intelligence described by Ellis, and to the prediction of academic success.

Besides the complexities involved in defining proficiency, there are other issues to consider, as Brown (2004) pointed out. He explained that English language proficiency can be defined in a variety of ways, including purpose (p. 319). That is, English proficiency can be defined as the ability to handle North American academic English, or "Englishes of different countries, regions, cultures, religions, and even times" (p. 319). However it is defined, those who construct language tests have to consider the Englishes of the test takers and their local communities, the content of the test, the proctors, the raters and scorers, the community targeted by test decisions, the purpose of the decision, and the decision makers (p. 318). In short, the definition of proficiency can vary according to the purpose of the proficiency measure chosen and the English(es) included on that measure.

Just as defining proficiency is difficult, so is measuring it. For one thing, proficiency can be operationalized in more than one way. Researchers often equate it to scores on an in-house proficiency test, a writing sample, the TOEFL, or another well-known, standardized measure, as most did in the studies discussed below. In each case, the measure itself also must be assessed to determine that it is valid and reliable, and not biased towards any group of test takers. In addition, researchers have to consider whether to test proficiency in terms of productive or receptive skills, or both. In short, as Wakabayashi (2002) pointed out, measuring proficiency is difficult because of both test limitations and decisions about what sort of proficiency is necessary for a particular context (p. 635).

Many researchers believe that proficiency is a valuable predictor of academic success for ESL students. Considering Cummins' (1982) assertion that cognitive academic language proficiency takes 5 to 7 years to develop, it is plausible that students with more advanced proficiency, who would be able to focus more on their academic studies, and less on the language itself, would tend to be more academically successful. However, the studies reviewed suggest that the relationship between English proficiency and academic success appears to depend on a variety of factors, such as the type of proficiency measure used, the type of academic success measure used, the academic level of the student, and the student's major, to name a few.

Studies of English Proficiency Measures as Predictors of Academic Success

The measure of English proficiency most often used in the studies reviewed is the TOEFL test. As Rosenfeld, Oltman, and Sheppard (2004) pointed out, the main purpose of the TOEFL has always been to evaluate the English proficiency of students who want to study overseas, particularly in the United States and Canada. They noted that past studies have shown correlations of the TOEFL with overall language performance, interview scores, and writing samples, and that Educational Testing Service (ETS) researchers have continued to work on revising the TOEFL so that, among other goals, the test is able to give more information about the degree to which students will be able to handle English in an academic setting. To validate the TOEFL, these and other researchers have used the test as a predictor of criterion measures that are purported to measure the same construct (e.g., TOEFL scores and faculty ratings of the same ability). Rosenfeld et al. made the point that using TOEFL scores to predict course grades alone as an indication of the validity of the test would be inappropriate; at the same time, however, they acknowledged that proficiency in English is clearly a contributing factor in academic success (p. 98).

Although Japanese students applying to enter universities in Japan typically do not submit TOEFL scores, the TOEFL (or any other good test of English proficiency) is relevant in a number of ways. If scores on the TOEFL or other measures of English proficiency predict academic success for ESL students studying at English-medium universities in other countries, then it is likely that the

English proficiency of Japanese students applying to enter an English-medium university in Japan will predict their eventual academic success to a similar degree. Therefore, as noted above, studies related to this topic are explored in detail below. In order to clarify the picture, I first discuss the results of studies of undergraduate L2 students, and then graduate students. As I do so, I critique any points that might have led to questionable findings. Because studies vary widely, each is presented separately, and in chronological order.

Alderman (1981) discussed the possibility of a cut-off point in English proficiency that determined how well students were able to perform academically. That is, he claimed that English ability was a moderator variable in that the correlations between students' scores on academic aptitude tests in English and Spanish increased as their English proficiency increased. In this study for the Educational Testing Service, Alderman administered the SAT in English, a parallel measure of academic aptitude in Spanish called the PAA, and the TOEFL to 411 Puerto Rican secondary students whose L1 was Spanish. He found that a TOEFL score of over 500 points was the point at which students' scores on academic aptitude tests (of math and verbal abilities) in English began to match their scores on similar tests in their L1. Alderman said that the generalizability of his results might be limited, and that researchers should attempt to replicate these results with other populations.

Graham (1987) explored the question of whether the English language proficiency of ESL students attending an English-medium university (in primarily

American schools) could predict academic success. She summarized the research to that point in time, and presented studies showing a negative, mixed, and positive relationship between English proficiency and academic success. She pointed out that comparing these studies is difficult for three reasons. First, the definition of English proficiency often varies from study to study, sometimes meaning placement in a particular class or level, a score on a standardized test like the TOEFL, or a score on an in-house test. Second, although GPA is used most frequently as an indicator of academic success, it is imperfect for a number of reasons, including that it does not reflect how many courses students have taken, does not take into account “sympathy or goodwill grades” sometimes given to nonnative students, and it is a poor indicator in studies of graduate students, whose grade range is narrow. Third, researchers do not always make clear what the real significance of a finding is, over and above the correlation coefficient (p. 506). In fact, in reporting and interpreting the strength of correlation coefficients, the studies cited in Graham’s article often used different standards, rather than adhering to a consistent index, such as the one given by Field (2005), in which correlation coefficients of .10 are regarded as small, of .30, medium, and .50, large (p. 32). Comparing the conclusions of these studies is therefore not straightforward; the reader must carefully check the statistics underlying the conclusions.

After reviewing the studies, Graham concluded that researchers’ opinions were evenly divided about whether English proficiency predicted academic success or not (p. 512). Therefore, she said that studies of the relationship between English proficiency and academic success had not yet resulted in any definitive conclusions, especially in terms of admissions decisions, but that there may be a cut-off point at which those below that particular level of proficiency are significantly less able to achieve academic success (p. 505). She urged school administrations to determine what this minimum level might be for their own schools so that they can establish fair English proficiency requirements (p. 505). Like Alderman (1981), Graham regarded the cut-off point of language proficiency as an important area for further study.

Johnson (1988) studied 196 international undergraduates enrolled in the spring of 1986 at the University of Wisconsin at Green Bay in order to determine the relationship between the school's minimum TOEFL entrance requirement (500) and academic success (as measured by GPA), and between the minimum TOEFL requirement and credit hours earned. She correlated the overall mean of the pre-admission TOEFL (537) with the overall mean GPA (3.0) of these international undergraduates, not explaining why she did not correlate the entire ranges of both sets of scores. She found a statistically significant correlation ($r = .36$, $p < .01$). In addition, she compared the GPAs of students with TOEFL scores below 500 and above 500, to determine if 500 was a viable cut-off point of language proficiency. She found that students with TOEFL scores below 500 earned grades that were significantly lower than those with scores above 500, as shown by a z statistic ($z = -3.77$, $p < .01$) (p. 165). However, she cited some groups of students who were exceptions. Japanese students, for example, had a mean of 490 on the TOEFL, but their GPA was comparable to that of Malaysian students, whose mean was 553. It should be noted, though, that those results were based on data gathered from 12 Japanese students and 94 Malaysians. Comparisons with such small numbers may or may not reflect reality.

Johnson also found that TOEFL subscores correlated with GPA. Results for Structure and Written Expression and Vocabulary and Reading Comprehension were both statistically significant ($r = .43$, $p < .01$, and $r = .36$, $p < .01$, respectively), while correlations with Listening scores were not (p. 166). She suggested that these results highlighted the importance of reading to academic success.

Johnson also found the correlation between TOEFL scores and credits earned to be statistically significant ($r = .80$, $p < .01$); however, as she noted, this result was based on all international undergraduates enrolled, whose amount of time at the university and numbers of credits varied greatly. Therefore, she selected 27 students and divided them into the two levels of proficiency used earlier (above 500, and below 500). In general, she found that students who scored both below and above 500 performed adequately, but both had lower grades in courses that required heavy reading and writing work, compared to science labs, mathematics, or introductory courses. She also found that the lower proficiency group took a lighter than average course load and completed the courses they had signed up for. On the other hand, the higher group, though taking an average load (15 hours), tended to drop courses and receive incomplete grades, which Johnson said might have affected their overall academic performance. This statement, taken together with the fact that Johnson reported no correlation coefficient for this follow-up study, made it unclear whether lower or higher proficiency students actually earned more academic credits.

Johnson emphasized that lower proficiency students were significantly less successful in academics by concluding her article with a quotation from a university instructor illustrating the difficulties low proficiency students encountered in classes. The instructor wrote about two students who both deserved an F, but who were receiving Ds for “determination and effort” (p. 167). He noted that the students encountered great difficulty asking questions or understanding basic tasks, and that they had been repeatedly guilty of plagiarism. In conclusion, Johnson warned that academic work at universities required an adequate level of English proficiency (p. 168).

Light and Wan (1991) studied 56 Soviet undergraduates studying as exchange students for one year in the United States from 1988 to 1989. The researchers asked the students to complete a timed writing sample, the TOEFL, and a pre- and post-study questionnaire (asking for demographic information, as well as a variety of rating information). The final inferential results were produced with the 48 students who returned all questionnaires (p. 180). The results showed that the summer 1988 TOEFL scores correlated moderately with students’ GPA of fall of the same year ($r = .33$, $p < .05$); the TOEFL scores also correlated with the number of courses successfully completed ($r = .25$, $p < .05$) (p. 180). Therefore, English proficiency was significantly correlated with academic success, in terms of grades and the number of courses completed. In addition, total words written on the timed writing sample correlated significantly with both English proficiency as measured by the TOEFL ($r = .69$, $p < .0001$) and with academic achievement, as measured by GPA ($r = .26$, $p < .05$) (p. 180). The researchers also discovered from additional information that the students' HSGPA correlated significantly with home-country college grades, ($r = .74$, $p < .001$) (p. 182). These high school grades also correlated moderately with the U.S. college GPA ($r = .29$, $p < .05$); therefore, there appeared to be an important relationship between earlier and later grades. Finally, Light and Wan made the point that the academic performance of these students was "remarkable," despite the fact that they had a TOEFL mean score of only 462 (Range = 303-613) (pp. 183-184). They discussed the importance of motivation, confidence, and the social interaction between these and other students on campus.

Christopher (1993) investigated the relationship between the English proficiency of ESL students and academic success in order to determine whether a direct measure of English proficiency (which she equated with a writing sample) or an indirect measure (equated with objective tests, such as TOEFL) would better predict later academic success. She used English language proficiency test scores (simply labeled as scores from an "objective test" (p. 94)) and writing sample scores as the independent variables in her study of 55 ESL students in a Canadian university. The dependent variables were GPA for all courses combined, GPA for humanities courses (in which students presumably had limited background knowledge), GPA for science and math courses (subjects the students had previously studied), and an additional measure of success, average accumulated credits per semester (AACPS), or the number of classes that students had successfully finished each semester and academic year; these credits and grades accumulated over a period of three semesters, or one academic year (p. 76). Christopher maintained that because AACPS could include the aspect of time, it could measure academic success differently than GPA (p. 43). In other words, students who were able to complete more class hours than other students in the same amount of time were completing academic work more successfully.

Christopher conducted correlation and multiple regression analyses and found that neither type of measure significantly predicted GPA; however, the direct measures explained 18.2% of the variance in AACPS, and the indirect measures, 10.3%. Even though there were statistically significant results for both measures, Christopher focused on the direct measure result only, perhaps because only this finding confirmed one of her hypotheses. In conclusion, she said that such results suggest that these proficiency measures predict academic success in a useful, if limited way, and that other factors yet undetermined must also be taken into account when attempting to predict academic success (p. 62).

Ward (1999) studied 100 bilingual Hispanic students who were in grades six through twelve. Although these students were not yet enrolled in a university, the results from the study are applicable to this study. Particularly interested in how a first language can benefit second language acquisition, Ward administered language proficiency tests in both Spanish and English, and obtained GPAs and English achievement test scores from the Iowa Tests of Basic Skills/Educational Development. She conducted correlational analyses with the data and reported a variety of findings, including the following: There was a statistically significant relationship between all levels of the English proficiency test and all of the achievement tests of the Iowa Tests (ranging from $r = .57$ to $r = .84$, all at $p = .0001$); there was a significant relationship between all levels of the English proficiency test and GPA in all subjects (ranging from $r = .33$ to $r = .64$, all significant at $p < .001$); and there were statistically significant correlations between students' Spanish proficiency test scores in reading and writing, and their English achievement test scores and GPA (pp. 340-342). Though Ward criticized GPAs as having low reliability because they are affected by teacher bias and student effort, these results suggest that students who are more proficient in English, and in L1 reading and writing, tend to reach higher levels of achievement in their school work and to have higher GPAs.

Al-Masawi and Al-Ansari (1999) tested 86 first- and second-year students majoring in English Language and Literature at the University of Bahrain in order to determine whether the TOEFL or the FCE (First Certificate of English test) was a better predictor of academic success, as measured by overall GPA and GPA in English courses. They used correlation and stepwise multiple regression to attempt to answer the question. Although they reported that all correlations of GPA and English GPA with the FCE scores were of "high magnitude" and that these results did not hold for the TOEFL, the results did not bear out their claim (p. 394). The results were similar for both tests. The correlations between GPA and TOEFL section subscores were .34 (listening comprehension), .51 (structure and written expression) and .42 (vocabulary and reading comprehension); the correlations between the seven FCE subscores and GPA were between .30 and .69 (p. 394). The average for the TOEFL section tests correlations was .42, and the average for the FCE section test correlations was .46. As for the correlations between English GPA and TOEFL subsections, they found the following: .57 (listening comprehension); .66 (structure and written expression); and .55 (vocabulary and reading comprehension). The correlations between the seven FCE subscores and English GPA were between .38 and .84; the highest correlations were with cloze and sentence transformation sections of the FCE. The average for the TOEFL section test correlations was .59, and the average for the FCE test correlations was .70. Although the FCE average was higher, the average correlations for sections on both tests with the English GPA were large, according to the criteria cited earlier from Field (2005). In addition, despite the researchers' claim, a number of the FCE correlations did not qualify as large. That is, 57% of the FCE correlations with GPA, and 29% of the correlations with English GPA were below .50.

Al-Masawi and Al-Ansari said that the stepwise multiple regression results indicated that the only significant predictors of both GPA and English GPA were two subscores of the FCE, the cloze and sentence transformation sections. In addition, the researchers reported that scores for Section Two (structure and written expression) of the TOEFL were also good predictors of students' English GPA (p. 397). However, the statistics they reported as support were problematic for three reasons. First, they used stepwise regression, but this choice is not the best if there is previous research available about the topic under study. According to Field (2005), the stepwise method should only be used when there is no literature available on the topic from which to make a hypothesis about the results. Second, they used the term, "Partial r-square" in their regression tables (p. 395). This term was confusing because one typically expects to see either R^2 or partial correlation, not a combination of these terms (Green & Salkind, 2005, p. 288). Given the context of a stepwise regression analysis, they might have intended to report semi-partial correlation (Field, 2005, p. 160). Finally, the predictors they described as "good" (except for the FCE subsection, sentence transformation) were not strong; for example, the structure and written expression scores of the TOEFL accounted for only 3% of the variance in English GPA, and the cloze subsection of the FCE accounted for 6% of the variation in overall GPA (p. 397). On the other hand, they were probably more correct in saying that sentence transformation predicted GPA and English GPA well (with both reported as partial r-square at .42 and .63, respectively) (p. 395). Even so, the type of analysis used, confusion in terminology, and apparent misinterpretations reduced the credibility of the findings.

Al-Musawi and Al-Ansari used the results to argue that the FCE was a better predictor of academic success than the TOEFL for the students at their university, and therefore, for students studying English as a foreign language. They argued for increased teaching of grammar, and for the use of FCE despite its relatively low reliability. Although the results were mixed and at times unclear, the findings showed that English proficiency was linked to academic success, as measured by GPA and English GPA. In addition, results of this study indicated that grammar or sentence form test scores might predict academic success.

Bayliss and Raymond (2004) reported the results of two studies—one with undergraduates and one with graduate students. The undergraduate study is discussed in this section. The researchers studied 100 L1 and 36 L2 speakers of French who had completed high school and one previous year of college, as well as a French as a second language (FSL) test (developed at the University of Ottawa) before entering a three-year program in civil law in a Canadian college. Bayliss and Raymond wanted to know to what extent the FSL test results, as well as an admission GPA (from grades from one previous year of college) predicted academic success of the L1 and L2 French speakers, as measured by first-semester GPA and number of courses failed. They found that all parts of the FSL test (listening, dictation, reading comprehension, error correction, and composition) correlated significantly with first-semester GPA (r = .26 to .55) and with number of courses failed (r = -.24 to -.41) for the L1 French speakers; however, there was one exception—dictation and number of courses failed did not correlate significantly. For the L2 French speakers, the only significant correlations were between reading comprehension and first-semester GPA (r = .37). In other words, the better the reading comprehension skill of L2 French-speaking students, the higher their GPA was.

Bayliss and Raymond also performed a stepwise regression analysis in order to predict GPA and number of courses failed, from scores on the FSL test and one year of previous college grades. As noted, this method of regression is not the best choice when there is literature available on the topic; because they cited many relevant studies in their review of research, they probably should have formulated hypotheses and used a different type of regression, as Field (2005) recommended.

The researchers found that the reading comprehension score of the FSL test was the best predictor of GPA for all students ($R^2 = .25$, $p < .00$) and for L2 students ($R^2 = .14$, $p < .02$). In addition, reading comprehension scores predicted the number of courses failed by all students ($R^2 = .20$, $p < .00$) and by the L1 students ($R^2 = .21$, $p < .00$) (p. 42). They stated that the admission GPA (from one year of previous college courses) was not a predictor "of any substance" for GPA or for number of courses failed for all students because admission GPA accounted for just 1% of the variance (p. 41).

However, exactly how the researchers derived the percent of variance and their conclusions about admission GPA was unclear. As Green and Salkind (2005) noted, when discussing relative contributions of individual predictors, one should report statistics, such as bivariate correlations, partial correlations, and standardized regression coefficients; in addition, when reporting whether individual variables contributed significantly to the regression equation, one should report *t* values (p.

295). Though Bayliss and Raymond did not report partial correlations, they reported a statistically significant *t* value for admission GPA ($t = 2.23$, $p < .03$), and an R^2 value which indicated that admission GPA was a statistically significant predictor of GPA that accounted for an additional 3% of the variance in GPA (p. 42). These statistics indicate that, though admission GPA was not a strong predictor of GPA, it was statistically significant and accounted for 3% (not 1%) of the variance in GPA. As Zwick (2002) pointed out, measures that add even small increments in predictions can be "worthwhile" (p. 86). Because these results were not reported accurately, and other results were omitted by the researchers, there might be other problems in the discussion and interpretation of study results.

In conclusion, Bayliss and Raymond studied L1 French speakers and L2 English speakers who were studying in a French-medium program in Canada. They found a variety of results regarding the L1 group, the L2 group, and for all students. The primary finding they reported for the L2 students was that language proficiency, particularly reading comprehension, played a significant part in predicting academic success in terms of GPA.

At the undergraduate level, the L2 proficiency of second-language learners appears to be a variable of continuing interest for researchers. Though results are mixed, most studies suggest that L2 proficiency is linked to academic success in a number of ways. Researchers continue to be particularly interested in whether L2 proficiency predicts academic success or not, and whether there might be a minimum level, or cut-off point, of proficiency required for success.

In studies of international graduate students, English language proficiency appears to play a less important role in predicting academic success for a number of reasons. In the following studies, most of the graduate students had achieved high scores on the TOEFL (usually above the minimum 550 required for entrance). Therefore, English language proficiency probably affected their academic success less than other factors, such as knowledge of their specialty.

However, even if L2 proficiency predicts academic success for international graduate students, it might be difficult for researchers to demonstrate the effect. That is, if researchers use TOEFL scores as an independent variable and GPA as a dependent variable, the restricted range of scores (usually 550 and above for the TOEFL, and between 3.0 and 4.0 for GPA) would reduce correlations. In fact, such low correlations are reported in a number of the following studies. Nevertheless, other studies show that English proficiency continues to play a role of some importance, even for graduate students. As was true in the case of undergraduates, studies of graduate students vary widely; therefore, each is presented separately, and in chronological order. However, the study by Hale, Stansfield, and Duran (1984) is presented last because it was unlike the other studies in being a summary of TOEFL studies. As I discuss the following research, I continue to critique any points that might have led to questionable findings.

Hwang and Dizney (1970) used correlational analysis to investigate the degree to which total TOEFL scores predicted ESL course grades and first-semester GPA for 63 Chinese graduate students in their first term of study (in a variety of fields) at the University of Oregon. The researchers found a statistically significant correlation ($r = .66$, $p < .05$) between total TOEFL scores and ESL grades for the 20 students who were required to take these courses. However, they reported that none of the five correlations between total TOEFL scores and first-term GPA (in five majors) was statistically significant. Despite this claim, the correlation table they presented included one correlation of .69 between total TOEFL scores and the first-term GPA of students studying architecture. A correlation of .69 is large, according to the index cited by Field above (p. 476); therefore, the researchers made at least one oversight in reporting these statistics. In addition, Hwang and Dizney also commented that their sample was very small and perhaps not representative of other populations. Given these two concerns, the results might reveal little about the degree to which TOEFL scores predict graduate school success.

Sharon (1972) studied the admissions data, including TOEFL and Graduate Record Examination (GRE) scores, of 975 international graduate students from 24 American universities, in order to determine whether TOEFL scores significantly added to the predictive validity of the GRE Verbal test (GRE-V) (p. 426). He hypothesized that the TOEFL scores would moderate the relationship between the GRE-V and graduate school GPA in that prediction of graduate school grades would be more accurate for students with high TOEFL scores than for those with relatively low scores. To perform his analysis, he divided the students into three major fields: engineering, technology, and mathematics; natural sciences; and other

(p. 429). Sharon also adjusted the graduate school GPA for differences in grading (p. 428). Although he used regression to make this adjustment, the process was only generally described. After discussing the resulting predictions, Sharon concluded that the TOEFL might work as a moderator of the GRE in predicting the GPA of international graduate students, but that his hypothesis that prediction of GPA would be better for students who had earned high TOEFL scores than for those who had low TOEFL scores was only partially supported by the results (p. 431).

Ayers and Peters (1977) studied the records of 50 male Asians who had completed graduate programs in engineering, math, and chemistry in order to determine the relationship between pre-admission TOEFL scores and Graduate Record Examination (GRE) Verbal and Quantitative scores, and the students' final overall GPA. They found that the TOEFL and GRE Verbal and Quantitative scores were correlated significantly and highly ($r = .76$, $p < .01$, and $r = .64$, $p < .01$, respectively). However, this correlation between the TOEFL and GRE scores was based on a sample of only 15 students because only a portion of the total group had completed the GRE; thus, the generalizability of these results is limited. They also found that TOEFL scores were correlated significantly with GPA ($r = .40$, $p < .01$) and that students with TOEFL scores one-half of a standard deviation above the mean had significantly higher GPAs than students with TOEFL scores one-half standard deviation below the mean. Though they did not refer specifically to a cut-off point of language proficiency or discuss this result at any length, they were apparently trying to establish a general cut-off point. In addition, they conducted a stepwise multiple regression analysis using scores from the TOEFL and the GRE Verbal section and produced a regression equation that led them to conclude that the TOEFL can be used to predict the success of Asian students studying particular subject areas (p. 463).

Light, Xu, and Mossop (1987), in a study frequently cited in the literature, investigated 367 international graduate students majoring in a variety of subjects to determine the relationship between TOEFL scores and first-semester GPA and credits earned. They looked at correlations between pre-admission TOEFL scores, first-semester GPA, and graduate credits earned in the first semester, and were especially interested in the relationship between a TOEFL score of 550 (the university entrance requirement) and first semester GPA. They reported that overall TOEFL and GPA in the first semester correlated weakly ($r = .14$, $p < .05$), as did overall TOEFL and credits earned in the first semester ($r = .19$, $p < .02$) (p. 255).

In their explanation of the results, they said that the correlation between TOEFL and GPA, though statistically significant, was too small to be practically important and that a student’s TOEFL score reveals almost nothing that allows researchers to predict academic achievement. They concluded that the TOEFL did not effectively predict academic success, as measured by the GPA of these students in their first semester of study; however, they stated that the correlation between TOEFL and graduate credits earned was significant and important (p. 258).

If one examines the above conclusions and information from the study, questions arise. First of all, the authors' interpretation of correlation coefficients was inconsistent. Though they stated that a correlation of .14 (for TOEFL and GPA) was unimportant, they interpreted the correlation of .19 (between TOEFL and graduate credits earned) as meaningful. Both of these would be classified as small, according to the criteria by Field cited earlier. In addition, if the correlation coefficient is squared, one can determine the percent of variance in the dependent variable that is explained by the independent variable (Field, 2005, pp. 32-33). In this case, r^2 for each was .0196 and .0361, respectively. In other words, TOEFL scores accounted for 1.96% of the variance in GPA and 3.61% of the variance in graduate credits earned; alternatively, 98.04% of the variance in GPA was due to factors other than TOEFL scores, and 96.39% of the variance in graduate credits earned was also due to other factors. These effect sizes were similarly low, yet the researchers interpreted them quite differently.

In addition, Light, Xu, and Mossop determined through a *t*-test that specially admitted students with TOEFL scores below the cut-off score of 550 did not earn grades that were significantly lower than students with TOEFL scores over 550 (p. 257). However, by dividing all students into five TOEFL score ranges and using analysis of variance to examine differences among the groups, they found that there were significant differences among the five groups, although they did not use follow-up tests to determine where the group differences were. They also did not explain the rationale for the uneven TOEFL score ranges that they used for the analysis.

The study by Light, Xu, and Mossop put forth and answered interesting questions. They also made valuable recommendations, such as encouraging further research into other criteria for academic success, into students' own views, and into other variables, such as attitudes and motivation (p. 259). However, questions about a number of their conclusions reduce the credibility of the study.

Ayers and Quattlebaum (1992) investigated relationships among pre-admission TOEFL scores, GRE scores, and final GPA in their study of 67 Asian graduate students majoring in engineering (60 male and seven female students), with the primary purpose of determining whether academic success could be predicted by TOEFL or GRE scores. Though they found that TOEFL scores correlated significantly with all sections of the GRE (with Verbal, $r = .63$; Quantitative, $r = .30$; and Analytic, $r = .35$, all at $p < .01$), the correlation with GPA was not significant, at $r = .05$. Only the Quantitative Section of the GRE correlated significantly with GPA ($r = .32$, $p < .01$). Despite finding no significant correlation between TOEFL scores and GPA, the writers concluded that the TOEFL could still be considered useful for screening applicants for a minimum level of communication skills needed for work in an American university (p. 975). They also recommended studying other variables, such as motivation and students' background, as they might hold more importance for admissions (p. 975). However, there is no evidence in the study linked to these summary statements.

In considering the very low correlation between TOEFL scores and GPA found in this study, it is important to note that an unspecified number of students in this study came from India, and the others were described as native speakers of Chinese. This fact raises two potentially important issues. First, many people from India are extremely proficient in English, and can be classified as native speakers of English. Second, the Chinese speakers could have come from Hong Kong where many people are bilingual. In addition, the students' average TOEFL score was not reported. Thus, it is possible that these students had uniformly high scores. If so, it is not surprising that the correlation coefficients were so low. As noted earlier, a restricted range in one or more of the variables results in falsely low correlations. In short, the low correlation found in this study might have revealed little about the real relationship between English language proficiency and GPA for ESL students.

Vinke and Jochems (1993) studied 90 Indonesian engineers in a one-year postgraduate engineering course conducted in English in the Netherlands to determine the relationship between pre-admission TOEFL scores and academic success, which they operationalized as the average of seven examinations that were required to complete the program. In addition, they examined whether age affected students' academic success. They found that younger students tended to be more academically successful in their program and that TOEFL scores correlated strongly with academic success ($r = .51$, $p < .002$). However, the correlation varied by proficiency level: for students with TOEFL scores below 450, the correlation was only .12, and nonsignificant; for students with scores above 450, the correlation was the same as the overall result ($r = .51$, $p < .002$). From these results, the authors concluded that the lack of English proficiency of the students whose TOEFL scores were below 450 created a “real impediment” to their academic endeavors (p. 282).

Vinke and Jochems concluded that the overall correlation between TOEFL scores and academic success was “limited,” even though the correlation was statistically significant, and TOEFL scores accounted for approximately 26% of the variance in academic success (p. 283). Rather than stressing the significance of the findings, they focused more on explaining why TOEFL scores had not explained more of the variance in academic success. They suggested that: (a) the TOEFL might not be testing the language skills needed in an English-medium university, as Bachman (1991) had proposed, and (b) other non-academic factors, besides proficiency, were affecting academic success.

Like Alderman (1981), Graham (1987), and Johnson (1988), Vinke and Hochems believed that the differential results for low and high proficiency students were evidence of a cut-off point in English proficiency, above which students had a significantly better chance of being successful academically (p. 275). Using a *t*-test, they found that students who scored about 450 and above on the TOEFL had significantly higher grades than students who scored below 450, but they said that the specific cut-off point or preferable range of TOEFL scores for their own group of students still needed to be determined. They suggested that such a range would probably vary, depending on major or school, and recommended that schools establish these cut-off points as admission requirements for their own programs.

Bayliss and Raymond (2004) investigated whether an advanced ESL test could predict later academic success, as measured by GPA, upon completion of the MBA program (p. 29), and whether any subtest scores of the test were especially important predictors of academic success. Although this study was divided into several parts and involved other individuals, the participants of interest here were 34 Chinese students enrolled in an MBA program at a Canadian university. The advanced ESL test used by Bayliss and Raymond was the CanTEST, a communicative test with six subsections that students were given in China, one year before entering the university. The students took the CanTEST again in April, 2000, when those who were required to take the EAP program began at the university, and again in November, 2000, when the students were finishing their first MBA courses in a "sheltered" period in which only Chinese students were enrolled in classes.

Bayliss and Raymond found the following statistically significant correlations between CanTEST subsections and GPA: (a) listening subtest scores of 22 students who took the April test ($r = .49$, $p < .05$), and (b) reading comprehension and cloze subtest scores of 34 students who took the November test ($r = .62$, $p < .001$, and $r = .34$, $p < .05$, respectively). They pointed out that the correlation of .62 between reading comprehension and GPA meant that 38% of the variance in the GPA was accounted for by the reading comprehension score; in addition, they stated that a regression analysis they performed showed that only this November reading comprehension score was a significant predictor of GPA. (The authors did not specify the type of regression analysis used, nor did they provide a table of results.) Based on these findings, Bayliss and Raymond concluded that reading skills and overall English proficiency appear to be important factors in predicting academic success in the MBA program they studied.

Though the above conclusions might be correct, a number of points are problematic. First, the fact that the researchers used CanTEST scores obtained in November, after the students had completed the EAP program and a semester of graduate work, as predictors of final GPA is problematic. These test scores did not represent the proficiency level of applicants; they represented the proficiency level of students already enrolled in the program. Therefore, these November test scores should not have been used as predictors in the regression analysis; in other words, results with these scores did not demonstrate whether the English proficiency of applicants was valuable in predicting their future academic success or not. Another questionable point was the regression itself. The researchers did not discuss whether they had met the assumptions of regression, nor did they present relevant tables. In addition, the number of students (34 who had completed EAP courses, and 54 total students) was quite small. According to Field (2005), researchers who want to test the predictors (as was the case in this study) should use the following formula to determine the minimum number of participants needed: $104 + k$, where k is the number of predictors (p. 173). In this study, then, at least 106 participants should have been included in the study because the independent variables consisted of two CanTEST sections, the April listening subtest score and the November reading comprehension subtest score. These issues diminish the credibility of the results.

The study by Hale, Stansfield, and Duran (1984) was unlike the above studies, in that the authors summarized 82 studies published between 1963 and 1982, all of which focused on the TOEFL test. Because their work was done for the Educational Testing Service, they pointed out that the summaries were not intended as evaluations and that the summaries presented the point of view and interpretations of the original writers, not those of Hale et al.; in addition, they stated that because they were employed by the publishers of the TOEFL test, it would have been inappropriate to evaluate the studies (p. 7). Therefore, the authors did not argue for any particular claim about the TOEFL.

However, a number of researchers writing about the TOEFL cited this study quite differently. For example, Ayers and Quattlebaum (1992) said that Hale, Stansfield, and Duncan had summarized “over 100 studies” and their review of the research had caused them to doubt whether the TOEFL was valid for predicting graduate school success (p. 973). In addition, Light, Xu, and Mossop (1987) wrote about this study in a similar manner. Unfortunately, other authorities writing about the TOEFL followed suit and made the same mistake. Certainly, when researchers misrepresent sources in such a way, they reveal carelessness, at best, or dishonesty, at worst. In either case, the misrepresentations are worrying and cast doubt on their own research about the TOEFL.

It is difficult to derive firm conclusions from the above studies because they are not strictly comparable in many ways. The gender, age, level of study, and academic major of the participants vary according to the study. Different numbers of participants are involved, and the size of the sample is often too small for the type of analysis used. Different measures of proficiency and academic success are used, and different kinds of analyses are employed, from correlation to regression. But even more problematic than these differences in design and analysis are the varying ways that researchers have used statistical techniques, reported information, and interpreted their results. As noted above, they sometimes did not adhere to a standard way of interpreting findings (such as correlation coefficients), or they focused on one finding to the exclusion of another that was equally significant. These problems seem linked to researcher bias, and serve to undermine the credibility of a number of the above studies.

However, despite the probable weaknesses of the studies reviewed and the difficulty involved in comparing them, they are worth reviewing because they report findings about a broad range of variables. For example, English proficiency was operationalized as scores from the TOEFL, TOEFL section scores, reading tests, writing samples, as well as scores from other tests, such as the GRE and tests created by individual universities. These variables provide a valuable basis for planning further studies.

The above studies are also useful because findings about a variety of relationships among variables are reported. For example, for undergraduate students, English proficiency (as measured by TOEFL) was often significantly correlated with academic success, and all proficiency measures were correlated with the number of credits earned. For graduate students, the results were more mixed. At times, proficiency, as measured by TOEFL, was significantly correlated with measures of academic success and with credits earned (when studied). In other studies, particularly those of graduate students of advanced English proficiency, the limited range of proficiency scores and GPA tended to reduce or eliminate the predictive power of English proficiency. Though one-to-one comparisons between studies are not possible because of design differences, most results have indicated that English proficiency predicts later academic success of international students in an English-medium university, to some degree.

Finally, as noted above, the TOEFL is not usually directly relevant to Japanese high school graduates applying to Japanese universities because many of these students have not taken an English proficiency test; if they have, the test is usually not the TOEFL. However, findings from studies concerning the TOEFL are relevant to the present study. Because English proficiency was often a significant predictor of academic success in the above studies, it is likely that proficiency also predicts the academic achievement of undergraduates of English-medium universities in Japan. In addition, officials at the university under study have considered requiring students to submit TOEFL scores in the admission application.

Therefore, examining the degree to which students' TOEFL scores (obtained in April of their first year at the university) predict later academic performance could inform not only the current study, but also the admissions process at this particular university.

Non-Cognitive Variables

As pointed out in the Introduction chapter, non-cognitive variables include a wide range of personal qualities, attitudes, behaviors, background factors, and conditions, all of which can affect the way people live their lives, and the degree of success they attain.

Using Non-Cognitive Variables to Predict University Success

Though cognitive variables have consistently predicted academic achievement, the prediction is far from complete. As Cutrona, Cole, Colangelo, Assouline, and Russell (1994) noted, somewhere between 60 to 70% of the variance in academic achievement has not yet been explained. Therefore, it is likely that studying non-cognitive variables can lead to a better understanding of factors that contribute to academic success of students. Two often-studied non-cognitive variables are examined here: gender and family background (in terms of parents' education, one component of socioeconomic status).

Gender.

Gender has been shown to be related to success in school in a number of ways. Researchers have noted that women generally perform better than men in school. For example, Perez (2004) stated that female students tend to receive better high school and college grades than men. In a similar vein, Astin and Oseguera (2005) noted that women are more likely to continue on through college to attain their bachelor’s degrees. However, women tend to score lower on admissions tests, such as the SAT, and the result is that their grades in college are underpredicted by regression equations used by admission officials at universities. Young (2004) suggested reasons for lower test scores, and the subsequent underprediction of women’s university grades: women do less well than men on timed multiple-choice test items, women might be negatively affected by the roles assigned to women in test questions, and women might be handicapped by tests which assign a penalty for guessing because they are less likely to guess under these conditions (p. 296). Young claimed that the main reason for the underprediction of women’s college grades has to do with courses that men and women select. He said that a greater number of women take courses in the humanities and social sciences, while more men take engineering and natural science courses. Because grades are typically higher in the humanities and social sciences, women tend to receive higher grades, and men tend to receive lower grades than predicted by admissions criteria. Young said that the underprediction of women’s grades in college is mostly caused by such differences in course selection, but also noted that researchers and admissions officials should continue to study this underprediction, as it might continue to lead to bias against women in admissions decisions. Young recommended that admissions officials be aware of this underprediction, and consider making adjustments in favor of women.

Parents' education.

Socioeconomic status, especially as represented by parents' education and family income, has frequently been shown to be related to success in school. As Zwick (2004) pointed out, every measure of academic achievement is closely tied to educational level and income (p. 213). Income is related to quality of schooling, extracurricular activities, opportunities, home environment, and self-confidence. Holloway, Fuller, Hess, Azuma, Kashiwagi, and Gorman (1990) pointed out that children in Japan who come from poorer families often feel that they are not able to make good use of schooling, and their parents believe that the children are unable to compete with children from wealthier families (pp. 205-206). Family income must certainly impact students' academic achievement in Japan; however, university data related to income are usually regarded as personal and maintained in strict confidence, and therefore were not available for this study.

The educational level of the parents is another aspect of socioeconomic status that affects academic achievement. For example, Tinto (1975) maintained that students who continue on through college tend to have more educated parents with higher expectations for them than those who drop out (p. 100). In addition, the parents' educational level is predictive of their children's school grades. LeTendre (1996) said that the educational level of parents was significantly related to students' grades in his study (p. 207). Similarly, Jürges and Schneider (2004), in an international study of student achievement in mathematics, found a strong link between parents' level of education and their children's academic performance.

Because not all studies of parents' education report comparable statistical results, it is sometimes difficult to evaluate the impact of parents' education, relative to typically strong predictors such as HSGPA and test scores; however, the following two studies provide more information regarding the relative contribution of parents' education to the prediction of GPA. Zwick and Green (2007) reported average correlations of HSGPA and parents' education within high schools. The correlation between HSGPA and the educational level of fathers and mothers was .18 and .14, respectively (p. 15). Zwick, Brown, and Sklar (2004) also examined the relationship of parents' education (represented by the number of years of education of the more highly educated parent) with first-year university achievement. They reported regression analysis results for separate cohorts from 1996 to 1999 at seven separate University of California campuses. Regression coefficients for parents' education ranged from -.03 to .10; of the 29 coefficients reported, nine were statistically significant (p. 25). In reporting overall results, they said that adding parents' income and education increased the proportion of explained variance (R^2) by an average of .004 (p. 12).

Some researchers have found that one parent affects the children's academic achievement more than the other parent. For example, Astin and Oseguera (2005) found that the father's level of education was slightly more predictive of his children's finishing four- or six-year degrees, although the years of education of both parents were significant predictors. However, Travis and Kohli (1995) found that the father's level of education did not relate to the educational level that the children attained, but that the mother's education did; mothers who were more educated had more educated children (p. 504). In a similar vein, Nora, Barlow, and Crisp (2005) found that the education of mothers in low-income families strongly affected whether their children would return to college after the second year: if mothers had completed an undergraduate degree, their children were more likely (by 57%) to return for the third year of college than children whose mothers had not earned a degree (p. 133). These and other studies have demonstrated that the parents' level of education certainly impacts their children's academic performance in a number of ways. Therefore, it is likely that parents' education also predicts the academic achievement of Japanese university students, at least to some degree.

Gaps in the Literature

A number of challenges for educational researchers and admissions officers have been pointed out in the above review of the literature. In Japan, no studies have been devoted to predicting university success, based on HSGPA and other application information. Therefore, admissions officials in Japan are generally unaware of the potential value of such information. In addition, TOEFL test scores, though generally regarded as reliable and valuable by university officials in other countries, are not currently used for university admissions in Japan; therefore, it is unknown whether the TOEFL is useful in predicting success in an international university in Japan or not. Another kind of information that is available to admissions officials is high school *hensachi* rankings. No research has been conducted to determine whether these rankings are related to success in university or not. It is also unknown to what degree non-cognitive variables affect university success in Japan.

The literature review also reveals many gaps in knowledge about Japanese university entrance examinations; however, students continue to be granted access to universities in Japan on the basis of Center Examination and university entrance examination scores. Little has been written about the Center Examination in English, and what has been written about university entrance examinations reveals weaknesses. University faculty members assigned to construct the tests typically lack testing expertise, and the tests they write, not designed to measure any particular construct, appear to lack sufficient validity and reliability. Whether any of these examinations predicts university success or not is currently unknown.

The current study is designed to answer, or begin to answer, these questions, and it is hoped that the study results will not only contribute to a theoretical understanding of the larger picture of student success, but also be of practical value. If university officials were to learn more about the relative value of application information and other variables, and give more weight in the admissions process to the cognitive measures that have been shown to predict college success, their decisions would surely improve. In addition, they might be able to use their increased understanding about the importance of non-cognitive variables to identify potentially at-risk students and provide academic support.

Purposes of the Study

The primary aim of the present study is to evaluate both cognitive and non-cognitive variables in terms of their relative usefulness in predicting university success in Japan. The following variables, most of which are included in application information, will be examined in the first part of the study: HSGPA; high school grade factor scores; total and section scores of the ITP TOEFL; high school *hensachi* rankings; and the non-cognitive variables, gender and parents' education. Hierarchical multiple regression analyses are used to examine the degree to which each predicts success in the EAP and regular university programs.

The second purpose is to investigate to what degree Center Examination scores and entrance examination results used at the university under study predict college success. The Center Examination section scores, four types of university entrance examinations, and numerical scores from the examinations (including the English test, Japanese test, English essay, and Japanese and English interview scores) are examined in further hierarchical multiple regression analyses to determine how valuable each type and measure is for predicting success in the EAP and regular university programs.

The third purpose of the study is to construct a “big picture” of the ways in which variables interact over time to account for student success in both the EAP and regular university program. One aspect of academic success over time examined is timely and exemplary completion of program requirements. Therefore, logistic regression is used to identify variables that are most related to students completing the EAP program on time, graduating on time, and graduating with honors. A second aspect of the larger picture of academic success regards the ways in which variables interact over time, from before matriculation to the first year of university, and on to graduation. This “big picture” of academic success is examined through path analysis.

It is hoped that this study will inform admissions officials and provide educational researchers more information derived from regression and other statistical procedures that have not so far been used to understand university success in Japan. Each procedure provides a different perspective from which to view and better understand student success.

Research Questions

Research Question 1: To what extent do HSGPA, ITP TOEFL scores obtained in April, at the start of students’ first year at the university, and *hensachi* rankings predict EAP GPA, as well as first-year GPA and final GPA in the regular university program? Based on previous research, it is hypothesized that HSGPA and ITP TOEFL scores predict the students' performance in the EAP program, as well as first-year GPA and final GPA in the regular university program. In addition, based on research that suggests that level of high school predicts university achievement, it is hypothesized that *hensachi* rankings are not significantly related to the EAP GPA, but do predict first-year GPA and final GPA.

Research Question 2: To what extent do high school grade factor scores predict EAP GPA, as well as first-year GPA and final GPA in the regular university program? Even though the high school grades are unidimensional in the sense that they measure academic performance in similar ways, it is hypothesized that high school grades can be factored into three factors (representing more specific types of abilities), and that these factors vary in the extent to which they predict GPAs.

Research Question 3: To what extent do ITP TOEFL section scores predict EAP GPA, as well as first-year GPA in the regular university program? It is hypothesized that the section scores vary in the extent to which they predict GPAs.

Research Question 4: To what extent do the non-cognitive variables, gender and parental education predict HSGPA, EAP GPA, first-year GPA, and final GPA in the regular university program? It is hypothesized that gender and parental education both predict the students' HSGPA, as well as EAP GPA, first-year GPA, and final GPA.

Research Question 5: To what extent do Center Examination scores predict EAP GPA, as well as first-year GPA and final GPA in the regular university program? Based on the information obtained regarding the Center Examination, it is hypothesized that only the English section scores predict EAP GPA, first-year GPA, and final GPA.

Research Question 6: To what extent do type of entrance examination passed and entrance examination test scores (from the English test, Japanese test, English essay, and Japanese and English interview) predict EAP GPA, as well as first-year GPA and final GPA in the regular university program? Based on the research regarding the unsystematic manner in which these tests are constructed and the widespread criticism of entrance examinations, it is hypothesized that only English essay exams predict EAP GPA, first-year GPA, and final GPA.

Research Question 7: To what extent do HSGPA, high school grade factor scores, ITP TOEFL scores, *hensachi* rankings, and non-cognitive variables predict finishing the EAP program on time (FOT)? Based on previous research, it is hypothesized that HSGPA (and one or two, but not all, high school grade factor scores), TOEFL scores, and non-cognitive variables predict FOT.

Research Question 8: To what extent do HSGPA, high school grade factor scores, ITP TOEFL scores, *hensachi* rankings, and non-cognitive variables predict university graduation on time (GOT) (in four years) and graduation with honors (GWH)? Based on previous research, it is hypothesized that HSGPA (and one or two, but not all, high school grade factor scores), ITP TOEFL scores, and non-cognitive variables predict GOT and GWH.

Research Question 9: How do non-cognitive variables, HSGPA, ITP TOEFL scores, *hensachi* rankings, Center Examination scores, and university entrance examination scores interrelate in a path analysis and contribute to academic performance in the EAP program and in the first year of regular university study over time? It is hypothesized that: (a) father and mother's education contribute to performance in the HSGPA; (b) HSGPA contributes to EAP GPA and first-year GPA; (c) ITP TOEFL scores, Center Examination English section scores, and university examination English essay scores contribute to EAP GPA; (d) ITP TOEFL scores, EAP GPA, university examination English essay scores, and *hensachi* rankings contribute to academic performance in first-year GPA, and constitute parts of an acceptable path model.

CHAPTER 3

METHODS

Participants

The participants were members of five cohorts of students who enrolled from 2004 to 2008 in an English-medium university in eastern Japan. Of the 718 total participants, 217 were male, and 501 were female. When they entered as freshmen, they ranged in age from 18 to 33, though 95% were either 18 or 19. Participants were Japanese, and attended high school in Japan. Their English proficiency level, upon entrance to the university, ranged from a total score of 347 to 653, as measured by two TOEFL ITP tests administered in April, at the beginning of their freshman year (M = 472.23, SD = 39.43 (N = 701); M = 482.71, SD = 35.54 (N = 532), respectively). Participants were similar in terms of formal English education: They had previously studied English for six years. None of those studied were officially designated as "returnees," a common term for Japanese students returning from living abroad; however, through contact with students, I learned that some had lived overseas with their families, or studied abroad, usually for one year, in junior high or high school. The percentage of students with overseas experience appeared to be small though it was not possible to confirm this impression because complete university records were not available.

The university in question requires that entering students take the TOEFL for placement purposes and study in the EAP program for at least one semester, in order to pass the highest level of the program. Students with lower TOEFL scores might be required to study in the EAP program for two to three semesters. After completing the EAP program successfully, students receive a passing mark that allows them to first enroll in general education courses (which require approximately one year to complete), and later, in courses to complete a major in either International Relations or Global Business Studies. All courses are taught in English. Students are also required to spend one year of study abroad, usually in their sophomore or junior year. In many ways, these students are similar to ESL students studying in universities overseas though the fact that they live in their own country for most of their college life, and are able to use their native language freely outside of class means that they have a comparatively more comfortable and stress-free lifestyle. However, the school life of these students and ESL students is similar: They both must use English in class, the types of courses they study are quite similar, and the workload is substantial.

Permission to conduct this study was obtained from the President of the university, and data regarding cognitive variables (e.g., test scores, *hensachi* rankings, entrance examination scores, HSGPA, and semester GPAs), as well as gender, were obtained from the Admissions, Student Records, and Student Affairs offices. In addition, students were asked to complete a questionnaire regarding non-cognitive variables and were informed about the study orally and in writing. Directions on the questionnaire explained that answering the questions was entirely voluntary.

Materials

A description of seven measures used in the study follows.

High School Grade Point Average (HSGPA)

High school seniors in Japan typically take the following nine courses: Japanese, civics, history, math, science, health education, home economics, art, and foreign language (English, and very occasionally, another foreign language). Students can also take elective courses in religion or other areas, but the majority of students do not. Individual high school course grades in Japan are in the form of a number from 0 (the lowest possible grade) to 5 (the highest possible grade), and the HSGPA is calculated by multiplying each number by the number of hours of course credit, and then dividing by the total number of courses. This HSGPA is also used as a basis for designating an overall letter grade (A, B, C, etc.) on the student's transcript; this overall designated letter grade was not used in the current study.

Center Examination

As noted above, the Center Examination is the only standardized national entrance examination in Japan, and was originally modeled on the SAT. It includes five subject areas: civics, geography and history (100 points), Japanese language (200 points), foreign language (250 points for the English test, and 200 points for tests of other languages), science (100 points), and mathematics (100 points). Students who apply for general admission to the university under study are required to submit Center Examination scores, but the type of scores required varies, depending on the type of entrance examination. For the “A” examination, they submit Japanese, English, and one other section score of their choice; for the “B” examination and the “Gap Year Examination” (for students who matriculate in September, rather than April), they submit all five section scores; and for the “C” examination, they submit only English scores. The Admissions Office examinations (AO I and AO II) do not require Center Examination scores. In addition, the Admissions Office at the university under study alters the Japanese and English test section scores of the Center Examination, in order to balance score totals between entrance examinations and to emphasize English skills: the score on the Japanese test section is halved, and therefore worth 100 points, and the score on the English test section is reduced proportionately from 250 points to 200 points.

There has been little research on the examination; therefore, little is known about its validity and reliability. As of the date of this study, the National Center for University Entrance Examinations (at www.dnc.ac.jp) has not reported any relevant information or statistics in English. However, as noted previously, Ito (2005) reported that the English section of the Center Examination had a reliability of .82 (p. 102), and stated that another Japanese researcher had reported reliability coefficients between .94 and .96 (p. 93). As for validity, Ito reported that all but one section of the English test showed “satisfactory validity” (p. 91) based on his analysis; however, much more study is needed in order to make confident reliability and validity claims about the test.

Because only the classes of 2005 to 2008 submitted Center Examination scores, this variable is used in a pullout study after the initial regression analysis. In addition, because the English section test of the Center Examination was changed in 2006 in order to include a listening component, results are given for 2005 to 2008, as well as 2006 to 2008, in order to determine whether the addition of the listening component increased the predictive validity of the English section test.

University Entrance Examinations

The university in question administers a number of entrance examinations, each with somewhat different requirements. If students fail one entrance examination, they may take one or more later tests. The general admission examinations are the A, B, and C examinations, which are conducted in February and March. The February tests (types A and B) are tests of Japanese and English reading, writing, and translation. The A and B examinations are not multiple-choice tests, and the content tends to vary somewhat from test to test. However, in general, the Japanese test typically includes a one- to two-page reading in Japanese, often including graphs and tables. On this test, students might be asked to give a reading response of about 500 Japanese characters (about one-half a page), or they might be directed to give three or four shorter responses to the reading, sometimes including translations of particular sections into English. The English test includes approximately four pages of readings, about which students write answers to short comprehension and translation questions, as well as an essay of about 200 words.

One or two of the A and B examinations I reviewed included only one question requiring students to write a reading response of more than 250 words. In other words, in some cases, the A and B examinations are short essay tests.

The third type of general admissions entrance examination is the C examination given in March. This examination requires students to write an essay in English of 300 to 500 words in response to a paragraph-long prompt in English. The Gap Year Examination is also conducted in March for students who want to enter the university in September, rather than April; this examination requires students to participate in an interview conducted in Japanese and English.

Other than the general admissions examinations, the university conducts Admissions Office examinations (AO I and AO II). These examinations are intended for special students, such as principal-recommended students, transfer students, and working adults. The AO I examination is conducted in November, and most students who take it are either transfer students, working adults, or principal-recommended students. The smaller-scale AO II is conducted in August, and is principally offered for returnee students. Both AO examinations require students to write an English essay of 300 to 500 words in response to a reading prompt of less than one page and participate in an interview in both Japanese and English. For the current study, sufficient data were available for the A, B, and AO examinations.

Ultimately, these examinations are scored pass/fail, but numeric scores are also produced for: (a) the English reading, writing, and translation test scores (from the A and B examinations), (b) the Japanese reading, writing, and translation test scores (from the A and B examinations), (c) English essay scores (from the C, AO I, and AO II examinations), and (d) interview scores (from the AO I, AO II, and Gap Year examinations). Each of these is scored in a particular way, and the scoring method varies at times according to the entrance examination type. In addition, the scoring allocation is published in brochures for the general admissions examinations, but not for the AO examinations. Procedures in constructing and scoring the AO examinations are considered secret; therefore, information given here is accordingly general.

As explained, scores from the reading, writing, and translation tests in both English and Japanese are derived from the A and B entrance examinations. How these tests are produced is kept in the strictest confidence. Even the admissions official I spoke with did not know which faculty members had written these tests in that year; she said that the reason for such confidentiality is to protect both test writers and test content. The faculty members who wrote the tests and senior admissions officials are responsible for scoring the tests; the maximum score for the English test is 200 points, and for the Japanese test, 100 points.

In order to produce essay scores, appointed faculty members from the EAP department rate essays, and the scores are then converted by the Director to either 100 or 200 points, depending on the type of examination. On the day of the entrance examination, applicants are given an English essay to read and then prompted to write a reaction to a question about the essay. Completed essays that meet word requirements are then rated by EAP faculty members, after a grade norming session. Two faculty members rate each essay for content and conventions. A third rater may also rate the essay, in the case that the first two ratings differ by more than a designated amount. Ratings are then converted by the Director to either a 100- or 200-point scale, depending on the examination type.

Interview scores are derived from fifteen to twenty-minute interviews conducted by three faculty members. The initial and final segments are conducted in Japanese, but each candidate must also answer questions in English at a designated point in the interview. After each interview, the faculty members confer and produce a rating, which is noted by the lead interviewer. The lead interviewer is responsible for recording ratings and relaying final decisions to admissions officials.

In order to use entrance examinations as variables in this study, type of examination passed (type A, B, AO I, or AO II) was coded using dummy codes. Numeric scores were also studied, depending on the number of each available. Scores used in the study included: English test scores from the A and B examinations (based on a 200-point scale); Japanese test scores from the A and B examinations (based on a 100-point scale); English essays from the AO I and AO II examinations (based on a 100-point scale); and interview scores from the AO I and AO II examinations (based on a 100-point scale).

The entrance examinations taken by the class that entered in 2004 were not equivalent to the current examinations, according to Admissions Office staff members. They reported that students did not write essays, participate in interviews, or submit Center Examination scores in 2004; these became requirements for the various entrance examinations, as previously described, from 2005 on. Accordingly, entrance examination type and scores are not used in any investigation that includes the class that matriculated in 2004. As is typical for entrance examinations in Japan, no information about reliability or validity has been established for the A, B, AO I, or AO II examinations.

***Hensachi* Rankings**

As previously explained, these rankings are considered an index of the academic level of high schools (and therefore, of the students enrolled there), and university admissions officials consult published lists on the Internet in order to obtain and record the *hensachi* rankings for the high schools of applicants to the university. Officials explained that students who have relatively low HSGPAs might be admitted to the university if they come from high schools with relatively high *hensachi* rankings, and that students who come from schools with lower *hensachi* rankings must have relatively high HSGPAs in order to be admitted. The sites primarily used by officials at the university under study included two: http://hennsati.seesaa.net/article/47027227.html and http://kintaro.boy.jp. Officials also used rankings from http://momotaro.boy.jp on the rare occasions when high schools were not listed on the first two sites. *Hensachi* rankings are practically identical on these three sites, but occasionally vary by one or two points. The rankings can range from 20 to 80, though most fall within the range of 30 to 70; the mean value is 50, and one standard deviation is 10 (Newfields, 2006, December).

ITP TOEFL

The ITP TOEFL is used by educational institutions for program decisions, such as placement and exit. It includes listening comprehension, structure and written expression, and reading comprehension sections, all in a multiple-choice format. According to the official website (http://www.ets.org) the range of possible scores for each of the three section tests and for total score follows: Section I and II, 31 to 68; Section II, 31 to 67; total score, 310 to 677. Test reliability information published by ETS, and based on tests given between July, 2002, and June, 2003, was as follows: total test, .95; listening comprehension, .90; structure and written expression, .87; and reading, .88 (Educational Testing Service, 2006).

In this study, the TOEFL scores from the two tests administered in early April of students' freshman year are used as a measure of English proficiency before beginning university. Because the tests were administered only a few days apart, it was possible to produce an index of reliability based on equivalent forms by correlating the two results. The reliability for the total score was .76 (N = 532). Section scores were available for the classes that entered in 2007 and 2008 (N = 309). The equivalent forms reliability of each section was as follows: listening, .72; structure and written expression, .68; and reading, .54. These coefficients are rather low.

The reason for such low reliability of the total and section scores might be that most students took the test for the first time, and they might have been anxious, especially during the first administration. On the second administration, students were probably more relaxed and more able to demonstrate their abilities. In fact, the data suggest that this inference is reasonable; the following are statistics for each administration: TOEFL 1, $M = 472.23$, $SD = 39.43$ ($N = 701$); TOEFL 2, $M = 482.71$, $SD = 35.54$ ($N = 532$). Even though more scores were available for the first administration because the class that entered in 2004 took the test only one time, the second TOEFL score was used in this study because those scores were probably more reliable than those from the first administration. However, because both TOEFL scores were not available for all students, regression was used to predict the missing second scores from the first set of scores. The predicted second score was imputed for 169 students for the sample consisting of 2004 to 2008 students. However, in the 2005 to 2008 sample of students ($N = 557$), the predicted second TOEFL score was imputed for only 25 students.

Student Questionnaire

The participants were asked to complete the Student Questionnaire (See Appendix A for the original Japanese version, Appendix B for the English translation.). They were asked about a variety of non-cognitive demographic variables that have been connected, either through research or popular belief, to academic success. Before completing the questionnaire, the students were informed about the study orally and in writing. Directions on the questionnaire explained that answering was entirely voluntary. Initially, students who chose to answer responded using their names and identification numbers. Accordingly, care was taken to ensure that these responses were kept confidential at all times. The identifying information was necessary initially in order to match their responses to other data; however, after the data were matched, all names and identification numbers were deleted. As noted above, participants were then assigned a reference number. Only the Student Affairs Office and Student Records Office retained the identifying information. However, these offices retained only university information, not data gathered from the Student Questionnaire.

Variables used from the questionnaire for this study included only self-reported educational levels of mothers and fathers. The variables, mother education and father education, were coded such that individual education of each parent was recorded as years of study (as in high school educated being equivalent to 12).

University GPAs

As already noted, all students complete at least one level of the EAP program and then go on to take courses first in general education, and then in their major (either International Relations or Global Business). One year of their study takes place in a partner university abroad. In this year abroad, students earn credits, but their university GPA is a result of only their coursework in general education and major courses taken in Japan. Compared to large, more diverse universities, the students' course of study is relatively similar. The GPA in both the EAP and in later courses is calculated as it is in the American system, with an A being equivalent to 4.0 (However, as noted earlier, the EAP GPA is maintained by the EAP department and is not included in the later university GPA of students.). In order to graduate from the university, students must earn a minimum 2.0 university GPA and 124 credits. Outstanding final GPAs are awarded honors: Summa cum laude is awarded to GPAs of 3.85 and above; magna cum laude, to GPAs of 3.75 to 3.84; and cum laude, to GPAs of 3.6 to 3.74.

Data Collection

HSGPA

The HSGPA of students was retrieved from the computer records of the Academic Affairs and Student Records offices, except for the first class of students (2004). The high school records for the class that matriculated in 2004 were obtainable only by going through transcripts stored on paper in notebooks maintained by the Admissions Office. In order to obtain these missing HSGPA records, I was given permission to view these transcripts and transfer these data to the computer records that contained all other students' high school records. Staff from the Admissions Office assisted in this work, which required approximately two months (September, 2008 to November, 2008). All data were checked and double-checked as they were entered. In the process of obtaining these HSGPA records, an opportunity to obtain individual high school course grades, as well as high school absences, for the classes entering from 2004 to 2008 (N = 692) arose, and was taken. The work was done in the Admissions Office records office so that all material would be kept confidential; because of scheduling difficulties this additional data collection required another six months to complete (October, 2008 to April, 2009). Though the students could be identified in this phase of the research, the final records were labeled only with a reference number, and not with student identification numbers or names.

Center Examination and University Entrance Examinations

The participants' Center Examination and university entrance examination scores were retrieved from the computer records of the Admissions Office, except for the first class of students who matriculated in 2004. (Descriptions of these tests are included in the Materials section.) These data were retrieved in a two-month period, from October through December, 2008.

Hensachi Rankings

Using two lists of *hensachi* rankings (which provided essentially identical information) obtained from admissions officials in Fall of 2009, a research assistant and I entered a ranking for the high school of every student into the data record (N = 713). In some cases, there were several rankings for one school. In that case, the highest ranking was used because highest rankings were most relevant for liberal arts universities, according to the Admissions Office official. In addition, approximately 15 of the schools were missing from the lists. The admissions official checked the entire list and added rankings for the 15 schools. In most cases, these schools were not listed because of recent name changes. In two or three cases, the admission official had to consult a third list (http://momotaro.boy.jp) to confirm the *hensachi*. For this group of students for whom *hensachi* rankings were available (N = 713), the range was 35 to 76 (M = 58.26, SD = 6.80). The range and mean were higher than averages cited earlier.

ITP TOEFL

The ITP TOEFL scores of students were retrieved from the computer records of the Student Records office. These data were retrieved in a two-month period, from October through December, 2008.

Student Questionnaire

The student questionnaire was administered to students from 2007 through 2008. Students were asked to voluntarily complete the questionnaire in large-group meetings, such as during university orientation, in university classes, or by letter (in the case of graduates). Over this two-year period, 731 questionnaires were collected.

University GPAs

The semester GPAs and credits accrued of all students who were admitted from 2004 to 2008 were provided by the Student Records Office in Fall of 2008.

These university records also included the EAP grades, but records for approximately 20 students were missing and had to be obtained directly from the EAP office. In addition, the cumulative GPAs and accrued credits of graduating seniors were obtained, beginning in Spring, 2008, and continuing until Fall, 2009.

In order to assess reliability and dimensionality, I requested records with specific course and instructor information. However, due to university policies regarding student records, it was not possible to obtain such records for the general education or major courses. Nevertheless, it was possible to obtain EAP records for the class of 2008. Of these students, 110 were in Level 3, and 51 were in Level 1 or Level 2. For students in Level 3, only three EAP courses were used to calculate averages (Reading, Writing, and Listening), but for students in Levels 1 and 2, four courses were used (Reading, Writing, Listening, and Speaking). Because reliability estimates are dependent on the number of courses, a better estimate typically would result from an analysis of the four courses. Therefore, it was decided to assess dimensionality and reliability using the information from the 51 Level 1 and 2 students.

Procedures

After collecting the above data, including: HSGPA, and individual high school grades and absences; Center Examination scores for the classes that entered from 2005 to 2008; type of entrance exam passed and entrance examination scores from English and Japanese tests, English essays, and interviews for the classes that entered from 2005 to 2008; *hensachi* rankings of student high schools; two TOEFL scores from April of each student's freshman year; TOEFL section scores for the classes that began in 2007 and 2008; questionnaire information (including mother and father education); EAP cumulative average and number of semesters required for EAP program completion; first-year regular university GPA (the cumulative GPA students earned in two semesters after finishing the EAP program); final regular university GPA; number of semesters needed for graduation; and number of students who graduated with honors, I evaluated the data to ensure that they were ready for analysis, according to recommendations of Tabachnick and Fidell (2007, pp. 60-92).

Initially, I checked the dimensionality and reliability of GPAs. As noted above, the only data available for this examination were obtained from the EAP program. The specific EAP data used to investigate dimensionality and reliability came from the four EAP course grades of 51 Level 1 and 2 students who entered in 2008. I used principal components analysis in order to evaluate the dimensionality of these grades. The results are presented in Table 3.

These results show that the three grades are fundamentally unidimensional and account for 45.79% of the variance. According to Armor (1974), any factor that accounts for 40 to 60% of the variance is a good solution; therefore, these results look favorable in this sense. The reliability of these grades can be estimated using the formula for theta cited in the literature review. With four course grades and an eigenvalue from the principal components analysis of 1.832, the resulting theta is .61. This value is moderate and fits within the .50 to .80 range that Pedhazur (1997) reported as being often obtained in non-experimental research.

Table 3. *Factor Loadings from Principal-Components Analysis of Four EAP Course Grades: Communalities, Eigenvalue, and Percentage of Variance*

Course	Component 1	Communality
Reading	.77	.59
Writing	.72	.52
Listening	.58	.34
Speaking	.62	.39
Eigenvalue	1.83	
% of variance	45.79	

High school grades were also assessed for dimensionality and reliability. The factor analysis of nine cumulative high school grades (in Japanese, history, civics, math, science, health education, art, foreign language, and home economics) resulted in one factor and an eigenvalue of 5.01, accounting for 55.65% of the variance. Using the formula for theta, the reliability of high school grades was calculated to be .90.

The general education and major courses could not be examined for unidimensionality in this way because the necessary information was unavailable, but it can be argued that it is likely that the grades for these courses follow a generally unidimensional pattern similar to that of the EAP courses. First, the courses are similar to the EAP classes in that students continue to participate in general education and major courses by making use of the same English reading, writing, listening, and speaking skills. In addition, and as already noted, students choose one of only two majors, International Relations or Global Business. Both majors build on the same general education course requirements, and students in one major also take a number of courses in the other major. Therefore, the course work is similar. Even the majors themselves are related; both build primarily on the disciplines of the humanities and business. In conclusion, it is likely that the GPA of students in both majors follows the pattern observed with the EAP data, and is fundamentally unidimensional.

The reliability of the GPA based on general education and major courses could be roughly estimated using the method proposed by Werts, Linn, and Jöreskog (1978), which was previously discussed. As Werts et al. explained, it is possible to calculate the reliability of any semester GPA if the correlations between semester GPAs exhibit a simplex form. When the simplex form is present, correlations between GPAs from two semesters decrease as the time between them increases; under these conditions the reliability of any semester GPA can be calculated with the formula shown in the literature review.

The correlations between semester GPAs are shown in Table 4. These correlations are all significant, and after the fifth semester are large, most likely a result of students taking fewer courses during their senior year. The correlations reveal a pattern that is essentially simplex in nature. The pattern exemplified by the students who completed semester 8 is a good example. This set of correlations from the first to the eighth semester is for the same group of 396 students. As is characteristic of a simplex pattern, the correlations gradually decrease as time between semesters increases. For example, the correlation between the GPAs of semesters 7 and 8 was .89, but the correlation dropped gradually to .85, .80, .57, .35, .23, and .23, as time between the semesters increased.

Table 4. *Intercorrelations of Semester GPAs*

Measure	1	2	3	4	5	6	7	8
1. Semester 1	---							
2. Semester 2	.48*	---						
3. Semester 3	.24*	.48*	---					
4. Semester 4	.19*	.38*	.64*	---				
5. Semester 5	.19*	.34*	.48*	.78*	---			
6. Semester 6	.17*	.24*	.33*	.55*	.77*	---		
7. Semester 7	.21*	.30*	.41*	.67*	.91*	.96*	---	
8. Semester 8	.23*	.23*	.35*	.57*	.80*	.85*	.89*	---

Note. *$p < .05$. Semester GPAs are based on different numbers of students, depending on number enrolled at the time the data were obtained. N for each semester follows: 1, N = 718, 2, N = 718, 3, N = 556; 4, N = 556; 5, N = 556; 6, N = 401; 7, N = 396; and 8, N = 396.

Because these correlations reveal a simplex pattern, it is possible to estimate the reliability of the GPA for any semester. Using the method proposed by Werts, Linn, and Jöreskog (1978), a number of estimates for each semester can be obtained that are roughly equal within sampling error limits. However, it is clear that results will vary, depending on the size of correlations between semesters. To obtain one conservative estimate of reliability of these GPAs, the reliability coefficient of semester 6 was calculated by averaging the coefficients obtained from correlations most distant in time (semesters 6, 1, and 8) and those closest in time (semesters 6, 5, and 7). The resulting coefficient was an average of .62 and .82, or .72. This is considered a rough estimate of the reliability of the general education and major course GPA.

In the evaluation of the data as a whole, some of the data (particularly GPAs) were found to be negatively skewed; however, the variables were not transformed for a number of reasons. First, the sample size is large, and, as Tabachnick and Fidell (2007) noted, “In a large sample, a variable with statistically significant skewness often does not deviate enough from normality to make a substantive difference in the analysis”; examining the level of skewness as well as the “visual appearance of the distribution” (p. 80) is often sufficient. Though negatively skewed, these distributions generally look normal. In addition, Tabachnick and Fidell pointed out that if all variables are skewed to approximately the same moderate level (as these were), transformation does not result in significant improvement in the analysis of these variables (p. 87). Finally, transforming variables is not recommended in all cases because of difficulties in interpretation that arise, especially when the scales of the variables are commonly used and understood (p. 86). This is the case with GPA; therefore, it is preferable to leave the data in their original form.

Another preliminary analysis concerned the *hensachi* variable. University admissions officials in Japan commonly assume that *hensachi* rankings moderate HSGPA; in other words, they believe that a 4.0 grade point average from a high school that has a low *hensachi* ranking is not equivalent to the same HSGPA from a highly rated high school. Therefore, the *hensachi* variable was examined to determine whether *hensachi* rankings moderate HSGPA, as is commonly believed in Japan, or not. This assumption was tested with multilevel linear modeling, and the HLM software, student version 6.08, was used for the evaluation (Raudenbush, Bryk, Cheong, & Congdon, 2004).

Multilevel linear modeling is an analysis that allows researchers to examine nested or clustered data, which is common in educational research, as when students are clustered within the same classroom or school. Because group members tend to be more like one another than those from different groups, results of statistical analyses of such groups are affected by the group relationship. For example, when researchers attempt to use ordinary regression with clustered data, regression coefficient standard errors will usually be underestimated, leading researchers to believe their results are meaningful when they are not (a Type I error) (Pedhazur, 1997, p. 692). In contrast, researchers who use multilevel analysis with clustered data are able to take account of both the students and the group within which they are nested, and can answer a variety of questions, including whether predictors at the level of the school can "moderate regression slopes" between a predictor at the level of the student and achievement (Schreiber & Griffin, 2004, p. 24). This type of "cross-level" inference, of whether *hensachi* rankings moderate between HSGPA and first-year GPA in university, is the focus of the multilevel modeling analysis of *hensachi* rankings.

However, it must be noted that the data from the current study, which were not collected for the purpose of multilevel linear modeling, are not ideal for this particular analysis because most students are not clustered in groups within high schools. Porter (2005) recommended that clusters consist of 10 members, and that larger group size was preferable (p. 28). However, in the sample used, 295 students were “clustered” within 213 high schools. As is clear from these numbers, in many cases, only one student represented a particular high school. Additional reasons why the data were not ideal for this analysis were that the data were collected from neither a random nor a national sample, and that running the multilevel model with this data took more than 7000 iterations, suggesting a poor fit to the data (Lakin, 2010, p. 4). Therefore, the multilevel linear modeling analysis of the *hensachi* rankings presented here must be regarded as a preliminary investigation and simulation only.

As Tabachnick and Fidell (2007) explained, multilevel modeling is carried out in steps. The first step is the intercepts-only, or null model, a model that does not include predictors and which tests for mean differences between the groups on the dependent variable. This first model generates an intraclass correlation coefficient (ICC) which is derived from variance components. The size of the ICC is an indication of whether multilevel modeling is indicated. Lakin (2010) noted that an ICC of less than .05 suggests that multilevel modeling is not indicated, although researchers might still use the analysis for theoretical reasons. The next step of multilevel modeling is a second model which adds the “first-level predictors,” or individual difference predictors, to the null model (p. 789). In the current study, HSGPA is used as the individual difference predictor. In the third model the “second-level predictor” or the variable at the level of the group, is added to the second model. The second-level predictor in the current study is the *hensachi* ranking. In this third model, the *hensachi* ranking becomes the independent variable while the dependent variable becomes the slope produced by the HSGPA in the second model. If the results show that the random effect of this model is significant for the HSGPA slope, then *hensachi* rankings do moderate the effect of HSGPA, as is commonly assumed in Japan. If the effect is not significant, the assumption is not supported.

The initial analysis of the null model indicated that ICC values were small for data of the current study. For EAP and final GPA, the ICC values were .02 and .01, respectively. For first-year GPA, the ICC was .30. The multilevel modeling analysis was pursued with first-year GPA, as a simulation, to determine whether results supported the assumption that *hensachi* ranking moderated HSGPA. Results showed that the coefficient and t-ratio for *hensachi* ranking were .01 and 1.76 ($p = .08$), respectively. The random coefficient was not significant ($p > .50$), indicating that the common assumption that HSGPA varies with *hensachi* ranking was not supported. In other words, admissions officials in Japan should regard the same HSGPA from different high schools as roughly equivalent. Figure 1 presents the level-one equations graphing, which demonstrates the finding visually.

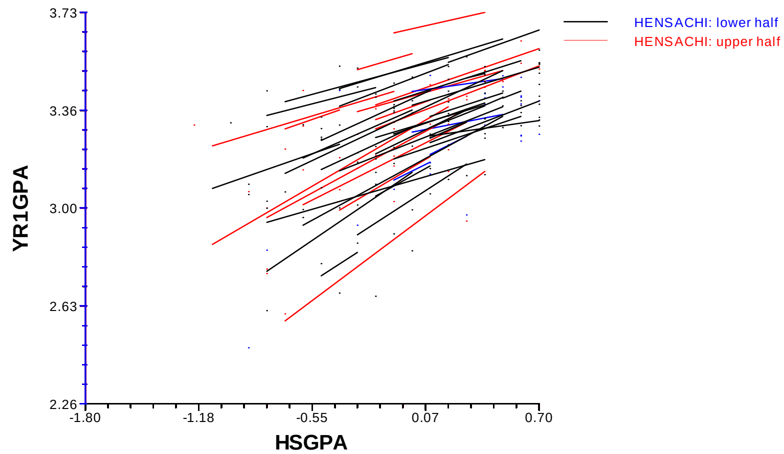

Figure 1. Level-one equation graphing displaying the relation of HSGPA and first-year GPA for high schools (N = 213) ranked at lower half and upper half of *hensachi* rankings

The slopes for the lower and upper *hensachi* rankings are roughly parallel, indicating that the relationship between HSGPA and first-year GPA are roughly the same among the 213 high schools sampled. In other words, *hensachi* rankings do not moderate HSGPA. Again, the results of this analysis are only a simulation; however, the results give some justification for including *hensachi* rankings as a separate predictor variable, along with the individual difference variables.

Analyses

In researching how well variables predict later academic performance, a number of issues must be clarified. For example, the definition of college success must be made clear. As Burton and Ramist (2001) pointed out, first-year GPA is used as a “proxy for success in college” in most predictive validity studies (p. 1). In the current study, the issue is more complex because students can enter the EAP program at level 1, 2, or 3, with each level requiring one semester of study. Therefore, both cumulative EAP GPA and first-year regular university course GPA (defined as the cumulative GPA derived from two semesters of regular university study after completion of the EAP program) are used as measures of college success. In addition, it was possible to obtain the final cumulative GPA of students who graduated, so that measure is used as well. Hierarchical multiple regression is used initially to determine which variables predict first-year cumulative EAP GPA, first-year regular university course GPA, and final cumulative GPA for students who enrolled in the university from 2005 to 2008. After this initial analysis, and an investigation of the predictive validity of entrance examinations, the definition of academic success is extended to include exemplary program completion, as defined by finishing the EAP program on time, graduating on time, and graduating with honors.

The study consists of three parts. In the first part of the study, a series of hierarchical multiple regression analyses are performed in order to answer the first four research questions. To answer the first research question, admissions information about the cohorts that matriculated from 2005 to 2008 are examined as potential predictors of EAP GPA, first year GPA in regular university courses, and final cumulative GPA of graduating students (the dependent variables). The independent variables are HSGPA, total ITP TOEFL score, and *hensachi* ranking. As Field (2005) pointed out, hierarchical multiple regression is appropriate when the researcher decides the order of variables to be entered into the model based on results of past research (p. 160). These variables are entered in the order listed above because previous research (Astin, 1993; Johnson, 1988; Light & Wan, 1991; Zwick, 2002) has indicated that HSGPA should be the strongest predictor, with TOEFL also contributing as well. In the case of *hensachi* rankings, no research has been conducted examining their predictive validity. However, because researchers have often made the point that high school level is predictive of academic success in university, it seems likely that *hensachi* rankings are related to university success (provided that *hensachi* rankings are accurate indicators of high school level).

In follow-up pullout studies that use the best predictors from the initial regression analysis, additional independent variables are examined in further hierarchical multiple regressions. As is true for other similar predictive validity studies, the aim is to determine the best set of predictors. The independent variables used in the pullout studies include: individual high school grade information for classes that entered from 2005 to 2008 (using high school grade factor scores, instead of HSGPA, to answer the second research question); ITP TOEFL section scores for classes that entered in 2007 and 2008 (using TOEFL section scores, instead of TOEFL total scores, to answer the third research question); and non-cognitive data (gender and parental education) derived from student questionnaires for classes that entered from 2005 to 2008 (in order to answer the fifth research question).

In the second part of the study, Japanese university entrance examinations are investigated. Hierarchical multiple regression is used to determine whether Center Examination section scores and university entrance examination type and scores predict academic performance in EAP courses, in the first year of regular university courses, and in the final cumulative GPA of graduating students, over and above predictors found in the first part of the study. Again, the aim of each research question is to determine the best set of predictors. Center Examination test scores for the classes that entered from 2005 through 2008 (and 2006 to 2008, for the English section test), type of university entrance examination passed at the university under study, and entrance examination Japanese and English tests, English essay, and interview scores for classes that entered from 2005 to 2008 are examined to determine the degree to which each predicts academic success over and above the predictors found in the first part of the study (thereby answering the fifth and sixth research questions).

In the third part of the study, the larger picture of academic success is examined in two ways. The first examination of this larger picture of success involves determining the group of variables that contribute most to timely and exemplary completion of program requirements. In order to examine exemplary program completion, a series of logistic regression analyses are performed in order to answer the seventh and eighth research questions. The seventh research question concerns the extent to which HSGPA, high school grade factor scores, total ITP TOEFL scores, *hensachi* rankings, and non-cognitive variables predict finishing the

EAP program on time (defined as completing each level of required study in one semester). Logistic regression is used to examine to what degree these independent variables predict finishing the EAP program on time (for classes that matriculated from 2005 to 2008), thereby answering the seventh research question.

The eighth research question concerns the extent to which HSGPA, high school grade factor scores, total ITP TOEFL scores, *hensachi* rankings, and non-cognitive variables predict graduating on time, and graduating with honors, for students in the 2005 and 2006 cohorts. Burton and Ramist (2001) stated that, along with first-year university GPA, long-term success is equally important (p. 1); certainly, most people would agree that graduating is evidence of academic success, despite the fact that many factors can affect whether students graduate or not. In addition, graduating on time is a more precise definition of long-term success. Though definitions of graduating on time vary, Astin (1993) stated that the strictest definition, graduating in four years, produces superior results in analyses because students who take longer to graduate possess qualities more like those of dropouts than of graduates (p. 192). Therefore, the definition of graduating on time in the current study is graduation within four academic years. The other measure of long-term success that is investigated in the current study is graduation with honors, which was defined as earning cum laude, magna cum laude, or summa cum laude upon graduation (with a cumulative GPA of 3.6 to 4.0).

The second examination of the larger picture of academic success involves determining the ways in which HSGPA, total ITP TOEFL scores, *hensachi* rankings, non-cognitive variables, Center Examination English section scores, university entrance examination English essay scores, EAP GPA, and first-year university GPA interact over time in accounting for university achievement. In order to examine this larger picture of academic success over time, all variables are combined in a path analysis to determine how they relate to one another over time and account for academic achievement in the EAP program, and in the first year of regular university study, thereby answering the ninth research question. Data from students who entered the university from 2005 to 2007 are used in the path model, and the analysis extends through students' first year of regular university study. Figure 1 presents the hypothesized relationship of non-cognitive variables (e.g., gender, father education, and mother education), HSGPA, total ITP TOEFL scores, Center Examination English section test scores, university entrance examination English essay scores, *hensachi* rankings, EAP cumulative GPA, and first-year GPA in the regular university program. Because scores for the various examinations are on different metrics, *z*-scores will be used for the path analysis.

Statistical Assumptions

The assumptions of hierarchical multiple regression have been checked. These include that all variables are quantitative, categorical, or dummy coded; that predictors have variance that is not zero; that the predictors do not correlate highly (resulting in multicolinearity); that the predictors are not correlated with some outside variables; that residuals remain the same at every level of the predictors

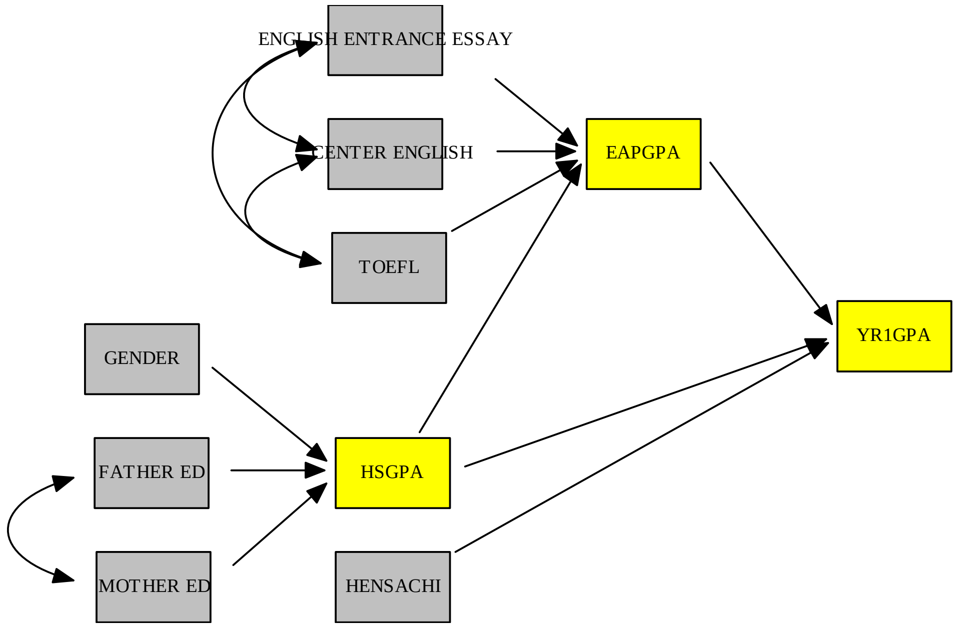

Figure 2. Hypothesized path model of academic achievement up to first-year GPA, including cognitive and non-cognitive variables.

(i.e., homoscedasticity); that error terms or residuals are uncorrelated; that errors are normally distributed; that all outcome values are independent; and that the model produced is linear (Field, 2005, pp. 169-170). Multicollinearity was not found to be a problem in any analysis; all VIF values were close to 1.0.

Logistic regression, used in the second part of the study, has almost no restrictions (Tabachnick & Fidell, 2007), and allows the researcher to use a wide variety of predictor variables (p. 441). The outcome variable should be discrete, which is appropriate for this second part of the study (determining which students graduate and which do not, as well as which students finish the EAP program on time, and which do not). However, Field (2005) warned that researchers must be careful of two situations with logistic regression. One problematic situation occurs when the researcher is doing an analysis that includes one or more empty categories. According to Field, this will result in an incorrect solution from SPSS, as well as a large standard error. He recommended that researchers look at breakdowns of all the categorical independent variables, and inspect the size of standard errors. The second problematic situation, known as perfect separation, occurs when the outcome variable is perfectly predicted. Perfect separation is often caused by too many variables being fitted to a small number of cases. Field suggested that researchers collect more data or use a simpler model to remedy this problem (pp. 265-266).

The main analysis using hierarchical multiple regression is conducted first. Based on previous research, the following independent variables regarding the classes that entered from 2005 to 2008 are entered in blocks in the order listed, using the forced entry method recommended by Field (2005, pp. 177-178): (a) HSGPA; (b) ITP total TOEFL scores; and (c) *hensachi* rankings. Four follow-up hierarchical multiple regression analyses, two logistic regression analyses, and a path analysis are conducted later.

According to Pedhazur (1997), Sewall Wright developed path analysis as a way to study the indirect and direct predictors of an outcome variable. A path diagram lays out the hypothesized causal relationship among exogenous and endogenous variables. While the variation in exogenous variables is assumed to be caused by factors outside the model, the variation in endogenous variables is assumed to be explained by variables inside the model. Arrows are drawn from the independent variables, or those assumed to be causes, to the dependent variables, those assumed to be effects. The flow of cause to effect in a recursive path model is unidirectional and from left to right. As a way to show that there are other outside causes of variance in variables, residuals are also included in the path diagram.

Pedhazur explained that recursive path analysis includes a number of assumptions, including the following: variable relationships are “linear, additive, and causal;” residuals are not correlated with variables to the left of them in the diagram; all variables of importance are included in the model; the cause-effect flow is unidirectional; variables are on an interval scale; and variables are measured without error (p. 771). In addition, because path analysis is based on multiple linear regression, those assumptions should also be met.

CHAPTER 4

RESULTS

Before examining the results for each research question, descriptive statistics are presented for all predictor measures (for students enrolling from 2005 to 2008) in Tables 5 to 10 below. Table 5 presents descriptive statistics for HSGPA, ITP TOEFL, and *hensachi* rankings. (It should be noted that the maximum HSGPA in Japan is 5.0.) These statistics show that the participants have generally high grade point averages (M = 4.22) and relatively high English proficiency (as measured by the ITP TOEFL (M = 484.20). In addition, their high schools are ranked nearly one standard deviation above the mean of 50.

Table 5. *Descriptive Statistics for HSGPA, ITP TOEFL Scores, and Hensachi Rankings*

	HSGPA	ITP TOEFL	*Hensachi*
M	4.22	484.20	58.44
SE	.02	1.51	.29
95% CI	[4.18, 4.26]	[481.25, 487.16]	[57.87, 59.00]
SD	.47	35.51	6.86
Skewness	-.63	.29	-.51
SES	.10	.10	.10
Kurtosis	.15	1.48	.29
SEK	.21	.21	.21

Note. N = 566 for HSGPA (range = 2.4-5.0); N = 557 for ITP TOEFL (range = 347-637); N = 568 for *Hensachi* (range = 35-76), for 2005 to 2008 students.

Table 6 presents descriptive statistics for the number of years of education for the participants' fathers and mothers, as reported by the participants on the questionnaire. Given that primary and secondary school education totals 12 years in Japan, the means for both the fathers (M = 14.37) and mothers (M = 13.55) indicate that many attended or graduated from a junior college or university, though the patterns are different for fathers and mothers. In the case of fathers, most finished university. However, the percentage of those who finished only high school was 35.42%; up to 14 years of schooling, 8.33%; and through university, 48.70%. As for mothers, most finished high school, not university. The percentage of those who finished only high school was 44.36%, up to 14 years of schooling, 31.08%, and through university, 21.81%. (In addition, about 7% of the fathers and 3% of the mothers completed either 9 years of schooling or more than 16, but those percentages are not reported here.) Besides the fact that more fathers finished university than mothers, one other trend is apparent here. The 14 years of schooling probably represents junior college or *senmon daigaku*, colleges for specialty or technical training in a variety of subjects, such as secretarial work, travel consulting, aesthetics, and other fields. While only 8.33% of the fathers finished this level of schooling, 31.08% of the mothers finished at this junior college level. This finding is not surprising because junior colleges have traditionally fulfilled the function of finishing schools for women (McVeigh, 1997).

Tables 7 and 8 present the descriptive statistics for the Center Examination section tests. Table 7 presents descriptive statistics for the English, Japanese, and Social Studies section tests. Although results for the Japanese and Social Studies

Table 6. *Descriptive Statistics for the Number of Years of Education for the Participants' Fathers and Mothers*

	Father's education	Mother's education
M	14.37	13.55
SE	.12	.09
95% CI	[14.13, 14.61]	[13.36, 13.73]
SD	2.25	1.73
Skewness	-.24	.41
SES	.13	.13
Kurtosis	-.55	-.27
SEK	.27	.26

Note. *N* = 329 for Father's Education (range = 9-20); *N* = 340 for Mother's Education (range = 9-20) for 2005 to 2008 students.

Table 7. *Descriptive Statistics for the Center Examination English, Japanese, and Social Studies Section Tests*

	English	Japanese	Social studies
M	179.18	74.27	81.08
SE	.62	.52	.58
95% CI	[177.95, 180.41]	[73.24, 75.30]	[79.95, 82.22]
SD	10.85	9.42	9.34
Skewness	-.45	-.45	-.69
SES	.14	.14	.15
Kurtosis	.01	.12	.32
SEK	.28	.27	.30

Note. *N* = 303 for the English test (maximum score = 200, range = 140.80-200) for 2006 to 2008 students; *N* = 325 for the Japanese test (maximum score = 100, range = 43-98); *N* = 264 for the Social Studies test (maximum score = 100, range = 51-98) for 2005 to 2008 students.

section tests are presented for students who matriculated from 2005 to 2008, the results for the English section test are presented for students who matriculated from 2006 to 2008 because the English test was changed in 2006 with the incorporation of a listening section.

Table 8 presents descriptive statistics for the section tests in math and science. For all but the English tests, the participants had mean scores of approximately 75% to 80%, and the range of scores was wide. For the English test, the mean score was 90%, and the range was narrower, indicating that this test was less challenging for the majority of the participants.

Table 8. *Descriptive Statistics for the Center Examination Math and Science Section Tests*

	Math	Science
M	76.62	75.81
SE	.97	.80
95% CI	[74.70, 78.53]	[74.23, 77.40]
SD	12.94	11.01
Skewness	-.33	-.38
SES	.18	.18
Kurtosis	-.25	-.37
SEK	.36	.35

Note. *N* = 178 for Math (maximum score = 100, range = 43-100); *N* = 189 for Science (maximum score = 100, range = 48-97), for 2005 to 2008 students.

Tables 9 and 10 present descriptive statistics for university entrance examination tasks. Table 9 presents the descriptive statistics for English and Japanese test scores. Compared to the Center Examination English section test, the university English entrance examination test appears to have been a more challenging test. The mean was only approximately 50%, with a maximum score of about 80%. Participants scored higher on the Japanese test, with a mean score of 68%. The range for both tests was wide.

Table 9. *Descriptive Statistics for the University Entrance Examination Test Scores*

	English test	Japanese test
M	106.69	67.99
SE	1.33	.65
95% CI	[104.07, 109.30]	[66.70, 69.27]
SD	23.97	11.77
Skewness	-.64	-.64
SES	.14	.14
Kurtosis	.37	.65
SEK	.27	.27

Note. *N* = 325 for English Test (maximum score = 200, range = 26-162); *N* = 325 for Japanese Test (maximum score = 100, range = 18-93) for 2005 to 2008 students.

Table 10 presents descriptive statistics for two other scored entrance examination tasks, English essays and Japanese/English interview tests. Statistics are presented for students who matriculated from 2005 to 2008. The results indicate that the English essay was more challenging than the interview for many of the participants: the mean scores were approximately 69% and 84%, respectively. However, a number of students attained the maximum score for the essay and the interview. No students were able to attain the maximum score for the entrance examination tests in English and Japanese. These results show that the two tests were more challenging than the other two university entrance examination tasks, and the English test was the most challenging (with a mean score of only approximately 53%).

Table 10. *Descriptive Statistics for the University Entrance Examination Essay and Interview Test Scores*

	English essay (100-point version)	Interview test
M	69.10	84.22
SE	1.04	.68
95% CI	[67.05, 71.16]	[82.87, 85.56]
SD	14.40	9.43
Skewness	-.44	-.33
SES	.18	.18
Kurtosis	-.65	-.56
SEK	.35	.35

Note. *N* = 191 for the 100-point English essay (range = 33-100); *N* = 191 for the Interview Test (maximum score = 100, range = 60-100) for 2005 to 2008 students.

Research Question 1: Predicting EAP GPA, First-Year GPA, and Final GPA from HSGPA, ITP TOEFL Scores, and *Hensachi* Rankings

The first research question asked to what extent HSGPA, *hensachi* rankings, and TOEFL score obtained in April, at the start of the students' first year, predict EAP GPA, as well as first-year and final GPA in the regular university program. To answer the first research question, a hierarchical multiple regression analysis was conducted to predict EAP cumulative GPA from HSGPA, ITP total TOEFL score, and *hensachi* rankings for students who entered the university from 2005 to 2008. In this sample all participants were able to complete the EAP program, regardless of their initial level. This group of 551 students (386 female and 165 male students) also had HSGPA, *hensachi* rankings, and TOEFL scores. Descriptive statistics are presented in Table 11. Mean grades for high school and EAP were above average; however, the range extended from below average to excellent. According to the *Institutional Testing Program: Manual for Supervisors*, the mean ITP TOEFL score was at approximately the fourteenth percentile, but the range of scores was wide, extending from below the first to nearly the ninety-fourth percentile (p. 9).

In Table 12, the correlation and partial correlation of each predictor with cumulative GPA in the EAP program is shown. HSGPA and ITP TOEFL scores had moderate, statistically significant correlations with EAP GPA, and remained at a similar level, even when controlling for other predictors. Though the *hensachi* rankings did not have a statistically significant relationship with EAP GPA, the partial correlation of .07 and *p* value of .10 suggest that the relationship might still be of some importance.

Table 11. *Descriptive Statistics for EAP GPA, HSGPA, ITP TOEFL Scores, and Hensachi Rankings for 2005-2008 Students*

	EAP GPA	HSGPA	ITP TOEFL	*Hensachi*
M	3.10	4.23	483.78	58.48
SE	.02	.02	1.50	.29
95% CI	[3.05, 3.14]	[4.19, 4.26]	[480.83, 486.73]	[57.91, 59.06]
SD	.51	.47	35.27	6.85
Skewness	-.42	-.63	.28	-.53
SES	.10	.10	.10	.10
Kurtosis	.14	.21	1.51	.36
SEK	.21	.21	.21	.21

Note. *N* = 551 for EAP GPA (range = 1.06-4.00); HSGPA (range = 2.40-5.00); ITP TOEFL (range = 347-637); and *Hensachi* (range = 35-76).

The correlations in Table 12 were adjusted for attenuation, according to the formula found in Cohen, Cohen, West, and Aiken (2003), in which the correlation of two values is divided by the square root of the product of the reliability coefficient of each measure (p. 56). Given the reliabilities estimated for HSGPA, EAP GPA, and TOEFL (.90, .61, and .77, respectively), the adjusted correlation for HSGPA and EAP GPA was .36, and that for TOEFL and EAP GPA was .32.

Table 12. *The Bivariate and Partial Correlations for HSGPA, ITP TOEFL Scores, and Hensachi Rankings with EAP GPA*

Predictor	Correlation with EAP GPA	Partial correlation with EAP GPA
HSGPA	.27*	.27
ITP TOEFL	.22*	.21
Hensachi	-.001	.07

Note. * $p < .001$. $N = 551$.

Table 13 presents the results of the multiple regression analysis. The results show that HSGPA accounted for a significant amount of variability in EAP GPA, $R^2 = .07$, $F(1, 549) = 41.50$, $p < .001$. A second analysis was performed to determine whether ITP total TOEFL scores predicted EAP GPA over and above HSGPA. ITP total TOEFL scores accounted for a statistically significant amount of variability in cumulative EAP GPA, R^2 change = .04, $F(1, 548) = 27.24$, $p < .001$. Although the *B* value for TOEFL was only .003, Cohen, Cohen, West, and Aiken (2003) made the point that these values have limitations due to the different scales of the variables, and that it is often easier to interpret β as a result. They suggested extending the values beyond two decimal points, as shown in Table 13. The third analysis using the *hensachi* rankings was not statistically significant in predicting EAP cumulative GPA, over and above HSGPA and ITP total TOEFL scores, R^2 change = .004, $F(1, 547) = 2.69$ ($p = .10$).

Table 13. *Multiple Regression Results for Predicting EAP GPA for the 2005-2008 Students*

	B	SEB	β
Model *1*			
Constant	1.88	.19	
HSGPA	.29	.05	.27*
Model 2			
Constant	.44	.33	
HSGPA	.28	.04	.26*
ITP TOEFL	.003	.00	.21*
Model 3			
Constant	.08	.40	
HSGPA	.31	.05	.28*
ITP TOEFL	.003	.00	.20*
Hensachi	.01	.00	.07

Note. * $p < .001$. $R^2 = .07$ for Model 1; R^2 change = .04 for Model 2 ($p < .001$); R^2 change = .004 for Model 3 ($p = .10$).

A second hierarchical multiple regression analysis was conducted to predict the first-year regular university GPA from HSGPA, ITP total TOEFL score, and *hensachi* ranking for students who entered the university from 2005 to 2007. The sample was limited to these students because only this group had completed one year of university study. Descriptive statistics are presented in Table 14. The mean first-year university GPA of this group of 409 students was 3.16 (*SD* = .48). Although the GPA distribution was negatively skewed, the scores were not transformed for reasons already explained.

Table 15 presents the correlation and partial correlation between each predictor and first-year university GPA. The variables HSGPA and *hensachi* ranking had moderate to small statistically significant correlations with first-year GPA.

Table 14. *Descriptive Statistics for First-Year University GPA for the 2005-2007 Students*

	First-year university GPA
M	3.16
SE	.02
95% CI	[3.12, 3.21]
SD	.48
Skewness	-.79
SES	.12
Kurtosis	.59
SEK	.24

Note. N = 404 (range = 1.17 to 4.00).

After controlling for other predictors, correlations for HSGPA and *hensachi* ranking both increased. Two correlations in Table 15 can also be adjusted for attenuation, as already noted. Given the reliabilities estimated for HSGPA, TOEFL, and university GPA (.90, .77, and .72, respectively), the adjusted correlations for HSGPA and first-year university GPA were .32, and .15 for TOEFL and first-year university GPA.

Table 15. *The Bivariate and Partial Correlations for HSGPA, ITP TOEFL Scores, and Hensachi Rankings with First-Year GPA*

Predictor	Correlation with first-year GPA	Partial correlation with first-year GPA
HSGPA	.26*	.34
ITP TOEFL	.11	.09
Hensachi	.18*	.28

Note. * $p < .001$. N = 387.

Table 16 shows the results of the second hierarchical multiple regression analysis. HSGPA accounted for a significant amount of variability in first-year

GPA, $R^2 = .07$, $F(1, 385) = 28.22$, $p < .001$. A second analysis was conducted to determine whether ITP total TOEFL scores predicted first-year GPA over and above HSGPA. The ITP total TOEFL scores accounted for a significant amount of variability in first-year GPA, R^2 change = .01, $F(1, 384) = 5.32$, $p < .05$. Again, the value of *B* for TOEFL score was only .001, but the advice of Cohen, Cohen, West, and Aiken (2003) was followed. The third analysis included *hensachi* rankings. These rankings accounted for a significant amount of variability in first-year GPA, R^2 change = .07, $F(1, 383) = 31.81$, $p < .001$.

Table 16. *Multiple Regression Results for Predicting First-Year GPA of the 2005-2007 Students*

	B	SEB	β
Model 1			
Constant	2.02	.22	
HSGPA	.27	.05	.26**
Model 2			
Constant	1.22	.41	
HSGPA	.28	.05	.27**
ITP TOEFL	.002	.00	.11*
Model 3			
Constant	-.10	.46	
HSGPA	.36	.05	.35**
ITP TOEFL	.001	.00	.08
Hensachi	.02	.00	.28**

Note. $^*p < .05$. $^{**}p < .001$. $R^2 = .07$ for Model 1; R^2 change = .01 for Model 2 ($p < .05$); R^2 change = .07 for Model 3 ($p < .001$).

A third hierarchical multiple regression analysis was conducted to predict final GPA in the university program from HSGPA, ITP total TOEFL score, and *hensachi* ranking, for students who entered the university from 2005 to 2006. The sample was limited to these students because only this portion of the overall sample had graduated. Descriptive statistics are presented in Table 17. The mean final university GPA of this group of 171 students was 3.32 (*SD* = .34) on a 4-point scale.

Table 17. *Descriptive Statistics for Final GPA for the 2005-2006 Graduating Students*

	Final university GPA
M	3.32
SE	.03
95% CI	[3.27, 3.38]
SD	.34
Skewness	-.29
SES	.19
Kurtosis	-.59
SEK	.37

Note. *N* = 171 (range = 2.46 to 4.00)

In Table 18, the correlation and partial correlation of each predictor with final GPA are shown. The variables HSGPA and *hensachi* rankings had statistically significant correlations with final GPA. After controlling for other predictors, HSGPA and *hensachi* rankings both increased in their ability to predict final university GPA.

The correlations in Table 18 were adjusted for attenuation, according to the formula found in Cohen, Cohen, West, and Aiken (2003, p. 56). Given the reliabilities estimated for HSGPA, ITP total TOEFL score, and university GPA (.90, .77, and .72, respectively), the adjusted correlations were .35 for HSGPA and final university GPA, and .18 for ITP total TOEFL score and final university GPA.

(Correlations were adjusted for attenuation when answering the first research question. However, answers to all other research questions are based on unadjusted correlations. It should be borne in mind that the higher adjusted correlations are probably closer to representing actual correlations of variables because of problems in the data, such as lack of perfect reliability of variables and range restriction, that lower correlations.

Table 18. *The Bivariate and Partial Correlations for HSGPA, ITP TOEFL Scores, and Hensachi Rankings with Final GPA*

Predictor	Correlation with final GPA	Partial correlation with EAP GPA
HSGPA	.28**	.34
ITP TOEFL	.13	.11
Hensachi	.15*	.24

Note. * $p < .01$. ** $p < .001$. $N = 165$.

Table 19 presents the results of the multiple regression analysis. The results show that HSGPA accounted for a significant amount of variability in final university GPA, $R^2 = .08$, $F(1, 163) = 13.98$, $p < .001$. A second analysis was performed to determine whether ITP total TOEFL scores predicted final university GPA over and above HSGPA. As was true in the case of first-year GPA, ITP total TOEFL scores did not reach statistical significance in accounting for variability in final university GPA, R^2 change = .02, $F(1, 162) = 2.80$, $p = .10$. The third analysis using *hensachi* rankings was statistically significant in predicting final university

GPA, over and above HSGPA and ITP total TOEFL scores, R^2 change = .05, $F(1, 161) = 9.81$, $p < .01$.

Table 19. *Multiple Regression Results for Predicting Final GPA of the 2005-2006 Students*

	B	SEB	β
Model 1			
Constant	2.44	.24	
HSGPA	.21	.06	.28**
Model 2			
Constant	1.84	.43	
HSGPA	.21	.06	.28**
ITP TOEFL	.001	.00	.13
Model 3			
Constant	1.04	.49	
HSGPA	.26	.06	.35**
ITP TOEFL	.001	.00	.10
Hensachi	.01	.00	.24*

Note. R^2 = .08 for Model 1; R^2 change = .02 for Model 2 (p = .10); R^2 change = .05 for Model 3 ($p < .01$).
*p < .01. **p < .001.

Research Question 2: Predicting EAP GPA, First-Year GPA, and Final GPA from High School Grade Factor Scores

Research Question 2 asked to what extent high school grade factor scores predict EAP GPA, as well as first-year and final GPA in the regular university program. As was discussed in the procedural descriptions in the Methods section, high school grades were determined through a factor analysis to be fundamentally unidimensional with one general factor accounting for 55.65% of the variance. However, it was hypothesized that separate types of academic abilities or intelligence might underlie this general academic ability factor. This hypothesis reflects the work of a number of researchers who believe that intelligence is undergirded by more specific and distinct types of ability. For example, Gardner (1990) discussed seven kinds of intelligence: logical/mathematical, linguistic, spatial, kinesthetic, musical, interpersonal, and intrapersonal (pp. 933-935). In addition, Kline (1994) noted that researchers have consistently found three factors underlying intelligence, a more specific general intelligence factor, verbal ability, and spatial ability (p. 12). Therefore, based on a consideration of such theories, as well as conjectures about individual high school grades, it was hypothesized that three academic factors underlie the high school grades; specifically, a language ability factor, a numerical ability factor, and a spatial ability factor.

A maximum likelihood factor analysis was conducted using the high school grades of students who entered the university from 2004 to 2008 to identify and extract factors underlying the high school grades. Based on the scree plot provided by an earlier principal components analysis and the significance testing provided by the maximum likelihood factor analysis, it was determined that three factors were present. The three factors were rotated using Varimax, an orthogonal method of rotation. Even though it was hypothesized that these ability factors were correlated, the Varimax method was chosen because of ease in creating and using factor scores, and in reporting results, as recommended by Tabachnick and Fidell (2007). The rotated solution is shown in Table 20. Four high school subjects load strongly on the first factor: Japanese, History, Civics, and Foreign Language. These subjects require a variety of abilities, but ability to work with language is of prime importance. The second factor has two strong loadings: Mathematics and Science (though it should be noted that Science also loads onto the first factor at .47). These subjects require numerical and analytical ability. The last factor has three loadings from Physical Education, Art, and Home Economics. Home Economics also loads onto the first factor at .41, but Physical Education and Art load strongly onto the third factor. These subjects require spatial, kinesthetic, and creative abilities.

Table 20. *Summary of Factor Loadings for the Three-Factor Solution for High School Grades*

High school grade	Factor 1	Factor 2	Factor 3
Japanese	**.74**	.32	.37
History	**.66**	.37	.26
Civics	**.65**	.39	.19
Mathematics	.36	**.72**	.22
Science	.47	**.71**	.29
Physical education	.26	.29	**.42**
Art	.15	.10	**.63**
Foreign language	**.54**	.26	.32
Home economics	.41	.32	**.54**

Note. *N* = 649. Boldface indicates highest factor loadings.
Factor 1 = Language Ability; Factor 2 = Numerical Ability;
Factor 3 = Spatial Ability.

High school grade factor scores were also created through the above analysis. Based on the loadings on each factor, they were given the following labels: Factor 1, Language Ability; Factor 2, Numerical Ability; and Factor 3, Spatial Ability. These factor scores were examined to determine how each could add to the prediction of EAP GPA, first-year GPA, and final GPA. In Table 21, the correlations and partial correlations of the three factor scores with EAP GPA are listed. The Language Ability factor scores and the Spatial Ability factor scores have moderate, statistically significant correlations with EAP GPA. After controlling for other predictors, the correlations with the two ability factors decrease slightly. The Numerical Ability factor scores did not have a statistically significant correlation with EAP GPA.

Table 21. *The Bivariate and Partial Correlations for Ability Factor Scores and ITP TOEFL Scores with EAP GPA*

Predictor	Correlation with EAP GPA	Partial correlation with EAP GPA
Language ability factor scores	.21*	.14
Numerical ability factor scores	.13	.08
Spatial ability factor scores	.20**	.16

Note. *$p < .01$. ** $p < .001$. $N = 502$.

Table 22 shows the results of the hierarchical multiple regression analysis for students who matriculated from 2005 to 2008. Initially, the three ability factors accounted for a significant amount of variability in EAP GPA, $R^2 = .07$, $F(3, 498) = 12.30$, $p < .001$. Of the three ability factor scores, the Spatial Ability factor accounted for the most variability in the EAP GPA. The second analysis determined that ITP total TOEFL scores predicted EAP GPA over and above the three ability factors. The ITP total TOEFL scores accounted for a significant amount of variability in EAP GPA, R^2 change = .05, $F(1, 497) = 26.44$, $p < .001$.

Because the numerical factor scores were not significant predictors of EAP GPA in the first model, the regression analysis was repeated without the numerical factor scores. The results are presented in Table 23. In the second model the

Language Ability and Spatial Ability factor scores and ITP TOEFL scores were significant predictors of EAP GPA. Because the VIF and tolerance statistics indicated no multicollinearity in this second model, the Language Ability factor scores and ITP TOEFL scores can be seen as representing different types of language ability. The Language Ability factor scores might represent the ability to use language, including the L1, in a wide variety of courses and situations. These factor scores might also represent a language/general intelligence factor, as suggested by Kline (1994).

Table 22. *Multiple Regression Results for Predicting EAP GPA from Three Ability Factor Scores and ITP TOEFL Scores*

	B	SEB	β
Model 1			
Constant	3.08	.02	
Language ability factor scores	.09	.03	.14*
Numerical ability factor scores	.05	.03	.07
Spatial ability factor scores	.10	.03	.15*
Model 2			
Constant	1.55	.30	
Language ability factor scores	.09	.03	.14*
Numerical ability factor scores	.05	.03	.07
Spatial ability factor scores	.11	.03	.16**
ITP TOEFL	.003	.00	.22**

Note. R^2 = .07 for Model 1; R^2 change = .05 for Model 2 (p < .001).
*p < .01. ** p < .001.

The three ability factor scores were next examined to determine how each added to the prediction of first-year GPA. In Table 24, the correlations and partial correlations of the three ability factor scores with first-year GPA are shown. The Language Ability factor scores and the Numerical Ability factor scores had small to moderate, statistically significant correlations with first-year GPA. After controlling for other predictors, the correlations with the two ability factors remained at a similar level. The Spatial Ability factor scores did not have a statistically significant correlation with first-year GPA. After controlling for other predictors, the correlation between the *hensachi* rankings and first-year GPA increased to a level similar to that of the numerical factor scores.

Table 23. *Multiple Regression Results for Predicting EAP GPA from Two Ability Factor Scores and ITP TOEFL Scores*

	B	SEB	β
Model 1			
Constant	3.08	.02	
Language ability factor scores	.10	.03	.16*
Spatial ability factor scores	.11	.03	.16*
Model 2			
Constant	1.55	.30	
Language ability factor scores	.10	.03	.16**
Spatial ability factor scores	.11	.03	.16**
ITP TOEFL	.003	.00	.22**

Note. R^2 = .06 for Model 1; R^2 change = .05 for Model 2 (p < .001).
*p < .01. ** p < .001.

Table 25 shows the results of the hierarchical multiple regression analysis. Initially, the three ability factors accounted for a significant amount of variability in first-year GPA, $R^2 = .08$, $F(3, 352) = 10.82$, $p < .001$. Of the three ability factor scores, the Numerical Ability factor accounted for the most variability in first-year GPA. The second analysis determined that *hensachi* rankings made a contribution to the prediction of first-year GPA over and above the three ability factors. The *hensachi* rankings accounted for a significant amount of variability in first-year GPA, R^2 change = .06, $F(1, 351) = 25.19$, $p < .001$.

Table 24. *The Bivariate and Partial Correlations for Ability Factor Scores and Hensachi Rankings with First-Year GPA*

Predictor	Correlation with first-year GPA	Partial correlation with first-year GPA
Language ability factor scores	.18*	.18
Numerical ability factor scores	.26**	.25
Spatial ability factor scores	.13	.08
Hensachi	.15**	.26

Note. *$p < .01$. ** $p < .001$. $N = 356$.

Because the Spatial Ability factor scores were not significant predictors of first-year GPA in the first model, the regression analysis was repeated without those factor scores. The results for this second regression analysis, which are presented in Table 26, show that the Language Ability factor, Numerical Ability factor, and *hensachi* rankings were all significant predictors of first-year GPA. In addition, although the relative contributions of the Language Ability factor scores, Numerical Ability factor scores, and *hensachi* rankings remained approximately the same as those presented in Table 25, the coefficient increased slightly for the Language Ability factor scores but did not do so for either the Numerical Ability factor scores or the *hensachi* rankings.

Table 25. *Multiple Regression Results for Predicting First-Year GPA from Three Ability Factor Scores and Hensachi Rankings*

	B	SEB	β
Model 1			
Constant	3.14	.03	
Language ability factor scores	.06	.03	.11
Numerical ability factor scores	.14	.03	.23**
Spatial ability factor scores	.05	.04	.07
Model 2			
Constant	2.03	.22	
Language ability factor scores	.11	.03	.19*
Numerical ability factor scores	.15	.03	.25**
Spatial ability factor scores	.05	.04	.08
Hensachi	.02	.00	.26**

Note. R^2 = .08 for Model 1; R^2 change = .06 for Model 2 (p < .001).
**p* < .01. ** *p* < .001.

Table 26. *Multiple Regression Results for Predicting First-Year GPA from Two Ability Factor Scores and Hensachi Rankings*

	B	SEB	β
Model 1			
Constant	3.14	.03	
Language ability factor scores	.07	.03	.13*
Numerical ability factor scores	.14	.03	.23**
Model 2			
Constant	2.03	.22	
Language ability factor scores	.12	.03	.21**
Numerical ability factor scores	.15	.03	.25**
Hensachi	.02	.00	.26**

Note. R^2 = .08 for Model 1; R^2 change = .06 for Model 2 (p < .001).
**p* < .05. ** *p* < .001.

In the last analysis with the ability factor scores, the three ability factor scores were examined to determine how each could add to the prediction of the final GPA of the graduating students. In Table 27, the correlations and partial correlations of the three ability factor scores with final GPA are listed. The

Language Ability factor scores and the Numerical Ability factor scores had moderate, statistically significant correlations with final GPA. After controlling for other predictors, the correlation with the Language Ability factor scores increased, but that of the Numerical Ability factor scores decreased. The Spatial Ability factor scores did not have a statistically significant correlation with final GPA. After controlling for other predictors, the correlation between *hensachi* rankings and final GPA increased to a level similar to that of the two ability factor scores.

Table 27. *The Bivariate and Partial Correlations for Ability Factor Scores and Hensachi Rankings with Final GPA*

Predictor	Correlation with final GPA	Partial correlation with final GPA
Language ability factor scores	.25***	.31
Numerical ability factor scores	.23*	.18
Spatial ability factor scores	-.02	-.14
Hensachi	.10**	.26

Note. *$p < .05$. **$p < .01$. *** $p < .001$. $N = 145$.

Table 28 shows the results of the hierarchical multiple regression analysis. Initially, the three ability factors accounted for a significant amount of variability in final GPA, $R^2 = .10$, $F(3, 141) = 5.17$, $p < .01$. Of the three ability factor scores, the Language Ability factor accounted for the most variability in the final GPA. The second analysis determined that the *hensachi* rankings predicted final GPA over and above the three ability factors. The *hensachi* rankings accounted for a significant amount of variability in final GPA, R^2 change = .06, $F(1, 140) = 10.03$, $p < .01$.

Table 28. *Multiple Regression Results for Predicting Final GPA from Three Ability Factor Scores and Hensachi Rankings*

	B	SEB	β
Model 1			
Constant	3.31	.03	
Language ability factor scores	.10	.04	.24**
Numerical ability factor scores	.07	.03	.17*
Spatial ability factor scores	-.05	.04	-.11
Model 2			
Constant	2.53	.25	
Language ability factor scores	.14	.04	.36***
Numerical ability factor scores	.07	.03	.17*
Spatial ability factor scores	-.07	.04	-.14
Hensachi	.01	.00	.27**

Note. R^2 = .10 for Model 1; R^2 change = .06 for Model 2 (p < .001).
*p < .05. ** p < .01. ***p < .001.

Because the Spatial Ability factor score was not a significant predictor of final GPA in the first model (p = .09), the regression analysis was repeated without that factor score. The results for this second regression analysis are presented in Table 29. In the second model, the Language Ability factor scores, Numerical Ability factor scores, and *hensachi* rankings were significant predictors of final GPA. In addition, although the relative contribution of the Language Ability factor scores, Numerical Ability factor scores, and *hensachi* rankings remained approximately the same as that presented in Table 28, the coefficient decreased most for the Language Ability factor scores while the coefficient of the Numerical Ability factor score increased slightly, and that of the *hensachi* rankings decreased.

Table 29. *Multiple Regression Results for Predicting Final GPA from Two Ability Factor Scores and Hensachi Rankings*

	B	SEB	β
Model 1			
Constant	3.31	.03	
Language ability factor scores	.08	.03	.20*
Numerical ability factor scores	.07	.03	.17*
Model 2			
Constant	2.59	.25	
Language ability factor scores	.12	.04	.30**
Numerical ability factor scores	.07	.03	.18*
Hensachi	.01	.00	.25**

Note. R^2 = .09 for Model 1; R^2 change = .05 for Model 2 ($p < .01$).
*$p < .05$. ** $p < .01$.

Research Question 3: Predicting EAP GPA and First-Year GPA from ITP TOEFL Section Scores

The third research question asked to what extent ITP TOEFL section scores predict EAP GPA as well as first-year GPA in the regular university program. To answer the third research question, a hierarchical multiple regression analysis was conducted to predict EAP GPA from HSGPA and ITP TOEFL section scores, and first-year GPA from HSGPA, *hensachi* rankings, and ITP TOEFL section scores. The scoring range for each section of the ITP TOEFL is as follows: Section I, 31 to 68; Section II, 31 to 68; and Section III, 31 to 67. Descriptive statistics concerning the ITP TOEFL section scores are presented in Table 30.

In Table 31, the correlation and partial correlation of each predictor with EAP GPA is listed. Because the TOEFL Section III range is slightly different from that of Section I and II, all section scores were converted to *z* scores for the analysis.

Table 30. *Descriptive Statistics for ITP TOEFL Section Test Scores*

	Section I listening comprehension	Section II structure and written expression	Section III reading Comprehension
M	48.16	48.58	48.62
SE	.28	.28	.25
95% CI	[47.61, 48.71]	[48.02, 49.14]	[48.12, 49.12]
SD	4.89	5.00	4.48
Skewness	.27	.36	-.86
SES	.14	.14	.14
Kurtosis	.56	.76	1.92
SEK	.28	.28	.28

Note. $N = 310$. Score ranges for sections are: Section I, 31-68; Section II, 31-68; and Section III, 31-67.

HSGPA, TOEFL Section I (Listening Comprehension), and TOEFL Section III (Reading Comprehension) had moderate, statistically significant correlations with EAP GPA. After controlling for other predictors, the correlation with HSGPA increased, but the Section I correlation decreased slightly from .32 to .27, and the Section III correlation decreased from .28 to .18. Even though both Section I and III were significantly correlated with EAP GPA, the partial correlations suggested that the Section I scores were more strongly related to EAP GPA.

Table 31. *The Bivariate and Partial Correlations for HSGPA and ITP TOEFL Section Test Scores with EAP GPA*

Predictor	Correlation with EAP GPA	Partial correlation with EAP GPA
HSGPA	.26**	.29
TOEFL section I	.32**	.27
TOEFL section II	.22	.03
TOEFL section III	.28*	.18

Note. $^{**}p < .01$. $^{**}p < .001$. $N = 309$.

Table 32 presents the results of the multiple regression analysis. HSGPA and ITP TOEFL section scores were analyzed in one model for EAP GPA because these variables were already found to be significant predictors of EAP GPA. The results showed that the set of variables accounted for a significant amount of variability in EAP GPA, $R^2 = .22$, $F(4, 304) = 20.82$, $p < .001$. However, the TOEFL Section II scores did not contribute significantly to the prediction of EAP GPA. Only listening comprehension and reading comprehension accounted for a significant amount of variability in EAP GPA, with listening comprehension contributing to the prediction of EAP GPA to a greater degree.

A hierarchical multiple regression analysis was conducted to predict the first-year regular university GPA from HSGPA, the *hensachi* ranking, and the ITP TOEFL section scores for students who entered the university in 2007. The sample was limited to these 148 students because only this group had completed one year

Table 32. *Multiple Regression Results for Predicting EAP GPA from HSGPA and ITP TOEFL Section Test Scores*

	B	SEB	β
Model 1			
Constant	1.77	.23	
HSGPA	.29	.06	.27**
TOEFL Section I	.13	.03	.27**
TOEFL Section II	.01	.03	.03
TOEFL Section III	.09	.03	.19*

Note. $R^2 = .22$ ($p < .001$).
*$p < .01$. **$p < .001$.

of university study. The TOEFL section scores were entered in a second model to determine whether any of them predicted first-year GPA over and above the known predictors, HSGPA and the *hensachi* ranking.

Table 33 presents the correlation and partial correlation between each predictor and first-year university GPA. HSGPA had a moderately large and statistically significant correlation with first-year GPA, but correlations with ITP section scores were not statistically significant.

Table 33. *The Bivariate and Partial Correlations for HSGPA, Hensachi Rankings, and ITP TOEFL Section Test Scores with First-Year GPA*

Predictor	Correlation with first-year GPA	Partial correlation with first-year GPA
HSGPA	.30*	.35
Hensachi	.10	.15
TOEFL section I	-.01	.01
TOEFL section II	.00	-.02
TOEFL section III	.14	.14

Note. * $p < .001$. $N = 145$.

Table 34 shows the results of the hierarchical multiple regression analysis. HSGPA and the *hensachi* ranking accounted for a significant amount of variability in first-year GPA, $R^2 = .12$, $F(2, 142) = 10.05$, $p < .001$. A second analysis was conducted to determine whether the ITP TOEFL section scores predicted first-year GPA over and above the known predictors, HSGPA and the *hensachi* rankings. As a group, the section scores did not account for a significant amount of variability in first-year GPA, R^2 change = .02, $F(3, 139) = 1.03$, $p = .38$. However, the Section III scores were closer to reaching statistical significance ($p = .10$) than were the

Section I and Section II scores. With another or larger sample, English reading comprehension might account for a significant amount of variability in first-year GPA.

Table 34. *Multiple Regression Results for Predicting First-Year GPA from HSGPA, Hensachi Rankings, and ITP TOEFL Section Test Scores*

	B	SEB	β
Model 1			
Constant	.47	.60	
HSGPA	.41	.10	.35**
Hensachi	.01	.01	.19*
Model 2			
Constant	.59	.62	
HSGPA	.42	.10	.36**
Hensachi	.01	.01	.15
TOEFL section I	.004	.05	.01
TOEFL section II	-.01	.05	-.02
TOEFL section III	.08	.05	.15

Note. R^2 = .12 for Model 1; R^2 change = .02 for Model 2 (p = .38).
*p < .05. **p < .001.

Research Question 4: Predicting HSGPA, EAP GPA, First-Year GPA, and Final GPA from Non-Cognitive Variables

The fourth research questions asked to what extent the non-cognitive variables, gender and parental education, predict HSGPA, EAP GPA, first-year, and final GPA in the regular university program. In order to answer this question, the non-cognitive variables were examined in the following ways: (a) gender and parental education were used as independent variables in the prediction of HSGPA; (b) HSGPA and ITP TOEFL were used in the first model, and then gender and parental education were entered in a second model to predict EAP GPA; (c)

HSGPA and *hensachi* rankings were used in the first model, and then gender and parental education were entered in a second model to predict first-year and final GPA. Because the non-cognitive variables were chosen on the basis of previous research, a hierarchical multiple regression was used in predicting all levels of GPA.

The first analysis examines the predictive validity of gender and parental education for HSGPA. The correlations and partial correlations of the continuous non-cognitive variables, father's education and mother's education, with HSGPA for students who matriculated from 2005 to 2008 are presented in Table 35. Father's education had a statistically significant, but negative, correlation with HSGPA. Mother's education did not have a statistically significant relationship with HSGPA.

Table 35. *The Bivariate and Partial Correlations for Father and Mother's Education with HSGPA*

Predictor	Correlation with HSGPA	Partial correlation with HSGPA
Father's education	-.15*	-.12
Mother's education	-.07.	.01

Note. ***p* < 001. **p* < .05. *N* = 324.

The results of the multiple regression analysis for the prediction of HSGPA from gender and parents' education are shown in Table 36. The results indicate that the set of gender and parents' education accounted for a statistically significant amount of variability in HSGPA, R^2 = .10, F(3, 320) = 11.59, p < .001. Females significantly outperformed males with a mean HSGPA of 4.29 (SD = .43), compared to the male mean HSGPA of 4.06 (*SD* = .53). In addition, the father's education accounted for a statistically significant amount of variability in HSGPA; as the father's education increased, HSGPA tended to decrease.

The next analysis examines the predictive validity of gender and parental education for EAP GPA. The first hierarchical regression model includes HSGPA and ITP TOEFL scores, and the second model includes the three non-cognitive variables, gender, father's education, and mother's education. Table 37 presents the correlation and partial correlation of the father and mother's education with EAP GPA. Neither parent's education had a statistically significant relationship with EAP GPA.

Table 36. *Multiple Regression Results for Predicting HSGPA from Female Gender and Parents' Education*

	B	SEB	β
Model 1			
Constant	4.42	.21	
Female gender	.27	.05	.28**
Father's education	-.03	.01	-.13*
Mother's education	.001	.02	.01

Note. R^2 = .10.
*p < .05. **p < .001.

Table 37. *The Bivariate and Partial Correlations for Father's Education and Mother's Education with EAP GPA*

Predictor	Correlation with EAP GPA	Partial correlation with EAP GPA
Father's education	.07	.07
Mother's education	.04	.03

Note. *N* = 315.

The regression analysis results are presented in Table 38. The results show that HSGPA and TOEFL accounted for a significant amount of variability in EAP GPA, $R^2 = .15$, $F(2, 312) = 27.03$, $p < .001$. A second analysis showed that non-cognitive variables accounted for a statistically significant amount of variability in EAP GPA, R^2 change = .06, $F(3, 309) = 7.48$, $p < .001$. Neither father's education nor mother's education accounted for a statistically significant amount of variability in EAP GPA. However, gender accounted for a significant amount of variability in EAP GPA; females outperformed males with a mean EAP GPA of 3.18 (*SD* = .48), compared to the male mean GPA of 2.90 (*SD* = .54).

Table 38. *Multiple Regression Results for Predicting EAP GPA from HSGPA, ITP TOEFL Scores, Gender, Father's Education, and Mother's Education*

	B	SEB	β
Model 1			
Constant	.02	.42	
HSGPA	.29	.06	.26*
ITP TOEFL	.004	.00	.27*
Model 2			
Constant	-.37	.47	
HSGPA	.24	.06	.21*
ITP TOEFL	.004	.00	.28*
Gender	.26	.06	.24*
Father's education	.02	.01	.07
Mother's education	.01	.02	.03

Note. $R^2 = .15$ for Model 1; R^2 change = .06 for Model 2 ($p < .001$).
*$p < .001$.

In the third analysis, I examined the predictive validity of gender and parents' education for first-year GPA. The first hierarchical regression model includes HSGPA and the *hensachi* rankings, and the second model includes the three non-cognitive variables, gender, father's education, and mother's education. Table 39 presents the correlations and partial correlations of the father and mother's education with first-year GPA. After controlling for father's education, the correlation of the mother's education with first-year GPA decreased from .04 to -.02; however, the correlation of the father's education increased slightly from .10 to .12 after controlling for the mother's education. The father's education did not account for a significant amount of variability in first-year GPA (p = .09).

Table 39. *The Bivariate and Partial Correlations for Father's Education and Mother's Education with First-Year GPA*

Predictor	Correlation with first-year GPA	Partial correlation with first-year GPA
Father's education	.10	.12
Mother's education	.04	-.02

Note. N = 188.

The regression analysis results are presented in Table 40. The results show that HSGPA and the *hensachi* rankings accounted for a significant amount of variability in first-year GPA, R^2 = .15, $F(2, 185)$ = 16.42, p < .001. A second analysis showed that non-cognitive variables as a set did not account for a statistically significant amount of variability in first-year GPA, R^2 change = .02, $F(3, 182)$ = 1.23, p = .30. Gender did not account for a significant amount of variability in first-year GPA; the mean GPA of females (3.18, *SD*, .46), though higher, was not significantly different from that of males (3.14, *SD*, .53). As for parents' education, the amount of variability in first-year GPA accounted for by the father's education was close to reaching statistical significance, indicating that this variable might have importance in predicting student achievement in the first year of university, with another, or larger sample.

Table 40. *Multiple Regression Results for Predicting First-Year GPA from HSGPA, Hensachi Rankings, Gender, Father's Education, and Mother's Education*

	B	SEB	β
Model 1			
Constant	.40	.49	
HSGPA	.37	.08	.35*
Hensachi	.02	.01	.32*
Model 2			
Constant	.01	.57	
HSGPA	.38	.08	.36*
Hensachi	.02	.01	.31*
Gender	.03	.07	.03
Father's education	.03	.02	.14
Mother's education	-.01	.02	-.02

Note. R^2 = .15 for Model 1; R^2 change = .02 for Model 2 (p = .30).
**p* < .001.

In the final analysis, I examined the predictive validity of gender and parents' education for final GPA. The data from students who matriculated from 2004 to 2008 were used so that the sample size would be more adequate. The first hierarchical regression model includes HSGPA and the *hensachi* rankings, and the predictors included in the second model are the three non-cognitive variables, gender, father's education, and mother's education. The correlations and partial correlations of father and mother's education with final GPA are presented in Table 41. After controlling for father's education, the correlation of mother's education with final GPA increased from -.01 to -.10. The correlation of father's education increased slightly from .14 to .17 after controlling for mother's education. The mother's education did not have a statistically significant relationship with final GPA; however, the father's education was nearly significant at p = .06.

Table 41. *The Bivariate and Partial Correlations for Father's Education and Mother's Education with Final GPA*

Predictor	Correlation with final GPA	Partial correlation with final GPA
Father's education	.14	.17
Mother's education	-.01	-.10

Note. N = 123.

The regression analysis results are presented in Table 42. The results show that HSGPA and *hensachi* rankings accounted for a significant amount of variability in final GPA, R^2 = .15, F(2, 120) = 10.85, p < .001. A second analysis showed that non-cognitive variables did not account for a statistically significant amount of variability in final GPA, R^2 change = .03, F(3, 117) = 1.23, p = .30. Gender did not account for a significant amount of variability in final GPA; the mean GPA of females (3.30, *SD*, .34), though higher, was not significantly different from that of males (3.27, *SD*, .37). The amount of variability in final GPA accounted for by father's education was not quite significant, at p = .06. As was the case with first-year GPA, the amount of variance accounted for by the father's education was close to reaching statistical significance, indicating that this variable might have importance in predicting student achievement in the final university GPA, with another, or larger sample.

Table 42. *Multiple Regression Results for Predicting Final GPA from HSGPA, Hensachi Rankings, Gender, Father's Education, and Mother's Education*

	B	SEB	β
Model 1			
Constant	.98	.52	
HSGPA	.33	.07	.42**
Hensachi	.02	.01	.32*
Model 2			
Constant	.89	.56	
HSGPA	.34	.08	.44**
Hensachi	.02	.01	.29*
Gender	-.02	.07	-.03
Father's education	.03	.02	.19
Mother's education	-.02	.02	-.11

Note. R^2 = .15 for Model 1; R^2 change = .03 for Model 2 (p = .30).
*p < .01. **p < .001.

Research Question 5: Predicting EAP GPA, First-Year GPA, and Final GPA from Center Examination Scores

The fifth research question asked to what extent Center Examination scores predict EAP GPA, as well as first-year GPA and final GPA in the regular university program. To answer this research question, hierarchical multiple regression analyses were conducted to predict EAP GPA from HSGPA, ITP total TOEFL scores, and Center Examination scores; first-year GPA from HSGPA, the *hensachi* rankings, and Center Examination scores; and final GPA from HSGPA, the *hensachi* rankings, and Center Examination scores. Descriptive statistics for the Center Examination scores were presented in Tables 7 and 8. Because the five Center Examination subject tests were not all scored on the same scale, *z* scores were used in the analyses.

In Table 43, the correlations and partial correlations of all predictors with EAP GPA are listed for students who matriculated from 2005 to 2008. Only HSGPA had a statistically significant correlation with EAP GPA. After controlling for other predictors, HSGPA remained at a similar level. The ITP TOEFL and Center Examination scores did not have a statistically significant relationship with EAP GPA.

Table 43. *The Bivariate and Partial Correlations for HSGPA, ITP TOEFL Scores, and Center Examination Section Test Scores (2005 to 2008) with EAP GPA*

Predictor	Correlation with EAP GPA	Partial correlation with EAP GPA
HSGPA	.46**	.42
ITP TOEFL	.17	.10
Center English	-.00	-.03
Center Japanese	-.09	-.03
Center social studies	-.07	-.07
Center math	-.10	-.06
Center science	.01	.02

** $p < .001$. $N = 153$.

The results of the multiple regression analysis are presented in Table 44. The results show that the set of HSGPA and ITP TOEFL scores accounted for a significant amount of variability in EAP GPA, $R^2 = .22$, $F(2, 150) = 20.50$, $p < .001$. A second analysis was performed to determine whether the Center Examination scores predicted EAP GPA over and above HSGPA and ITP TOEFL scores. The Center Examination scores did not account for a statistically significant amount of variability in EAP GPA, R^2 change = .01, $F(5, 145) = .30$, $p = .91$. Because TOEFL scores did not significantly predict EAP GPA in the first model, the regression analysis was repeated without TOEFL scores. These results are presented in Table 45. The results for both HSGPA and Center Examination section tests are similar.

Table 44. *Multiple Regression Results for Predicting EAP GPA from HSGPA, ITP TOEFL Scores, and Center Examination Section Test Scores (2005 to 2008)*

	B	SEB	β
Model 1			
Constant	.23	.67	
HSGPA	.49	.08	.44*
ITP TOEFL	.002	.00	.08
Model 2			
Constant	.19	.72	
HSGPA	.47	.08	.43*
ITP TOEFL	.002	.00	.10
Center English	-.01	.05	-.03
Center Japanese	-.01	.04	-.02
Center social studies	-.03	.04	-.06
Center math	-.03	.04	-.06
Center science	.01	.04	.02

Note. R^2 = .22 for Model 1; R^2 change = .01 for Model 2 (p = .91)
*p < .001.

Because the English section test of the Center Examination was changed in 2006 by the addition of a listening component, the analysis was repeated for students who matriculated in 2006 and after in order to determine whether the results for the English section test improved. In Table 46, the correlations and partial correlations of all predictors with EAP GPA are listed for classes entering from 2006 to 2008. Because the ITP TOEFL scores were statistically significant predictors of EAP GPA in answer to the first research question, they were retained in this analysis.

Table 45. *Multiple Regression Results for Predicting EAP GPA from HSGPA and Center Examination Section Test Scores (2005 to 2008)*

	B	SEB	β
Model 1			
Constant	.89	.35	
HSGPA	.51	.08	.46*
Model 2			
Constant	.93	.35	
HSGPA	.50	.08	.45*
Center English	.003	.04	.01
Center Japanese	-.02	.04	-.04
Center social studies	-.03	.04	-.06
Center math	-.02	.04	-.04
Center science	.004	.04	.01

Note. R^2 = .21 for Model 1; R^2 change = .01 for Model 2 (p = .93)
*p < .001.

Table 46. *The Bivariate and Partial Correlations of HSGPA, ITP TOEFL Scores, and Center Examination Section Test Scores (2006 to 2008) with EAP GPA.*

Predictor	Correlation with EAP GPA	Partial correlation with EAP GPA
HSGPA	.48**	.46
ITP TOEFL	.23	.13
Center English	.25*	.19
Center Japanese	-.06	.04
Center social studies	-.08	-.12
Center math	-.15	-.12
Center science	.16	.17

*p < .05. ** p < .001. N = 127.

The results of the multiple regression analysis using Center Examination scores from only 2006 to 2008 are presented in Table 47. HSGPA and ITP TOEFL scores accounted for a significant amount of variability in EAP GPA, R^2 = .25, F(2, 124) = 20.85, p < .001. A second analysis was performed to determine whether the Center Examination scores predicted EAP GPA over and above HSGPA and ITP

TOEFL scores. In this analysis, the Center Examination section scores accounted for a statistically significant amount of variability in EAP GPA, R^2 change = .07, $F(5, 119) = 2.41$, $p < .05$. These results indicate that the Center English section test showed an increase in predictive validity after the addition of the listening component in 2006. The Center science section test also improved to a nearly significant level, at $p = .06$.

Table 47. *Multiple Regression Results for Predicting EAP GPA from HSGPA, ITP TOEFL Scores, and Center Examination Section Test Scores (2006 to 2008)*

	B	SEB	β
Model 1			
Constant	-.41	.70	
HSGPA	.47	.08	.45**
ITP TOEFL	.00	.00	.16*
Model 2			
Constant	.07	.75	
HSGPA	.46	.08	.43**
ITP TOEFL	.002	.00	.12
Center English	.09	.04	.17*
Center Japanese	.02	.04	.04
Center social studies	-.05	.04	-.10
Center math	-.06	.04	-.11
Center science	.09	.05	.15

Note. R^2 = .25 for Model 1; R^2 change = .07 for Model 2 ($p < .05$)
*$p < .05$. **$p < .001$.

Hierarchical multiple regression analyses were next conducted to predict first-year GPA from HSGPA, the *hensachi* rankings, and Center Examination section test scores. The number of students in this analysis was 103, slightly less than the recommended 111; Field (2005) recommends a minimum sample size of 104 + k (the number of variables) (p. 173). In Table 48, the correlations and partial correlations of all predictors with first-year GPA are listed for classes entering from 2005 to 2008. HSGPA, the *hensachi* rankings, and the Center Examination math section test had statistically significant correlations with first-year GPA. After controlling for other predictors, each statistically significant correlation increased.

Table 48. *The Bivariate and Partial Correlations for HSGPA, Hensachi Rankings, and Center Examination Section Test Scores (2005 to 2008) with First-Year GPA*

Predictor	Correlation with first-year GPA	Partial correlation with first-year GPA
HSGPA	.29***	.40
Hensachi	.16**	.32
Center English	-.01	.07
Center Japanese	-.05	-.08
Center social studies	.02	.01
Center math	.14*	.23
Center science	-.02	-.03

*$p < .05$. ** $p < .01$.***$p < .001$. $N = 103$.

The results of the multiple regression analysis for the classes that entered the university from 2005 to 2008 are presented in Table 49. The results show that HSGPA and *hensachi* rankings accounted for a significant amount of variability in first-year GPA, $R^2 = .16$, $F(2, 100) = 9.39$, $p < .001$. A second analysis was performed to determine whether the set of Center Examination section test scores predicted first-year GPA over and above HSGPA and the *hensachi* rankings. In this second analysis, the Center Examination section test scores were not significant in predicting first-year GPA, over and above HSGPA and *hensachi* rankings, R^2 change = .05, ($F(5, 95) = 1.18$, $p = .33$. Nevertheless, the Center Examination math section test accounted for a statistically significant amount of variability in the first-year GPA.

Table 49. *Multiple Regression Results for Predicting First-Year GPA from HSGPA, Hensachi Rankings, and Center Examination Section Test Scores (2005 to 2008)*

	B	SEB	β
Model 1			
Constant	.07	.74	
HSGPA	.42	.11	.39***
Hensachi	.02	.01	.29**
Model 2			
Constant	-.18	.74	
HSGPA	.45	.11	.42***
Hensachi	.02	.01	.32**
Center English	.03	.05	.06
Center Japanese	-.04	.05	-.08
Center social studies	.01	.05	.01
Center math	.10	.04	.21*
Center science	-.02	.05	-.03

Note. R^2 = .16 for Model 1; R^2 change = .05 for Model 2 (p = .33).
*p < .05. **p < .01. ***p < .001.

The multiple regression analysis predicting first-year GPA from HSGPA, the *hensachi* rankings, and Center Examination section test scores was repeated with the Center Examination scores from 2006 to 2008, although the sample size was inadequate (N = 77). Although not presented in tabular form here, results indicated that the zero-order and partial correlations of the English, math, and science section tests were at a similar level of approximately .15, indicating the kinds of changes in the English and science section tests shown in Table 46. However, none of these correlations was statistically significant. In addition, none of the Center Examination section tests from 2006 to 2008 accounted for a statistically significant amount of variability in the first-year GPA.

The last analysis using the Center Examination section scores was performed to determine whether these scores added to the prediction of final GPA, over and above the two predictors (HSGPA and *hensachi* rankings) found significant in the answers to the first research question. In Table 50, the correlations and partial correlations of the predictor variables, HSGPA, the *hensachi* rankings, and Center Examination section test scores, with final GPA are listed. However, these results should be interpreted with caution because data from only 38 students were used. Because this sample size is inadequate, the results are presented here only for the sake of comparison.

Table 50. *The Bivariate and Partial Correlations for HSGPA, Hensachi Rankings and Center Examination Section Test Scores (2005 to 2008) with Final GPA*

Predictor	Correlation with final GPA	Partial correlation with final GPA
HSGPA	.35*	.47
Hensachi	.05	.24
Center English	.11	.12
Center Japanese	-.01	-.07
Center social studies	.25	.34
Center math	-.09	-.09
Center science	.15	-.02

Note. $*p < .01$. $N = 38$.

The regression results in Table 51 show that HSGPA and the *hensachi* rankings accounted for a significant amount of variability in final GPA, $R^2 = .16$,

$F(2, 35) = 3.40$, $p < .05$. A second analysis was performed to determine whether the Center Examination section test scores predicted final GPA over and above HSGPA and the *hensachi* rankings. The set of Center Examination section test scores was not significant in predicting final GPA, over and above HSGPA and the *hensachi* rankings, R^2 change = .12, ($F(5, 30) = 1.04$, $p = .41$. However, the social studies section test came close to statistical significance, at $p = .06$.

Because the *hensachi* rankings were not significant predictors of final GPA in the first model in Table 51, the regression analysis was repeated without the *hensachi* rankings. However, the results did not differ greatly from those presented in Table 51; accordingly, they are not presented here. In addition, the multiple

Table 51. *Multiple Regression Results for Predicting Final GPA from HSGPA, Hensachi Rankings, and Center Examination Section Test Scores (2005 to 2008)*

	B	SEB	β
Model 1			
Constant	1.15	.94	
HSGPA	.38	.15	.43*
Hensachi	.01	.01	.22
Model 2			
Constant	.89	.96	
HSGPA	.44	.15	.51**
Hensachi	.01	.01	.23
Center English	.03	.05	.11
Center Japanese	-.02	.06	-.06
Center social studies	.13	.07	.33
Center math	-.02	.05	-.07
Center science	-.01	.05	-.02

Note. R^2 = .16 for Model 1; R^2 change = .12 for Model 2 ($p = .41$).
*$p < .05$. **$p < .01$.

regression analysis predicting final GPA from HSGPA, the *hensachi* rankings, and the Center Examination section test scores from 2006 to 2008 was not repeated because the sample size was far from adequate ($N = 12$).

Research Question 6: Predicting EAP GPA, First-Year GPA, and Final GPA from Entrance Examination Type and Test Scores

The sixth research question asked to what extent type of entrance examination passed and entrance examination test scores (from English tests, Japanese tests, English essays, and oral interview) predict EAP GPA, as well as first-year and final GPA in the regular university program. To answer this research question, a series of hierarchical multiple regression analyses were conducted to predict EAP GPA, first-year GPA, and final GPA of graduating students from variables found to be significant in the answer to the first research question, together with (a) type of entrance examination passed (types A, B, and AO I and II) and (b) scores from entrance examination tests (English test, Japanese test, English essay, and oral interview scores). The results from the analysis concerning the type of entrance examination passed are presented first.

A hierarchical multiple regression analysis was conducted to predict EAP cumulative GPA from HSGPA, ITP total TOEFL score, and entrance examination passed. Of the 503 students who took one of these four entrance examinations and then matriculated from 2005 to 2008, 172 passed type A; 153 passed type B; 62 passed type AO I; and 116 passed type AO II. As explained in the Methods section, the A and B examinations are tests of Japanese and English reading, writing, and translation, while the AO I and II examinations require students to write an essay in English and participate in an interview in both Japanese and English. Type of entrance examination passed was dummy coded, and the AO I examination was used as a baseline. Dummy coding requires that one of the variables be regarded as a baseline and then excluded from the actual analysis; the results then are considered in reference to this variable. Usually this reference variable is the one believed most important, or the one that includes the majority of the participants (Field, 2005, pp. 208-209). In the case of the entrance examinations, however, the AO I type was chosen because the fewest participants had taken it (therefore allowing more students to remain in the analysis). To be sure that this choice of reference variable was appropriate, I repeated the analysis with the other tests as the reference variable and found that the results did not change appreciably. For the analysis, Types A, B, and AO II were entered as a block of variables, after HSGPA and the ITP TOEFL score.

The results of the regression analysis are presented in Table 52. The results show that HSGPA and ITP TOEFL total scores accounted for a significant amount of variability in EAP GPA, $R^2 = .11$, $F(2, 549) = 35.48$, $p < .001$. In the second analysis, type of entrance examination passed did not account for a statistically significant amount of variability in EAP GPA, over and above HSGPA and ITP total TOEFL scores, R^2 change = .01, $F(3, 546) = 1.23$, $p = .30$. Type of entrance examination was not a significant predictor of EAP GPA.

Table 52. *Multiple Regression Results for Predicting EAP GPA from HSGPA, ITP TOEFL Scores, and Type of Entrance Examination Passed*

	B	SEB	β
Model 1			
Constant	.44	.33	
HSGPA	.28	.04	.26*
ITP TOEFL	.003	.00	.21*
Model 2			
Constant	.54	.34	
HSGPA	.30	.05	.27*
ITP TOEFL	.003	.00	.20*
Examination A	-.10	.06	-.09
Examination B	-.10	.06	-.09
Examination AO II	-.06	.07	-.05

Note. R^2 = .11 for Model 1; R^2 change = .01 for Model 2 (p = .30). N = 552.
*p < .001.

A hierarchical multiple regression analysis was next conducted to predict first-year university GPA from HSGPA, the *hensachi* rankings, and type of entrance examination passed. The results are presented in Table 53. The regression analysis results showed that HSGPA and the *hensachi* rankings accounted for a significant amount of variability in first-year GPA, $R^2 = .14$, $F(2, 385) = 32.16$, $p < .001$. A second analysis was performed to determine whether type of entrance examination passed predicted first-year GPA over and above HSGPA and the *hensachi* rankings. In the second analysis, type of entrance examination passed did not account for a statistically significant amount of variability in first-year GPA over and above HSGPA and *hensachi* rankings, R^2 change = .01, ($F(3, 382) = .69$, $p = .56$.

Table 53. *Multiple Regression Results for Predicting First-Year GPA from HSGPA, Hensachi Rankings, and Type of Entrance Examination Passed*

	B	SEB	β
Model 1			
Constant	.44	.34	
HSGPA	.36	.05	.35*
Hensachi	.02	.00	.29*
Model 2			
Constant	.41	.35	
HSGPA	.38	.05	.36*
Hensachi	.02	.00	.29*
Examination A	.01	.07	.01
Examination B	-.05	.07	-.04
Examination AO II	-.08	.07	-.06

Note. R^2 = .14 for Model 1; R^2 change = .01 for Model 2 (p = .56). N = 388.
*p < .001.

For the last analysis involving type of entrance examination passed, a hierarchical multiple regression analysis was conducted to predict final GPA of graduating students from HSGPA, the *hensachi* rankings, and entrance examination passed. The results are presented in Table 54. The regression analysis results showed that HSGPA and the *hensachi* rankings accounted for a statistically significant amount of variability in final GPA, $R^2 = .13$, $F(2, 163) = 12.53$, $p < .001$. In the second analysis, type of entrance examination passed did not account for a statistically significant amount of variability in final GPA over and above HSGPA and the *hensachi* rankings, R^2 change = .01, $F(3, 160) = .88$, $p = .45$. Type of entrance examination passed did not predict GPA.

Table 54. *Multiple Regression Results for Predicting Final GPA from HSGPA, Hensachi Rankings, and Type of Entrance Examination Passed*

	B	SEB	β
Model 1			
Constant	1.50	.37	
HSGPA	.26	.06	.35**
Hensachi	.01	.00	.25*
Model 2			
Constant	1.62	.38	
HSGPA	.25	.06	.34**
Hensachi	.01	.00	.23*
Examination A	-.05	.07	-.07
Examination B	.05	.08	.06
Examination AO II	-.04	.08	-.06

Note. R^2 = .13 for Model 1; R^2 change = .01 for Model 2 (p = .45). N = 166.
*p < .01. **p < .001.

To answer the latter half of the sixth research question, a series of hierarchical multiple regression analyses were conducted to predict EAP GPA, first-year university GPA, and final GPA of graduating students from variables found to be significant in the answer to the first research question, together with scores from the entrance examination tasks (tests administered during the A and B entrance examinations, which included the English test and Japanese test scores; and the AO I and AO II tests, which included the 100-point essay and interview scores). The test scores were analyzed separately from the essay and interview scores because each student had scores for either the A and B examinations, or the essays and interviews of the AO examinations, but not both.

The English test scores and Japanese test scores were first examined to determine whether they added to the prediction of EAP, first-year, and final GPA, over and above predictors found useful in the answers to the first research question.

Because the scores for the English test and Japanese test were on different scales, they were converted to *z* scores for the analyses. In Table 55, the correlations and partial correlations of all predictors (HSGPA, ITP TOEFL, and the English and Japanese tests) with EAP GPA are listed. The variables HSGPA and ITP TOEFL had moderate, statistically significant correlations with EAP GPA, and remained at a similar level after controlling for other predictors. Neither the English nor Japanese test scores had a statistically significant relationship with EAP GPA.

Table 55. *The Bivariate and Partial Correlations for HSGPA, ITP TOEFL Scores, and Entrance Examination Test Scores with EAP GPA*

Predictor	Correlation with EAP GPA	Partial correlation with EAP GPA
HSGPA	.35**	.34
ITP TOEFL	.17*	.13
English test	.10	.07
Japanese test	.05	.05

Note. *$p < .05$. ** $p < .001$. $N = 324$.

The regression analysis results are presented in Table 56. HSGPA and ITP TOEFL scores accounted for a significant amount of variability in EAP GPA, $R^2 = .14$, $F(2, 321) = 26.11$, $p < .001$. In the second analysis, the entrance examination test scores did not account for a statistically significant amount of variability in EAP GPA, over and above HSGPA and TOEFL scores, R^2 change = .01, $F(2, 319) = 1.32$, $p = .27$. Neither the English test scores nor the Japanese test scores were significant predictors of EAP GPA.

Table 56. *Multiple Regression Results for Predicting EAP GPA from HSGPA, ITP TOEFL Scores, and Entrance Examination Test Scores*

	B	SEB	β
Model 1			
Constant	.44	.47	
HSGPA	.36	.06	.34***
ITP TOEFL	.002	.00	.14**
Model 2			
Constant	.52	.47	
HSGPA	.36	.06	.34***
ITP TOEFL	.002	.00	.13*
English test	.03	.03	.06
Japanese test	.02	.03	.05

Note. R^2 = .14 for Model 1; R^2 change = .01 for Model 2 (p = .27).
*p < .05. **p < .01. ***p < .001.

The English test and Japanese test scores were examined to determine whether they added to the prediction of first-year GPA, over and above HSGPA and the *hensachi* rankings. In Table 57, the correlation and partial correlation of each predictor (HSGPA, the *hensachi* rankings, and entrance examination test scores) with first-year GPA are listed. HSGPA had a moderate and statistically significant correlation with first-year GPA. The *hensachi* rankings and the English

Table 57. *The Bivariate and Partial Correlations for HSGPA, Hensachi Rankings, and Entrance Examination Test Scores with First-Year GPA*

Predictor	Correlation with first-year GPA	Partial correlation with first-year GPA
HSGPA	.30**	.34
Hensachi	.03*	.17
English test	.13*	.14
Japanese test	.06	.04

Note. *p < .05. **p < *.001*. N = 227.

test scores both reached statistical significance, ($p < .05$). The Japanese test scores did not demonstrate a statistically significant relationship with first-year GPA.

The regression analysis results are presented in Table 58. HSGPA and the *hensachi* rankings accounted for a significant amount of variability in first-year GPA, $R^2 = .11$, F(2, 224) = 13.91. The second model that included the English and Japanese test scores was significant in predicting first-year GPA, over and above HSGPA, R^2 change = .02, $F(2, 222) = 3.01$, $p = 05$. Although the Japanese test was not a significant predictor of first-year GPA, the English test reached a level of statistical significance similar to that of the *hensachi* rankings.

In the final analysis of the entrance examination test scores, the English test scores and Japanese test scores were examined to determine whether they added to the prediction of final GPA, over and above HSGPA and the *hensachi* rankings. Because the sample size was 94, and smaller than the recommended 107 (Field, 2005, p. 173), the results should be interpreted with caution. Table 59 presents the correlations and partial correlations of all predictors (HSGPA, the *hensachi* rankings, and entrance examination test scores) with final GPA. As was found with first-year GPA, HSGPA had a moderately large and statistically significant correlation with final GPA; however, the *hensachi* rankings, the English test scores, and the Japanese test scores did not have statistically significant relationships with the final GPA.

Table 58. *Multiple Regression Results for Predicting First-Year GPA from HSGPA, Hensachi Rankings, and Entrance Examination Test Scores*

	B	SEB	β
Model 1			
Constant	.95	.49	
HSGPA	.36	.07	.36**
Hensachi	.01	.01	.16*
Model 2			
Constant	.45	.54	
HSGPA	.37	.07	.37**
Hensachi	.01	.01	.17*
English test	.003	.00	.14*
Japanese test	.002	.00	.04

Note. R^2 = .11 for Model 1; R^2 change = .02 for Model 2 (p = .05).
*p < .05. **p < .001.

Table 59. *The Bivariate and Partial Correlations for HSGPA, Hensachi Rankings, and Entrance Examination Test Scores with Final GPA*

Predictor	Correlation with final GPA	Partial correlation with final GPA
HSGPA	.37*	.39
Hensachi	.01	.07
English test	-.04	-.09
Japanese test	-.14	-.10

Note. *p < .001. N = 94.

In the initial regression analysis, it was found that the *hensachi* rankings did not account for a statistically significant amount of variation in final GPA; therefore, the analysis was repeated without them. These regression analysis results are presented in Table 60. The results show that the second model, which includes the English and Japanese test scores, was not a significant predictor of final GPA, over and above HSGPA, R^2 change = .02, $F(2, 90) = 1.15$ (p = .32).

Table 60. *Multiple Regression Results for Predicting Final GPA from HSGPA And Entrance Examination Test Scores*

	B	SEB	β
Model 1			
Constant	2.11	.32	
HSGPA	.29	.07	.37*
Model 2			
Constant	2.09	.33	
HSGPA	.30	.08	.39*
English test	-.04	.05	-.09
Japanese test	-.04	.04	-.10

Note. R^2 = .14 for Model 1; R^2 change = .02 for Model 2 (p = .32).
*p < .001.

The second set of entrance examination scores examined was the English essay and interview scores. These scores were converted to *z* scores and entered in a regression analysis to determine whether they could add to the prediction of EAP, first-year, and final GPA, over and above predictors found useful in the answers to the first research question. In Table 61, the correlations and partial correlations of the predictor variables, HSGPA, ITP TOEFL, and entrance examination scores from English essays and interviews, with EAP GPA are listed. HSGPA, TOEFL,

Table 61. *The Bivariate and Partial Correlations for HSGPA, ITP TOEFL Scores, and Entrance Examination Essay and Interview Scores with EAP GPA*

Predictor	Correlation with EAP GPA	Partial correlation with EAP GPA
HSGPA	.10*	.16
ITP TOEFL	.33**	.28
English essay	.30**	.29
Interview	.20	.13

Note. *p < .05. ** p < .001. N = 174.

and the English essay scores all showed statistically significant correlations with the EAP GPA. The interview score was close to reaching statistical significance, at $p = .08$.

Table 62 shows the results of the hierarchical multiple regression analysis. Initially, HSGPA and ITP TOEFL total scores accounted for a significant amount of variability in EAP GPA, $R^2 = .12$, $F(2, 171) = 11.65$, $p < .001$. The second analysis included the entrance examination English essay and interview scores. As a set, these scores accounted for a significant amount of variability in EAP GPA, R^2 change = .09, $F(2, 169) = 9.55$, $p < .001$; however, the two scores were not equally related to the EAP GPA. The essay scores had a statistically significant relationship with EAP GPA, but the interview scores did not quite reach statistical significance ($p = .08$).

Table 62. *Multiple Regression Results for Predicting EAP GPA from HSGPA, ITP TOEFL Scores, and Entrance Examination Essay and Interview Scores*

	B	SEB	β
Model 1			
Constant	.71	.55	
HSGPA	.13	.09	.11
ITP TOEFL	.004	.00	.33**
Model 2			
Constant	1.08	.55	
HSGPA	.17	.09	.14*
ITP TOEFL	.003	.00	.26**
English essay	.35	.09	.28**
Interview	.06	.04	.12

Note. $R^2 = .12$ for Model 1; R^2 change = .09 for Model 2 ($p < .001$).
*$p < .05$. ** $p < .001$.

The English essay and interview scores were next examined to determine whether they added to the prediction of first-year GPA, over and above predictors found useful in the answers to the first research question (HSGPA and the *hensachi* rankings). In Table 63, the correlations and partial correlations of the predictor variables, HSGPA, the *hensachi* rankings, and entrance examination English essays and interviews, with first-year GPA are listed. HSGPA and the *hensachi* rankings had moderately large, statistically significant correlations with first-year GPA, but neither the English essay scores nor the interview scores reached statistical significance in their relationship with first-year GPA.

Table 63. *The Bivariate and Partial Correlations for HSGPA, Hensachi Rankings, and Entrance Examination Essay and Interview Scores with First-Year GPA*

Predictor	Correlation with first-year GPA	Partial correlation with first-year GPA
HSGPA	.21*	.35
Hensachi	.34*	.42
English essay	.14	.11
Interview	.12	.08

Note. $^*p < .001$. $N = 133$.

The regression analysis results, presented in Table 64, showed that HSGPA and the *hensachi* rankings accounted for a significant amount of variability in first-year GPA, $R^2 = .22$, $F(2, 130) = 18.46$. The second model including the English essay and interview scores was not significant in predicting first-year GPA, over and above HSGPA and the *hensachi* rankings, R^2 change = .02, $F(2, 128) = 1.63$, $p = .20$.

Table 64. *Multiple Regression Results for Predicting First-Year GPA from HSGPA, Hensachi Rankings, and Entrance Examination Essay and Interview Scores*

	B	SEB	β
Model 1			
Constant	-.01	.54	
HSGPA	.38	.09	.34*
Hensachi	.03	.01	.44*
Model 2			
Constant	.07	.54	
HSGPA	.38	.09	.35*
Hensachi	.03	.01	.43*
English essay	.15	.12	.11
Interview	.03	.04	.08

Note. R^2 = .22 for Model 1; R^2 change = .02 for Model 2 (p = .20).
*$p < .001$.

The last analysis using the English essay and interview scores was performed to determine whether these scores added to the prediction of final GPA, over and above the predictors found useful in the answers to the first research question (HSGPA and the *hensachi* rankings). In Table 65, the correlations and partial correlations of the predictor variables, HSGPA, the *hensachi* rankings, and entrance examination scores from English essays and interviews, with final GPA are shown. However, these results should be interpreted with caution because sample size was only 53. Because this sample size is inadequate, the results are presented here only for the sake of comparison. HSGPA and the *hensachi* rankings had statistically significant correlations with final GPA, but neither the English essay scores nor the interview scores reached statistical significance in their relationship with final GPA.

Table 65. *The Bivariate and Partial Correlations for HSGPA, Hensachi Rankings, and Entrance Examination Essay and Interview Scores with Final GPA*

Predictors	Correlation with final GPA	Partial correlation with final GPA
HSGPA	.15*	.30
Hensachi	.27**	.37
English essay	-.05	-.02
Interview	.08	.09

Note. $*p < .05$. $**p < .01$. $N = 53$.

Table 66 presents the regression analysis results for predicting final GPA from HSGPA, the *hensachi* rankings, and entrance examination English essay and interview scores. The regression analysis showed that HSGPA and the *hensachi* rankings accounted for a significant amount of variability in final GPA, $R^2 = .16$, $F(2, 50) = 4.67$, $p < .05$. The analysis showed that the second model, which

Table 66. *Multiple Regression Results for Predicting Final GPA from HSGPA, Hensachi Rankings, and Entrance Examination Essay and Interview Scores*

	B	SEB	β
Model 1			
Constant	1.05	.76	
HSGPA	.28	.12	.33*
Hensachi	.02	.01	.41**
Model 2			
Constant	1.05	.78	
HSGPA	.27	.13	.32*
Hensachi	.02	.01	.41**
English essay	-.03	.23	-.02
Interview	.03	.05	.08

Note. $R^2 = .17$ for Model 1; R^2 change = .01 for Model 2 ($p = .81$).
$*p < .05$. $**p < .01$.

included English essay and interview scores, was not significant in predicting final GPA, over and above HSGPA and the *hensachi* rankings, R^2 change = .01, $F(2, 48) = .21$, $p = .81$. Though both scores were not significant predictors of final GPA, the interview scores exhibited a stronger relationship with final GPA.

Research Question 7: Predicting Finishing the EAP Program on Time (FOT) from HSGPA, High School Grade Factor Scores, ITP TOEFL Scores, *Hensachi* Rankings, Gender, and Parents' Education

The seventh research question asked to what extent HSGPA, high school grade factor scores, ITP TOEFL scores, the *hensachi* rankings, and the non-cognitive variables, gender and parental education, predict finishing the EAP program on time (FOT). In order to answer this research question, the variables HSGPA (and high school factor scores), ITP TOEFL scores, the *hensachi* rankings, and the non-cognitive variables, gender, father's education, and mother's education, were examined in a logistic regression analysis that predicted FOT. Because HSGPA, the ITP TOEFL scores, and the *hensachi* rankings have been shown by previous research and/or by results of this study to predict academic performance in a variety of ways, they were first entered as a block of variables into the logistic regression analysis. The three non-cognitive variables, gender, father's education, and mother's education, have also been shown to be linked by previous research, and by results of this study, to school achievement in a variety of ways; therefore, these three non-cognitive variables were examined as a second block of variables to determine whether they were also significant predictors of FOT. After the first analysis with HSGPA, the HSGPA variable was replaced in a second analysis of FOT with high school grade factor scores. The data of students who entered university from 2004 to 2008 were used so that the sample size would be adequate in the following analyses.

The results of the logistic regression analysis for FOT are presented in Table 67. The Hosmer-Lemeshow goodness of fit statistic was nonsignificant ($p = .18$). According to Field (2005) this non-significant result is desirable because it indicates that the model predicts real world data reasonably well (p. 254). In addition, the model chi-square statistic was significant. As Field explained, a significant model chi-square statistic is desirable because it indicates that a model that includes the predictors is significantly better than a model with only a constant (p. 237), in other words, a significant model chi-square indicates that the model is a significant fit to the data (p. 244).

Significant predictors of finishing the EAP program on time, in the order of most impact, are: HSGPA, father's education, and the ITP TOEFL score, as shown in Table 67. The odds ratios indicate that as these variables increase, the odds of students finishing the EAP program on time also increase. In addition, because the odds ratio confidence intervals for these variables do not cross 1, the interpretation of these results is reliable. Though the odds ratio and confidence interval values of the *hensachi* rankings suggest that they might also predict FOT, the confidence intervals for it and all other variables cross 1, making interpretation unreliable.

Table 67. *Summary of Logistic Regression Analysis Predicting FOT from HSGPA, ITP TOEFL Scores, Hensachi Rankings, Gender, and Parents' Education*

Variable	*B*	*SE*	Wald statistic	Odds ratio	95% CI
HSGPA	1.26*	.50	6.30	3.52	[1.32, 9.38]
ITP TOEFL	.03**	.01	20.98	1.03	[1.02, 1.04]
Hensachi	.05	.03	2.39	1.05	[0.99, 1.11]
Gender (female)	-.29	.47	.37	.75	[0.30, 1.89]
Father's education	.27*	.11	6.44	1.31	[1.06, 1.61]
Mother's education	-.06	.13	.20	.94	[0.73, 1.22]
Constant	-21.87**	4.73	21.38	.00	

Note. R^2 = .11 (Cox & Snell), .24 (Nagelkerke). Model χ^2 = 42.37, p < .001.
N = 366.
*p < .05. **p < .001.

In the second analysis of FOT, high school factor scores are used instead of HSGPA. The results of this second logistic regression analysis are presented in Table 68. The Hosmer-Lemeshow goodness of fit statistic was not significant (p = .36); this result indicates that the model predicts the real world data. The model chi-square statistic was significant; therefore, the model displayed significant fit to the data.

Significant predictors of finishing the EAP program on time, in order of most impact, are: the Numerical Ability factor score, father's education, and ITP TOEFL, as shown in Table 68. The odds ratios indicate that as these variables increase, the odds of students finishing EAP on time also increase. In addition, because the confidence intervals for the odds ratio for these variables do not cross 1, the interpretation of these results is reliable. The *hensachi* rankings, though not statistically significant, also might predict finishing EAP on time (though this interpretation might be unreliable because the lower confidence interval value is .99). Other variables were not statistically significant; in addition, their confidence intervals cross 1, making interpretation unreliable.

Table 68. *Summary of Logistic Regression Analysis Predicting FOT from High School Grade Factor Scores, ITP TOEFL Scores, Hensachi Rankings, Gender, and Parents' Education*

Variable	*B*	*SE*	Wald statistic	Odds ratio	95% CI
Language ability factor	.24	.28	.71	1.27	[0.73, 2.21]
Numerical ability factor	.65*	.27	5.69	1.91	[1.12, 3.24]
Spatial ability factor	.33	.32	1.10	1.39	[0.75, 2.59]
ITP TOEFL	.03***	.01	22.20	1.03	[1.02, 1.04]
Hensachi	.06	.03	3.15	1.06	[0.99, 1.13]
Gender (female)	-.50	.50	1.02	.61	[0.23, 1.60]
Father's education	.29***	.11	6.84	1.34	[1.08, 1.67]
Mother's education	-.10	.14	.51	.91	[0.69, 1.19]
Constant	-17.52***	3.89	20.27	.00	

Note. R^2 = .12 (Cox & Snell), .26 (Nagelkerke). Model χ^2 = 45.52, p < .001.
N = 346.
*p < .05. **p < .01. ***p < .001.

Research Question 8: Predicting Graduation on Time (GOT) and Graduation with Honors (GWH) from HSGPA, High School Grade Factor Scores, ITP TOEFL Scores, *Hensachi* Rankings, Gender, and Parents' Education

The eighth research question asked to what extent HSGPA, high school grade factor scores, ITP TOEFL scores, the *hensachi* rankings, and the non-cognitive variables, gender and parents' education, predict university graduation on time (GOT) (in four years) and graduation with honors (GWH). In order to answer this research question, HSGPA (and high school factor scores), ITP TOEFL scores, the *hensachi* rankings, and the non-cognitive variables, gender, father's education, and mother's education, were examined in a logistic regression analysis that predicts both GOT and GWH. Because HSGPA, ITP TOEFL scores, and the *hensachi* ratings have been shown by previous research and/or by results of this study to predict academic performance in a variety of ways, they were first entered as a block of variables into the logistic regression analysis. The three non-cognitive variables have also been shown to be linked by previous research, and by results of this study, to school achievement in a variety of ways; therefore, these three non-cognitive variables are examined as a second block of variables to determine whether they also are linked to GOT and GWH. After the first analysis with HSGPA, the HSGPA variable was replaced in a second analysis of GOT with high school grade factor scores. Afterwards, GWH was examined in the same manner. The data of students who entered the university from 2004 to 2008 were used so the sample size was adequate in the following analyses.

The results of the logistic regression analysis for GOT are presented in Table 69. The Hosmer-Lemeshow goodness of fit statistic was nonsignificant ($p = .59$; this result indicated that the model predicts real-world data reasonably well. In addition, the model chi-square statistic was significant, as noted; therefore, the model showed significant fit to the data. The only significant predictor of graduating on time was the ITP TOEFL score. The odds ratio indicated that as students' scores on the TOEFL increase, the odds of students graduating on time also increased. In addition, because the confidence interval for the odds ratio for the TOEFL did not cross 1, the interpretation of these results is reliable.

Table 69. *Summary of Logistic Regression Analysis Predicting GOT from HSGPA, ITP TOEFL Scores, Hensachi Rankings, Gender, and Parents' Education*

Variable	*B*	*SE*	Wald statistic	Odds ratio	95% CI
HSGPA	.72	.50	2.09	2.06	[0.77, 5.50]
ITP TOEFL	.02*	.01	11.67	1.02	[1.01, 1.04]
Hensachi	.04	.03	1.53	1.04	[0.98, 1.11]
Gender (female)	-.60	.45	1.75	.55	[0.23, 1.34]
Father's education	.05	.11	.22	1.05	[0.85, 1.30]
Mother's education	-.22	.13	2.97	.80	[0.62, 1.03]
Constant	-14.13*	4.61	9.49	.00	

Note. R^2 = .17 (Cox & Snell), .23 (Nagelkerke). Model χ^2 = 25.34, p < .001.
N = 134.
*p < .01.

In the second analysis of GOT, high school factor scores were used instead of HSGPA. The results of this second logistic regression analysis are presented in Table 70. The Hosmer-Lemeshow goodness of fit statistic was significant (p = .01); this result indicates some difficulty in predicting the real-world data. The model chi-square statistic was significant, as noted; therefore, the model showed a significant fit to the data.

The only significant predictor of graduating on time was the ITP TOEFL, as noted in Table 70. The odds ratio indicates that as students' scores on the TOEFL increased, the odds of graduating on time also increased. In addition, because the confidence interval for the odds ratio for TOEFL did not cross 1, the interpretation of these results is reliable. All other variables were not statistically significant predictors of GOT.

Table 70. *Summary of Logistic Regression Analysis Predicting GOT from High School Grade Factor Scores, ITP TOEFL Scores, Hensachi Rankings, Gender, and Parents' Education*

Variable	*B*	*SE*	Wald statistic	Odds ratio	95% CI
Language ability factor	.18	.28	.41	1.20	[0.69, 2.08]
Numerical ability factor	.21	.25	.70	1.24	[0.75, 2.03]
Spatial ability factor	.36	.36	.98	1.43	[0.70, 2.93]
ITP TOEFL	.03*	.01	12.48	1.03	[1.01, 1.04]
Hensachi	.04	.04	.92	1.04	[0.96, 1.11]
Gender (female)	-.39	.50	.62	.68	[0.25, 1.79]
Father's education	.07	.12	.34	1.07	[0.85, 1.34]
Mother's education	-.25	.14	3.27	.78	[0.60, 1.02]
Constant	-11.58	3.98	8.45	.00	

Note. R^2 = .18 (Cox & Snell), .23 (Nagelkerke). Model $\chi^2 = 24.40$, $p < .01$.
$N = 127$.
*$p < .001$.

To answer the second part of the eighth research question, graduation with honors (GWH) was examined using logistic regression. The results for GWH are presented in Table 71. The Hosmer-Lemeshow goodness of fit statistic was nonsignificant ($p = .92$); this result shows that the model predicts real world data well. In addition, the model chi-square statistic was significant; therefore, the model displayed significant fit to the data.

The significant predictors of graduating with honors were HSGPA and the *hensachi* rankings (see Table 71). HSGPA was overwhelmingly the best predictor. Interpretations concerning both HSGPA and the *hensachi* ratings are reliable because their confidence intervals did not cross 1. All other variables were not statistically significant predictors of GWH.

Table 71. *Summary of Logistic Regression Analysis Predicting GWH from HSGPA, ITP TOEFL Scores, Hensachi Rankings, Gender, and Parents' Education*

Variable	*B*	*SE*	Wald statistic	Odds ratio	95% CI
HSGPA	1.49*	.66	5.11	4.42	[1.22, 16.03]
ITP TOEFL	.01	.01	2.42	1.01	[1.00, 1.02]
Hensachi	.10*	.04	5.85	1.10	[1.02, 1.19]
Gender (female)	.17	.53	.10	1.19	[0.42, 3.36]
Father's education	.12	.13	1.88	1.13	[0.88, 1.45]
Mother's education	-.13	.16	.67	.88	[0.64, 1.20]
Constant	-20.20**	5.52	13.40	.00	

Note. R^2 = .04 (Cox & Snell), .11 (Nagelkerke). Model χ^2 = 13.83, p < .05.
N = 366.
*p < .05. **p < .001.

In the next analysis of GWH, high school factor scores were used instead of HSGPA. The results of this second logistic regression analysis for GWH are presented in Table 72. The Hosmer-Lemeshow goodness of fit statistic was not significant (p = .64); this result indicates that the model predicts the real world data. The model chi-square statistic was not statistically significant (p = .07); therefore, the model did not show significant fit to the data.

Significant predictors of graduating with honors were the Language Ability factor score, ITP TOEFL score, and the *hensachi* ranking (see Table 72). The Language Ability factor score was the most predictive of the three variables. The interpretation for all three predictors was reliable because their confidence intervals did not cross 1. Other variables were not statistically significant predictors of GWH.

Table 72. *Summary of Logistic Regression Analysis Predicting GWH from High School Grade Factor Scores, ITP TOEFL Scores, Hensachi Rankings, Gender, and Parents' Education*

Variable	*B*	*SE*	Wald statistic	Odds ratio	95% CI
Language ability factor	.75*	.39	3.92	2.16	[1.01, 4.62]
Numerical ability factor	.42	.35	1.46	1.53	[0.77, 3.05]
Spatial ability factor	-.22	.39	.32	.80	[0.37, 1.72]
ITP TOEFL	.01*	.01	4.11	1.02	[1.00, 1.03]
Hensachi	.08*	.04	3.89	1.09	[1.00, 1.18]
Gender (female)	.07	.60	.04	1.12	[0.35, 3.61]
Father's education	.05	.13	.06	1.03	[0.80, 1.34]
Mother's education	-.08	.17	.25	.92	[0.66, 1.29]
Constant	-14.61**	4.57	10.25	.00	

Note. R^2 = .04 (Cox & Snell), .12 (Nagelkerke). Model χ^2 = 14.39, p = .07.
N = 346.
*p < .05. **p < .01.

Research Question 9: Combining HSGPA, ITP TOEFL Total Scores, *Hensachi* Rankings, EAP GPA, First-Year GPA, and Final GPA in Path Models

The ninth research question asked how non-cognitive variables, HSGPA, ITP TOEFL scores, the *hensachi* rankings, Center Examination scores, and university entrance examination scores interrelate in a path analysis and contribute to academic performance in the EAP program and in the first year of regular university study over time. It was originally hypothesized that: (a) non-cognitive variables contribute to performance in the HSGPA; (b) non-cognitive variables contribute to performance in the EAP and first-year GPA; (c) HSGPA contributes to EAP and first-year GPA; (d) TOEFL scores, Center Examination English section scores, and university examination English essay scores contribute to the EAP GPA; (e) TOEFL scores, EAP GPA, university examination essay scores, and the *hensachi* rankings contribute to academic performance in the first-year university GPA, and constitute parts of a well-fitting path model. In order to answer this research question, HSGPA, ITP TOEFL scores, the *hensachi* rankings, EAP GPA, first-year GPA, and final GPA, were examined through a path analysis conducted with EQS version 6.1 (Bentler, 2005). The non-cognitive variables, Center Examination English scores, and university examination essay scores, which were parts of the original hypothesized model, were not included in the analysis because doing so would have reduced the sample size to an unacceptably low level. Instead of the originally proposed assortment of variables only cognitive variables related to academic achievement were included. This more specific focus for the path analysis is more appropriate for this study, which is the first to examine these variables.

The hypothesized model examined the predictors of academic achievement in university in terms of EAP GPA, first-year GPA, and final GPA. It was hypothesized that HSGPA would predict university achievement because this variable has been repeatedly shown to do so (e.g., Astin, 1993; Burton & Ramist, 1994; Zwick, 2002); that TOEFL would predict EAP GPA because the results of this study, as well as those of a number of other studies have indicated that TOEFL is a significant predictor of achievement in university, especially in EAP programs (Johnson, 1988; Light, Xu, & Mossop, 1987; Vinke & Jochems, 1993); that the *hensachi* rankings would predict first-year GPA because the results of this study have shown them to do so; and that all previous levels of university GPA would predict later university GPAs (in this case, the earlier and separate EAP GPA should predict the first-year university GPA) because the work of researchers studying GPA has indicated that this relationship exists (Werts, Linn & Jöreskog, 1978; Young, 2004). The hypothesized path model is shown in Figure 3.

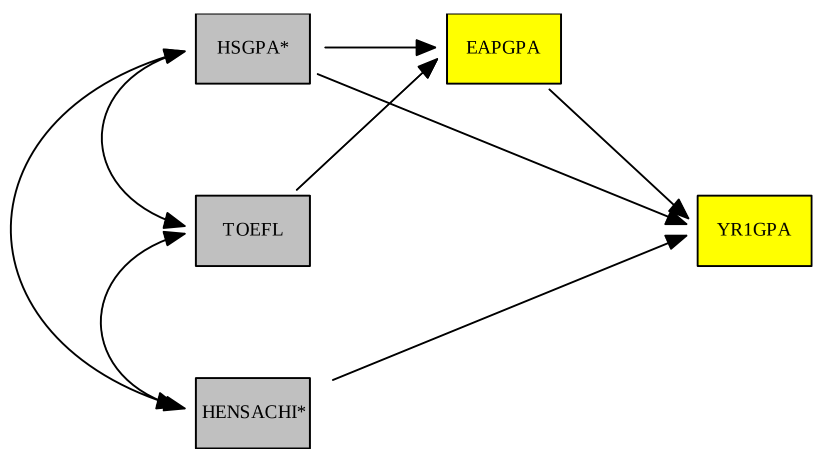

Figure 3. Hypothesized path model of academic achievement up to first-year GPA.

Table 73 presents the correlations among the variables used in the path analysis for first-year GPA. Measures that were expected to correlate significantly generally did so: HSGPA correlated with EAP GPA and first-year GPA; the *hensachi* rankings correlated with first-year GPA; the ITP TOEFL scores correlated with EAP GPA (and first-year GPA); and EAP GPA correlated with first-year GPA. It was also not surprising that HSGPA and *hensachi* rankings correlated negatively because, as explained earlier, students who were admitted to the university under study were expected to have high HSGPAs if they came from high schools that had relatively low *hensachi* rankings, while students from highly ranked high schools were still able to be admitted with relatively low HSGPAs. The variables that did not correlate significantly were not expected to do so (e.g., HSGPA and ITP TOEFL; EAP GPA and the *hensachi* rankings). The correlations shown in Table 73 suggest that the relationships specified in the path model are plausible.

Table 73. *Correlation Matrix of Variables Used in Path Analysis of First-Year GPA*

Measure	1	2	3	4	5
1. HSGPA	---				
2. *Hensachi*	-.37*	---			
3. ITP TOEFL	-.02	.13*	---		
4. EAP GPA	.24*	-.04	.07	---	
5. YR 1 GPA	.27**	.14*	.09*	.49**	---

Note. *$p < .05$. **$p < .01$. $N = 524$.

In order to evaluate the fit of a path model, Byrne (2006) discussed a number of fit indices, but focused on three. These fit indices follow, along with the values for each that indicate a well-fitting model: Comparative Fit Index (CFI) (> .95), Standardized Root Mean Square Residual (SRMR) (< .05), and Root Mean Square Error of Approximation (RMSEA) and its 90% confidence interval (< .05 suggests good fit, and values up to .08 are acceptable) (pp. 94-101).

The hypothesized model displayed good fit to the data, minimum fit function chi-square = 2.37, $p = .31$; CFI = 1.0; SRMR = .02; RMSEA = .02 (confidence interval, .00 to .09), although the confidence interval extends slightly beyond .08. The final model and path coefficients are presented in Figure 4.

Figure 5 presents the hypothesized path model for final GPA. It was hypothesized that HSGPA would predict all levels of university achievement because this variable has been repeatedly shown to do so (e.g., Astin, 1993; Burton & Ramist, 1994; Zwick, 2002). The ITP TOEFL total score was hypothesized to predict EAP GPA because the results of this study, as well as those of a number of previous studies have indicated that the TOEFL is a significant predictor of university GPA, especially in EAP programs (Johnson, 1988; Light, Xu, & Mossop, 1987; Vinke & Jochems, 1993). It was also hypothesized that the *hensachi* rankings would predict first-year GPA and final GPA because the results of this study have shown them to be significant predictors of university achievement, and all previous levels of university GPA predict later university GPAs (e.g., Werts, Linn & Jöreskog, 1978; Young, 2004).

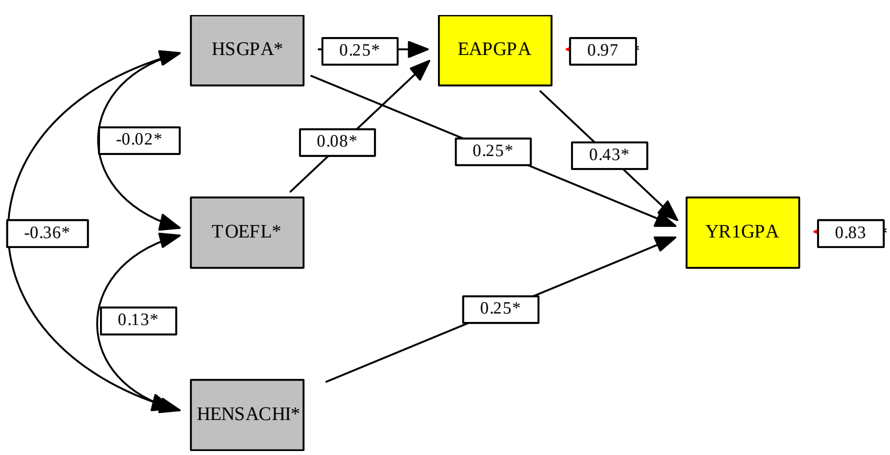

Figure 4. Standardized solution for the hypothesized path model of academic achievement up to first-year GPA.

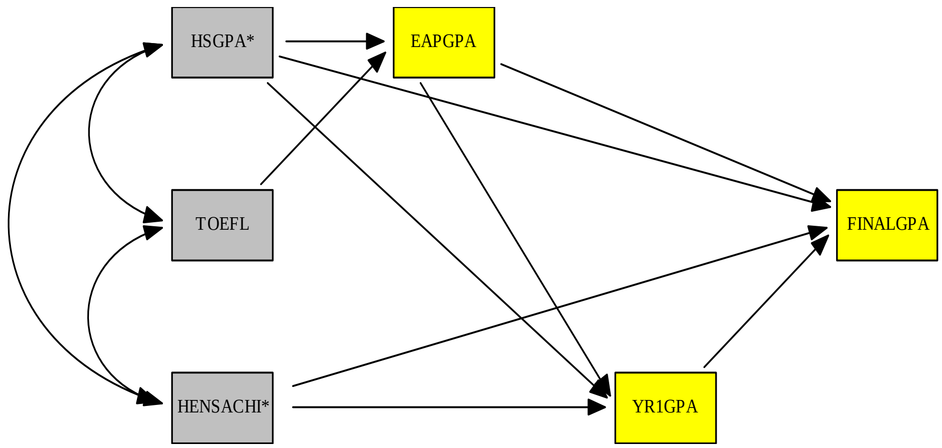

Figure 5. Hypothesized path model of academic achievement up to final GPA with two *hensachi* ranking paths.

Table 74 presents the correlations among the variables used in the path analysis for academic achievement up to the final GPA. As was found with the correlations among the variables for the first path analysis, measures that were expected to correlate significantly did so: HSGPA correlated significantly with EAP GPA and first-year GPA, *hensachi* rankings correlated with first-year GPA, ITP TOEFL correlated with EAP GPA (and first-year GPA); the earlier and separate EAP GPA correlated with first-year GPA. Those variables that did not correlate significantly were not expected to do so (e.g., HSGPA and ITP TOEFL; EAP GPA and *hensachi* rankings). Although *hensachi* rankings did not correlate significantly with final GPA, the path from *hensachi* rankings to final GPA was left in the model, based on results from the hierarchical multiple regression results of this study (see Tables 18 and 19).

Table 74. *Correlation Matrix of Variables Used in the Path Analysis of Final GPA*

Measure	1	2	3	4	5	6
1. HSGPA	---					
2. *Hensachi*	-.38*	---				
3. ITP TOEFL	.01	.13*	---			
4. EAP GPA	.21*	-.01	.09*	---		
5. YR 1 GPA	.28**	.13*	.18*	.39**	---	
6. Final GPA	.30**	.09	.19**	.45**	.73**	---

Note. *$p < .05$. **$p < .01$. $N = 295$.

The hypothesized model for academic achievement up to the final GPA showed good fit to the data, minimum fit function chi-square = 8.93, $p = .03$; CFI = .99; SRMR = .04; RMSEA = .08 (confidence interval, .01 to .15). The model path coefficients are presented in Figure 6. However, because the RMSEA was marginal, and the confidence interval for RMSEA exceeded the acceptable level of .08, the model was reconsidered. In light of the correlations in Table 74, the path from *hensachi* rankings to final GPA was deleted, and a path from ITP TOEFL to first-year GPA was added. Although ITP TOEFL scores also correlated significantly with final GPA and were shown to relate to graduating on time (see Tables 69 and 70), it was decided to add only one path from ITP TOEFL to first-year GPA, in keeping with the theory used throughout the study, that English proficiency is related most to success early in the student's university career. The revised hypothesized model for academic achievement up to the final GPA is presented in Figure 7.

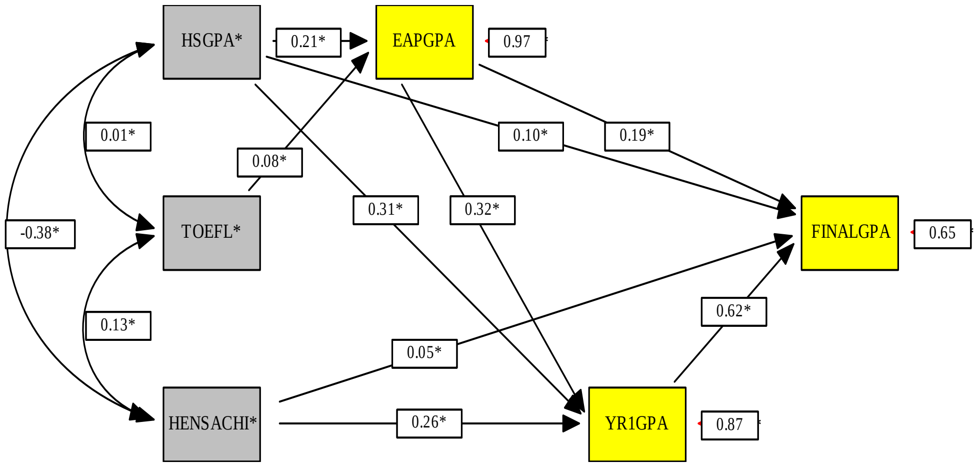

Figure 6. Standardized solution for the hypothesized path model of academic achievement up to final GPA with two *hensachi* ranking paths.

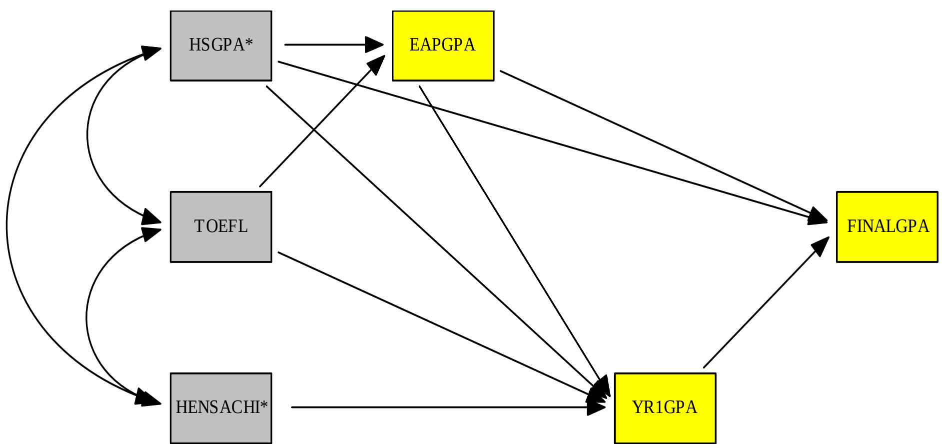

Figure 7. Revised hypothesized path model of academic achievement up to final GPA with two ITP TOEFL paths.

The revised hypothesized model for academic achievement through the final GPA showed improved fit to the data, minimum fit function chi-square = 4.69, p = .20; CFI = 1.0; SRMR = .02; RMSEA = .04 (confidence interval, .00 to .12). All indices improved for the revised model and indicate good fit. The confidence interval for RMSEA also improved, but still extended beyond .08. The final model and the path coefficients are presented in Figure 8.

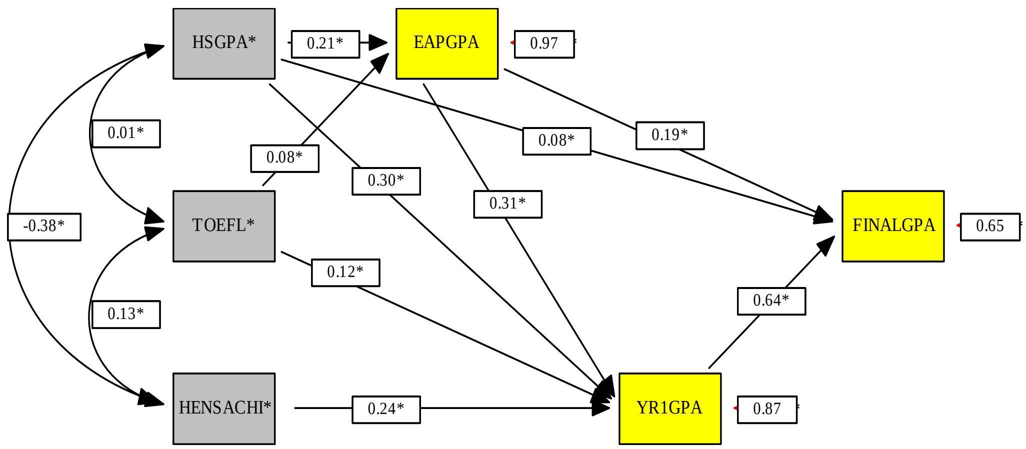

Figure 8. Standardized solution for the revised hypothesized path model of academic achievement up to final GPA with two ITP TOEFL paths.

CHAPTER 5

DISCUSSION

Research Question 1: Predicting EAP GPA, First-Year GPA, and Final GPA from HSGPA, ITP TOEFL Scores, and *Hensachi* Rankings

The first research question asked to what extent the variables of central interest in this study, HSGPA, ITP TOEFL scores obtained in April at the start of students' first year, and *hensachi* rankings, predict EAP GPA, first-year GPA, and final GPA in the regular university program. Based on previous research, it was hypothesized that HSGPA and ITP TOEFL scores predict the students' performance in the EAP program, as well as first-year GPA and final GPA in the regular university program. In addition, based on research indicating that level of high school predicts university achievement, it was hypothesized that *hensachi* rankings are not significantly related to the EAP GPA, but do predict first-year and final GPAs.

Prediction from HSGPA

The results confirm the hypothesis that HSGPA is a significant predictor of EAP GPA, first-year GPA, and final GPA. In order to examine the results in terms of past studies, similar statistics must be compared. Multiple correlations have often been presented in studies of HSGPA, and these correlations were sometimes presented as unadjusted; at other times, both unadjusted and adjusted correlations were presented. In order to compare relatively equivalent correlations, unadjusted correlations are compared. The studies presented in Table 1 reported unadjusted correlations of first-year GPA with HSGPA of .39 and .36. In the current study, the unadjusted correlations of EAP GPA and first-year GPA with HSGPA are .27 and .26 ($p < .001$), respectively. The results for these two GPAs are similar, perhaps reflecting the academic nature of the EAP program, and its similarity to the regular university program. These two correlations are both lower than those shown in Table 1. However, it should be recalled that the studies presented in Table 1 were large-scale, national studies of the predictive validity of the SAT. Ramist, Lewis, and McCamley-Jenkins (1994) studied over 40,000 students from 30 colleges, while Kobrin, Patterson, Shaw, Mattern, and Barbuti (2008) studied more than 150,000 students from 110 colleges. Clearly, such samples would have a much larger range in HSGPA, test scores, and university GPA, and higher resulting correlations than a sample from one university.

A study conducted by Hoffman and Lowitzki (2005) serves to demonstrate the point that small-scale studies of one university tend to produce smaller correlations than large-scale studies. These researchers studied 522 students at one university and found an overall correlation of .30 between HSGPA and first-year university GPA, a result similar to those found in the current study. In short, the results of this study are like those of similar smaller scale studies in that the correlations found are conservative estimates of the actual relationship between HSGPA and university GPA, which can be more accurately determined in large-scale studies. Nevertheless, even though the correlations found in this study between EAP GPA and first-year GPA with HSGPA are conservative estimates, they are statistically significant and indicate that HSGPA accounts for a portion of later academic success, both in an EAP program and in the regular university program.

The results of this study can also be compared to those of previous studies of final GPA (see Table 2). Average unadjusted correlations of HSGPA with final GPA were .49 in the 1930 to 1980 studies, and .42 in the 1980 to mid-1990 studies. In the current study, the unadjusted correlation of final GPA with HSGPA is .28 ($p < .001$). This result confirms the hypothesis regarding final GPA, but the correlation itself is lower than those reported in Table 2. However, it should be recalled that the correlations presented in Table 2 are average weighted correlations; these average weighted correlations were derived from eleven studies and more than 9,000 students (for the 1930 to 1980 data) and twelve studies and more than 25,000 students (for data gathered after 1980). The correlations found in individual studies in each group varied, and the smallest correlation found in studies before 1980 was .34, and after 1980, .30. These correlations are very close to the .28 correlation obtained in this study.

Although the correlations of HSGPA with EAP GPA, first-year GPA, and final GPA confirm the hypothesis that HSGPA predicts short- and long-term academic performance in an international university in Japan, the correlations, especially for first-year GPA and final GPA, are lower than might be expected.

There are at least two possible reasons for these lower correlations. The first reason concerns the number of cohorts studied. Although data were available for the 2004 to 2008 cohorts, students in the 2004 cohort did not complete the same entrance examinations, nor did they submit Center Examination scores. Therefore, in order to make the sample comparable for most of the research questions, data from the 2005 to 2008 cohorts were used. However, the range for first-year GPA was slightly more restricted for the 2005 to 2008 cohorts, so the range was reduced by .03. This slight range restriction reduced the correlation between HSGPA and first-year GPA by .01, and, because first-year GPA was a central variable, this slight range restriction affected other parts of the study as well.

The second reason that correlations are somewhat lower than expected for first-year GPA is that a number of low-achieving students had to be omitted from the analysis, and this range restriction resulted in lower correlation coefficients. Of 402 students, three had repeatedly failed in the EAP program; thus, they had not completed two semesters of regular university work by the time the study data were collected. These three students were therefore not included in the first-year analysis. However, if the GPA of these students is added to the first-year analysis (using the GPA they established while taking electives in the EAP program), the correlation of HSGPA and first-year GPA is .30 (compared to .28, without these three students).

In conclusion, the correlations of HSGPA with EAP GPA, first-year GPA, and final GPA are all statistically significant, and confirm the hypothesis regarding

HSGPA for the first research question. As noted, it is likely that the correlations are slightly larger than those actually found, for a number of reasons, including range restriction. Adjusting these correlations for attenuation also increases them, as was discussed in the Results section. However, it is unclear to what degree such adjustment brings the correlation closer to the actual correlation between the variables.

The answer to this part of the first research question adds to the literature in a number of ways. These results are the first to illuminate the relationship of HSGPA to university success in Japan. As Zwick (2002), Astin (1993), and the vast majority of researchers have pointed out, HSGPA is the most powerful predictor of university GPA. Until now, HSGPA has not been regarded as central to university success by many admissions officials in Japan, who have held the persistent belief that *hensachi* rankings of high schools are more trustworthy than the HSGPA. However, the results of this study show that HSGPA is not moderated by *hensachi* rankings, but is instead the most powerful predictor of university GPA, just as it is in other countries. Therefore, Japanese admissions officials should reexamine their past policies and begin to investigate the degree to which HSGPA predicts achievement for their own students. Although the results of this study are based on students in an international English-medium university in Japan, they probably have relevance to other types of universities in Japan, provided that course grades at those institutions reflect academic achievement.

The results concerning the importance of HSGPA as a predictor of university achievement for this group of students also contributes to the literature in another sense. Because this study is based on data gathered in an English-medium university, the results might also contribute to predicting the success of Japanese students studying at universities in the United States and other countries where English is spoken. As Burton and Ramist (2001) noted, second language learners at American universities are a "very important and growing" group about whom little is known; only a few studies of first-year success, and none of long-term success, have involved them (p. 26). Therefore, the results of this study add to what is known about the importance of HSGPA for members of this little-researched group.

Prediction from ITP TOEFL Scores

The results of this study also show that ITP TOEFL scores are significant predictors of EAP GPA. Therefore, the relevant hypothesis of the first research question, that TOEFL scores would predict EAP GPA, first-year GPA, and final GPA, was only partially confirmed.

Comparing the results of the current study to previous studies in which TOEFL scores were used is difficult, given the many differences among those studies in terms of the number of participants, gender, age, level of study, academic major, measures used for independent and dependent variables, and the analyses used. Graham (1987) also noted the lack of comparability among these studies (pp. 508-510).

However, some general comparisons are possible because correlations of total TOEFL scores and GPAs were reported in many of those studies. (Those that reported correlations for TOEFL section scores are addressed in the discussion section for Research Question 3.) Though some researchers did not find significant results, those that found statistically significant correlations between total TOEFL score and GPA for undergraduate students were Johnson (1988) and Light and Wan (1991), who reported correlations of .36 ($p < .01$) and .33 ($p < .05$), respectively. Researchers who studied graduate students and who found significant correlations were Hwang and Dizney (1970); Ayers and Peters (1977); Light, Xu, and Mossop (1987); and Vinke and Jochems (1993). These researchers reported correlations of .66 ($p < .05$), .40 ($p < .01$), .14 ($p < .05$), and .51 ($p < .002$), respectively. (Hwang and Dizney reported the correlation of .66 for ESL course grades and TOEFL scores, while all others reported correlations for university GPA and TOEFL scores.)

As is clear, the correlations reported in these studies between total TOEFL score and GPA differ greatly. In addition, the sample sizes in the majority of these studies are small, as previously discussed in the review of the literature. In order to properly compare the results of the current study, it is necessary to choose the most exemplary studies from the review. Of all the studies described, the Light, Xu, and Mossop (1987) study seems most trustworthy and comparable to the current study; the sample of students studied was the largest ($N = 367$), and the straightforward method of correlating pre-admission TOEFL scores with the GPA of the first semester of graduate school was used. An important difference is that the Light et al. study involved graduate students. Therefore, one must expect that their correlation would be lower than that obtained for undergraduates because the higher entering English proficiency of graduate students typically restricts the range of TOEFL scores and thereby lowers the correlation between TOEFL and university GPA. Therefore, it is reasonable to expect the correlation with undergraduate GPA in the current study to be higher than the .14 correlation found by Light et al. A larger correlation for undergraduates was found, with the correlation between ITP TOEFL total scores and EAP cumulative GPA of .22 ($p < .001$).

In the regular university program, correlations with the ITP TOEFL total scores are not significant at the $p < .05$ level. First-year university GPA correlates with ITP TOEFL scores at .11 ($p = .09$) and with final GPA at .13 ($p = .10$). These correlations are close to significance, indicating that language proficiency continues to play a role of some importance for students. After all, as Rosenfeld, Oltman, and Sheppard (2004) explained, even though the goal of the TOEFL is not to predict academic success, “facility with English contributes to students’ success in a given course” (p. 3). It should be noted that the graduating students who earned the final GPAs had surpassed 550 on the TOEFL and spent one year of study abroad. In other words, their level of English proficiency was similar to that of the graduate students in Light, Xu, and Mossop’s (1987) study. Therefore, it is unsurprising that the correlations are nearly identical (.13, as compared to .14).

Although English proficiency played a role throughout the undergraduate years of the students in this study, the correlations between the TOEFL total scores and GPA drop over time and gradually lose statistical significance. This change in relationship might appear to be due to differences in the content and grading schemes of courses in the EAP program and the regular university program. Compared to the scope of the EAP program, which focuses on English language skill development, the regular university program focuses on content areas. In their first year of regular university study, students take courses such as anthropology, mathematics, education, chemistry, biology, microeconomics, and freshman English. Once they complete such basic education courses, students then take courses in one of two majors, Global Business or Global Studies. Each of the regular university program courses tends to have somewhat different class activities, assignments, assessment types, and percentages of course credit assigned for work completed. Accordingly, there are differences between not only EAP courses and regular university program courses, but also among regular university program courses, just as there are at all universities. However, there is an important similarity between all courses at the university under study. All courses are English-medium and require that students read, write, listen, and speak in English. In other words, it is reasonable to expect that English language ability, as reflected in the ITP TOEFL score, would predict students' grades in such courses. Such an expectation is implicit in the studies of Hwang and Dizney (1970), Ayers and

Peters (1977), Light, Xu, and Mossop (1987), Vinke and Jochems (1993), and all others discussed in the relevant section of the review of literature.

Why then do the correlations between the TOEFL total scores and GPA drop over time and gradually lose statistical significance? One reason previously discussed is that the range of TOEFL scores becomes increasingly restricted as students advance in their university studies. Another change that might occur during the years of regular university study is that students are reaching the kind of "cut-off point" that a number of researchers have suggested exists (e.g., Alderman, 1981; Johnson, 1988; Light, Xu, & Mossop, 1987; Vinke & Jochems, 1993). At this cut-off point, possibly between a total TOEFL score of 450 and 500, students become able to demonstrate their academic abilities in the second language. The results in this part of the study generally support the notion of a cut-off point. After students complete the EAP program, they are required to have earned a TOEFL score of at least 550 in order to qualify for the regular university program. In accordance with the idea that these students have reached or exceeded the cut-off point, the regression analysis results show that HSGPA, which is a reflection of students' abilities in their L1, becomes increasingly important in predicting GPA, while TOEFL scores gradually take on a less important role.

Although not related to the research questions or hypotheses in this study, a notable finding is that male students score higher than female students on the ITP TOEFL, and the mean differences are nearly significant on the second administration of the test, with male students scoring approximately 6 points higher

(p = .06). These results, though not statistically significant, are noteworthy because they are generally in agreement with the results of other studies that have indicated that male students tend to score better than female students on standardized tests (Zwick, 2002, p. 144).

These results regarding the predictive validity of total ITP TOEFL scores add to the literature in a number of ways. First, these results are the first to show the relationship of ITP TOEFL to success in an international university in Japan. Although these results might not be relevant to Japanese-medium universities, they are potentially useful for English-medium universities in Japan, as well as for universities in countries where English is spoken. Because the TOEFL is useful in predicting initial success in university and continues to have a relationship to later success, it behooves admissions officials in English-medium universities in Japan to require that students submit TOEFL scores during the application process. Using TOEFL scores, along with HSGPA, and other useful information in a regression equation, for example, might improve the admissions process and lead to selecting candidates with greater chances of academic success. In addition, asking students to submit TOEFL scores as part of the application process provides practical benefits to universities. At present, the university under study administers two separate TOEFL testing sessions when students begin in April as first-year students. These test administrations require hours of time from both faculty and students. If entering students were to submit TOEFL scores during the application process, these scores could be used for placement, saving the hours of faculty and student time needed for testing. In addition, the EAP department could use the time previously used for administering the ITP TOEFL to develop and eventually administer their own institutional placement test based on curricular goals. By using both ITP TOEFL scores and institutional placement test scores, administrators could improve the reliability of student placement within the program. In any case, having students submit TOEFL scores as part of the application process would be advantageous.

These results regarding the ITP TOEFL also contribute to the second language acquisition literature in general. As already noted, many studies of the TOEFL suffer from inadequacies. Although this study has its limitations, the sample size is adequate, and the results are reasonable, especially in light of the results of past, larger studies. In addition, the results of this study suggest possible connections with theories discussed previously concerning language proficiency and the TOEFL. For example, the finding that the ITP TOEFL scores are close to statistical significance in predicting first-year GPA and final GPA, indicating that English proficiency continues to play some role in predicting the academic success of students, suggests support for: Cummins (1980), who said that a component of language proficiency (CALP) is related to academic performance, and Ellis (1985), who said that CALP probably predicts academic performance, especially in traditional classrooms. The results also provide evidence that language proficiency moderates academic achievement, and that there might be a particular cut-off point at which students are able to demonstrate their academic abilities in an L2 to the same degree that they are able to do so in their L1.

Prediction from *Hensachi* Rankings

The results of this study show that the *hensachi* rankings are significant predictors of first-year GPA and final GPA. These results confirm the hypothesis that *hensachi* rankings would not predict EAP GPA, but would predict first-year GPA and final GPA.

Comparing these results to those of past studies is not possible because no known research on *hensachi* rankings as predictors of university GPA exists. As noted earlier, Newfields (2006) doubted that these rankings would have value in predicting university success. However, Linn's (1966) discussion of differences in American high schools and the routine adjustments for these differences by university admissions officials suggested that an indicator of high school level, if accurate, might have value in predicting later university success. In addition, he pointed out that using another standardized measure of academic achievement (such as the SAT) would help to counter differences in high schools. Again, if the *hensachi* rankings are reasonably accurate and standardized, these rankings might help predict university success.

As implied by Linn, the *hensachi* rankings did predict success in the regular university program. As discussed earlier, the academic success of students is better predicted by academic measures rather than English proficiency beginning in the first year of regular university study. These rankings, then, are predictive of success during this period in the regular university program, but not in the EAP program. Specifically, correlations between *hensachi* rankings and EAP GPA, first-year GPA, and final GPA are -.001 (*ns*), .18 ($p < .001$), and .15 ($p < .01$), respectively. HSGPA is a more important predictor, but the *hensachi* rankings play a statistically significant role in predicting regular university GPA.

Although *hensachi* rankings are useful in predicting academic performance, admissions officials and researchers have to keep in mind two important points. First, as noted in the earlier discussion of the ecological fallacy, these are rankings of students' high schools, not of students. Although it is true that the rankings generally reflect the individual accomplishments of students who passed the necessary entrance examination to enter a high-ranking school, it is also true that other students entered the same school in other ways. For example, many Japanese high schools have attached junior high schools and elementary schools. Students in such a system can gain admission to the next level of schooling more easily than those who are not in the system. Therefore, the *hensachi* ranking of a high school does not represent a uniform level of achievement for all students who attended that school. Another point that researchers and admissions officials should attend to concerns the need to examine whether *hensachi* rankings moderate HSGPA or not. In the current study, it was determined through multilevel modeling that they did not; however, in the future, this relationship could change. Therefore, this relationship should be reexamined whenever *hensachi* rankings are used.

In conclusion, the hypothesis concerning *hensachi* rankings of the first research question is confirmed. These rankings add significantly to the prediction of first-year GPA and final GPA. This finding, though not applicable to schools outside of Japan, is the first to indicate the effectiveness of these rankings in predicting college GPA in Japan. As previously pointed out, admissions officials must carefully consider how to make best use of these rankings, as well as to explore the possibility of obtaining the individual *hensachi* rankings, which are not currently used in the university application process in Japan.

Combined Prediction from HSGPA, ITP TOEFL Scores, and *Hensachi* Rankings

The variables examined in answering Research Question 1 are useful as individual variables in predicting university GPA. They are also useful in combination. For the EAP GPA, the multiple correlation of HSGPA, TOEFL, and the *hensachi* rankings is .34, with an R^2 of .12. In the case of first-year GPA, the multiple correlation for the three variables is .39, and the R^2 is .15. Finally, the multiple correlation for final GPA is .38, and the R^2 is .15. In other words, the combination of these three variables accounts for 12% of the variance in the EAP GPA, and 15% of the variance in both the first-year GPA and final GPA.

These results might appear disappointing because 85% to 88% of the variance in GPA is not accounted for, but these correlations are similar to those found for two useful predictors in the United States, HSGPA and the SAT total score. The results from Table 1 show that the unadjusted multiple correlations of these two variables with first-year university GPA that were reported in two studies are .48 and .44. Considering that the ITP TOEFL is not an academic aptitude test, and that the *hensachi* ranking is not an individual difference variable, a multiple correlation of HSGPA, ITP TOEFL scores, and *hensachi* rankings with first-year GPA of .39 is far from disappointing.

However, it is true that there are many factors that contribute to success in college which are not accounted for here. Examples include academic preparation, social support, adequate financial resources, motivation, personality characteristics, effective time management, number of extracurricular activities, and level of part-time employment. Although many of these are not accounted for here, this study makes a start at explaining academic success in Japan, with results that are not unlike those from other more well known studies.

Research Question 2: Predicting EAP GPA, First-Year GPA, and Final GPA from High School Grade Factor Scores

The second research question asked to what extent high school grade factor scores predict EAP GPA, as well as first-year and final GPA in the regular university program. Based on the work of Gardner (1990) and others who see intelligence as undergirded by more specific types of abilities, as well as Kline (1991), and other intelligence theorists who have noted that researchers have consistently found three factors underlying intelligence, it is hypothesized that high school grades can be factored into three factors (representing more specific types of abilities), and that these factors vary in the extent to which they predict GPAs. The results confirmed the hypotheses associated with Research Question 2.

The results confirm the theory that intelligence is composed of separate kinds of abilities. As hypothesized, the factor analysis of the high school grades reveals three such factors, Language Ability, Numerical Ability, and Spatial Ability. The results also confirm the hypothesis that these factor scores predict university GPA to varying degrees. Unfortunately, it is not possible to compare these results with other studies because no other similar studies were found.

Using High School Grade Factor Scores to Predict EAP GPA

The results indicate that the Language Ability and Spatial Ability factor scores are statistically significant predictors of EAP GPA (with unadjusted correlations of .21 and .20, respectively), with each contributing to the prediction to about the same degree. Results from the initial hierarchical multiple regression analysis reveal a multiple correlation of .26 of the three factor scores with EAP GPA, and an R^2 value of .07. When the three factor scores and ITP TOEFL total scores are combined, the multiple correlation is .34 and the R^2 value is .12. These values are similar to the multiple correlations of HSGPA and the ITP TOEFL total scores with EAP GPA (although the R^2 using factor scores is .01 higher).

The Language Ability and Spatial Ability factor scores significantly contribute to the prediction of EAP GPA, but the Numerical Ability factor score does not, a finding that is reasonable given the fact that no course in the EAP program is quantitative in nature. It is also reasonable that the Language Ability factor score is predictive of the EAP GPA because language is at the forefront in EAP classes, as students read, write, speak, and listen to English. However, an unexpected finding here is that the Spatial Ability factor scores significantly predict EAP GPA. Gardner (1983) explained that spatial intelligence includes “loosely related capacities” such as being able to use and transform mental images and recognize objects and scenes (p. 176). This type of intelligence is also related to memory, in that it allows a person to relate patterns to other patterns previously met, and to make sense of and reuse this knowledge (p. 195). In the EAP program, all classes are taught in English, and such an experience is new for nearly all students. During this period, there are many types of patterns that students are working to remember, from word families and presentation formats, to paragraph and essay patterns. Perhaps students must call on their spatial abilities more at this point in their college careers, as they become accustomed to working with images and words in a second language and managing classroom tasks in English. In addition, this ability might be more important for students until they reach the English proficiency cut-off point discussed earlier, in which they become able to use their academic abilities to approximately the same degree in their first and second languages.

Using High School Grade Factor Scores to Predict First-Year GPA

The results indicate that the Language Ability and Numerical Ability factor scores are statistically significant predictors of first-year GPA (with unadjusted correlations of .18 and .26, respectively), with the Numerical Ability factor scores contributing to a greater degree. The results from the initial hierarchical multiple regression analysis reveal a multiple correlation of .29 and an R^2 value of .08 for the three factor scores with first-year GPA. The combination of factor scores and the *hensachi* rankings produces a multiple correlation of .38 and an R^2 value of .15. These values are higher (by about .01) than those that result from using HSGPA instead of factor scores.

The Language Ability and Numerical Ability factor scores contribute significantly to the prediction of the first-year GPA. This finding is reasonable because at this point in their college careers, students begin taking basic education courses that are required for the two majors offered at the university. These courses include mathematics and science classes (e.g., algebra, statistics, and physics) and students typically complete these quantitative courses during this first year of regular university study. Perhaps for that reason, the Numerical Ability factor score becomes most important at this time. The Spatial Ability factor scores do not contribute significantly to the prediction of first-year GPA. This finding supports the theory that students have reached a cut-off point of language proficiency by the end of their first year of regular university study, and no longer depend as much on the Spatial Ability factor that was useful to them while in the EAP program.

Using High School Grade Factor Scores to Predict Final GPA

The results also indicate that the Language Ability and Numerical Ability factor scores are statistically significant predictors (with unadjusted correlations of .25 and .23, respectively) of the final GPA. Results from the initial hierarchical multiple regression analysis reveal a multiple correlation of .32 and an R^2 value of .10 for the three factor scores with final GPA. When the three factor scores and the *hensachi* rankings are combined, the multiple correlation is .40 and the R^2 value is .16. All values are higher by .03 or .04 than those obtained when HSGPA is used instead of factor scores. In this case, using separate factor scores resulted in a better prediction of GPA at all three levels, and particularly so for final GPA.

The Language Ability and Numerical Ability factor scores significantly contribute to the prediction of the final GPA, and each contributes approximately to the same degree. Like the first-year GPA results, these findings are reasonable because the overall coursework completed by graduates requires an integration of quantitative skills and language skills. For example, the Global Business majors take courses in accounting, economics, and banking; the Global Studies majors take courses related to economics, information technology, and business. Therefore, students are required to use and demonstrate both numerical and language abilities. As with first-year GPA, the Spatial Ability factor scores do not significantly predict final GPA. In addition, the correlation of the Spatial Ability factor scores with first-year GPA and final GPA are .13 and -.02, respectively. The fact that this correlation drops to such a degree is further support for the proposal that students have reached a level of language proficiency that no longer requires the kinds of abilities represented by the Spatial Ability factor scores. On the other hand, the Language Ability factor scores are important throughout their college careers. The Language Ability factor might be a combination of general intelligence and language ability that is necessary for success at all levels of study.

In conclusion, the findings confirm the hypothesis associated with Research Question 2, that high school grades can be factored into three factors and that these factors vary to the extent to which they predict EAP GPA, first-year GPA, and final GPA. For the most part, the findings for this research question are reasonable.

Gender Differences in High School Grade Factor Scores

Although the second research question did not specifically concern gender differences, it is worthwhile to point out a number of gender differences found. Female students have significantly higher Language Ability and Spatial Ability factor scores than male students, but their Numerical Ability factor scores are not significantly different. These results are reasonable for at least two reasons. First, the factor scores are derived from high school grades, and as Zwick (2002) pointed out, past research typically indicates that female students have higher GPAs in both high school and university (p. 143). Therefore, it is unsurprising that the female participants have significantly higher scores for two of the three factor scores. It is also reasonable that the female test takers have higher Language Ability factor scores because they tend to score higher on the ACT English and Reading tests and often score higher than male examinees on other reading and writing tests, though they are generally outperformed by male students on standardized tests (p. 144). This relatively greater ability with language seems connected to the higher Language Ability factor scores. However, the finding that female participants have higher Spatial Ability factor scores is not supported by previous empirical research and requires further study.

The findings contribute to the literature in a number of ways. First, to the best of my knowledge, this is the first study to create factors from high school grades. These findings might contribute to educational research in a number of practical ways. For example, educational researchers and curriculum planners might use this kind of information to consider and study more about how different kinds of intelligence are used in different types and levels of study. Understanding abilities that students need at different levels also could help curriculum planners and teachers evaluate educational goals and curricula in a more informed manner.

Research Question 3: Predicting EAP GPA and First-Year GPA from ITP TOEFL Section Scores

The third research question asked to what extent ITP TOEFL section scores predict EAP GPA and first-year GPA in the regular university program. It is hypothesized that the section scores vary in the extent to which they predict EAP GPA and first-year GPA. As is true for total ITP TOEFL scores, the research available did not offer definitive answers as to how sections scores relate to academic success. However, the general hypothesis associated with Research Question 3 was confirmed; the ITP TOEFL section scores predict academic success at university to different degrees.

For purposes of comparison, two studies previously discussed are briefly reviewed. Johnson (1988) and Al-Masawi and Al-Ansari (1999) presented correlations among TOEFL section scores and GPA. Johnson reported the following statistically significant correlations of TOEFL section scores with university GPA: Section II, Structure and Written Expression, .43, and Section III, Vocabulary and Reading, .36. The Section I Listening section was not significantly correlated with university GPA, according to Johnson. On the other hand, Al-Masawi and Al-Ansari reported correlations between university GPA and the section scores as follows: Section I, .34; Section II, .51; and Section III, .42. They also reported that the TOEFL section scores were highly correlated with English course grades. As noted above in the discussion of research on total TOEFL scores, the results from these two studies are not completely comparable. Johnson concluded that reading ability was important for academic achievement (p. 166), but Al-Masawi and Al-Ansari argued for the need to teach grammar to help students achieve academic success.

Predicting EAP GPA from ITP TOEFL Section Scores

Correlations among ITP TOEFL section scores and EAP GPA in this study were: Section I, .32 ($p < .001$); Section II, .22 (*ns*); and Section III .28 ($p < .01$).

The multiple regression results indicated that Section I scores are equal in importance to HSGPA. For EAP then, listening scores are of most importance, and reading is next. The Structure and Written Expression scores are not predictive of EAP GPA. This last finding differs from the findings of both Johnson (1988) and Al-Masawi and Al-Ansari (1999), and is surprising, given the fact that students are required to demonstrate grammatical ability in reading, writing, and TOEFL preparation courses, the grades of which compose their EAP GPA. However, the finding that reading predicts GPA is in agreement with the findings of the above studies, as well as the opinions of other researchers, such as Bayliss and Raymond (2004), who said that reading comprehension is the best predictor of GPA (p. 42).

The findings that TOEFL Listening Comprehension scores are important and that TOEFL Structure and Written Expression scores are not predictive of GPA, can be explained by considering the sample of students involved in this and other studies. Johnson conducted her study in the United States, Al-Masawi and Al-Ansari conducted theirs at the University of Bahrain, and the participants in this study were Japanese university students. In Japan, students study English grammar extensively in junior high and high school, but they generally do not use the language often. As a result, one might expect that the grammatical knowledge (at least that needed for taking tests) of students who enter university is relatively good, while their listening ability is more variable. If this supposition is true, the variance in listening ability would be wider and would result in higher correlations with

EAP GPA, than those between grammatical knowledge and EAP GPA. The supposition is supported by the results explained here (also see Tables 30 and 31).

Predicting First-Year GPA from ITP TOEFL Section Scores

The correlations among the ITP TOEFL section scores and first-year GPA, as reported in Table 33, are as follows: Section I, -.01 (*ns*); Section II, .00 (*ns*); and Section III .14 (*ns*). The multiple regression results indicated that none of the section scores are statistically significant predictors of first-year GPA. However, Section III comes close to significance (p = .10).

The findings for first-year GPA indicated that English listening and grammatical proficiency plays almost no role in predicting academic success, though reading comprehension remains a factor, albeit not a statistically significant one. These findings are in agreement with those regarding total TOEFL scores. As noted before, students at this point in their university careers have all attained 550 and above on the TOEFL, with the result that the range of English ability has become much more restricted than it was when the students were studying in the EAP program. Given this restricted range of English ability, it is unsurprising to find that English proficiency, as demonstrated in the section tests, is no longer a significant predictor of GPA. In addition, the multiple regression results in Table 34 show that HSGPA is the only significant predictor of first-year GPA; this result also suggests that students have passed the cut-off point described earlier, which means that they are able to use their academic abilities in the second language to approximately the same degree that they use them in their first language. Therefore, HSGPA, and not English proficiency, is the significant predictor of achievement for these students.

In conclusion, the hypothesis regarding ITP TOEFL section scores was confirmed in that the TOEFL section scores vary in the extent to which each predicts EAP GPA and first-year GPA. For EAP GPA, the Listening and Reading Comprehension section scores were statistically significant predictors of success; however, for first-year GPA, none of the section scores significantly predicted achievement, though Section III scores were closest to reaching statistical significance ($p = .10$).

Of the three ITP TOEFL section tests, Section III (Reading Comprehension) is most related to success in the regular university program. These scores correlated most highly with measures of academic ability or level, such as HSGPA, *hensachi* rankings, and first-year GPA. Although the final result discussed above showed that Section III scores were not statistically significant predictors of first-year GPA, it should be noted that an earlier analysis that included three students who were later excluded (mentioned in the discussion of the first research question) indicated that Section III was a statistically significant predictor of first-year GPA. In addition, because the regression results were similar for first-year GPA and final GPA, it is likely that reading comprehension continues to be related in some important way to students' success throughout their college careers. These findings might simply highlight the fact that students are required to read frequently in the university, and that those who read well early in their college careers generally continue to perform better than those who do not.

Although one might expect that listening and grammatical ability, as reflected in scores on the Section I and II tests, would also predict success in the first-year regular university program because students must be able to listen effectively in all classes and demonstrate grammatical understanding in the extensive writing required in their first-year Composition I and Composition II courses, scores from neither the Section I nor Section II tests were significant predictors of first-year GPA. The finding that Section III scores appear most predictive of later university success might indicate that the ITP TOEFL reading comprehension section test is unlike the Section I and II tests, in that it also assesses academic ability, in addition to assessing English reading proficiency.

Gender Differences in ITP TOEFL Section Scores

Although the third research question was not specifically concerned with gender differences, one finding concerning the performance of female and male participants was worth noting. As Zwick (2002) noted, female test takers often outperform their male counterparts on tests of reading and writing; however, the results for the third research question indicated that the male participants significantly outperformed the female participants on Section III, Reading Comprehension ($p < .001$). The mean for the male participants was 50.17 (SD = 4.36), and the mean for the female participants was 48.04 (SD = 4.39) $p < .001$).

Even though the differences appear small, they still might be important because reading comprehension appears to be more related to success in the regular university program than the other two ITP TOEFL section tests.

The results here add to the literature in a number of ways. As far as I know, these results are the first to show the relationship of ITP TOEFL section scores to success in an international university in Japan. Such results are potentially useful for English-medium universities in Japan, as well as for universities in countries where English is spoken. As noted earlier, admissions officials at English-medium universities in Japan should consider requiring applicants to submit TOEFL scores as part of the application process. Total TOEFL scores provide useful information, but section scores provide more precise information that can improve the prediction of GPA. For example, in the regression analysis using HSGPA and total ITP TOEFL scores to predict EAP GPA, the resulting multiple correlation was .34, and the R^2 value was .11. However, using only the Section I and III section scores, rather than the total ITP TOEFL scores resulted in a multiple correlation of .46 and an R^2 value of .21. In other words, if admissions officials were to use the TOEFL section scores that they knew were most predictive for students at their university in a regression formula, they could improve their predictions of students' grades. In this case, HSGPA and the two section scores accounted for 21% of the variance in EAP GPA (compared to only 11% when the total ITP TOEFL scores were used). It seems clear that information from ITP TOEFL section scores can be valuable for admissions purposes.

These results regarding ITP TOEFL section scores also contribute to the second language acquisition literature in general. As already noted, more information is needed about the relationship of TOEFL to academic success, both in general, and for particular subpopulations of students. Therefore, these results are useful in that they shed additional light on the relative value of each of the test sections, specifically for Japanese university students. In addition, these results provide additional information related to the notion of a cut-off point in language proficiency, at which students are able to demonstrate their abilities in an L2 to approximately the same degree that they are able to do so in their L1. The results indicate that the notion of a cut-off point holds true for Section I, and to some degree, for Section III, because these scores were significant predictors of GPA in the EAP program, but not in the later university program. In summary, these results add to what is known about the relationship of the ITP TOEFL section tests to university academic success, the relationship between the TOEFL section tests and the notion of a linguistic proficiency cut-off point, and the relative importance of each TOEFL section test for Japanese students.

Research Question 4: Predicting EAP GPA, First-Year GPA, and Final GPA from Non-Cognitive Variables

The fourth research question asked to what extent the non-cognitive variables, gender and each parent’s education, predict HSGPA, EAP GPA, first-year GPA, and final GPA. Based on past research, it was hypothesized that both gender and each parent's education would predict HSGPA, EAP GPA, first-year GPA, and final GPA. The results partially confirmed the hypotheses for gender because being female was a significant predictor of HSGPA and EAP GPA, but not of first-year GPA or final GPA. The results also partially confirmed the hypotheses for parents' education because the father's education was a statistically significant predictor of GPA at times. However, the mother's education was not a significant predictor of academic success.

Predicting HSGPA, EAP GPA, First-Year GPA, and Final GPA from Gender

The results of this study partially confirmed the hypotheses associated with the fourth research question, that gender predicts HSGPA, EAP GPA, first-year GPA, and final GPA. A comparison of the mean GPAs of male and female participants indicate that statistically significant gender differences exist for HSGPA and EAP GPA. The results from the hierarchical multiple regression analysis shown in Table 36 indicate that gender is a significant predictor of HSGPA, with gender alone accounting for 8% of the variance in HSGPA. The finding that female students outperform male students in high school is "typical," according to Zwick (2002, p. 143). The hierarchical multiple regression analysis results shown in Table 38 show that gender is also a significant predictor of EAP GPA. Gender accounted for an additional 5% of the variance in EAP GPA, over and above HSGPA and ITP TOEFL scores. In high school and the EAP program, female students significantly outperform male students. Researchers such as Zwick (2002) have cited a number of possible reasons why female students outperform their male counterparts, including more regular class attendance and greater effort in study (p. 24). Another reason why female students perform better in the EAP program is that they typically perform better in verbal tasks, as pointed out by Neubauer and Fink (2005).

The female participants maintained a slight edge in grade point average over the male participants throughout most of their university careers; however, beginning with the first-year of regular university study, the differences in female and male achievement even out, and are no longer significantly different. This result is unlike the findings of Astin and Oseguera (2005), Perez (2004), and others who have claimed that women tend to do better in university than men.

The finding that the female participants significantly outperform the male participants only up to the EAP program was unexpected given that it is reasonable to assume that female students, whose mean GPAs exceeded those of male students in every subject in high school, would continue to significantly outperform them in university. However, such an assumption is based on the premise that the male and female participants do not change in essential aspects during these high school and university years. In actuality, both female and male students are maturing and changing in a variety of ways during these years; however, based on the fact that young men generally mature more slowly than young women, the argument can be made that the male students are making the greatest changes in maturity during these years, and that these changes contribute to their increased academic success.

Comparing mean GPAs from high school through university graduation reveals improvements in the GPAs of the male participants, indicating increased attention to their school work, possibly as a result of increasing maturity. The female participants, on the other hand, seem to maintain approximately the same level of GPA that they earned in high school. For example, the mean HSGPA for the male participants is 4.06 (SD = .53) and that of the female participants is 4.29 (SD = .43) on a 5-point scale. In the university EAP program, the mean GPA for the male participants is 2.90 (SD = .54) and that of the female participants is 3.18 (SD = .48) on a 4-point scale. In the first year of regular university study, the mean GPA for the male participants is 3.14 (SD = .53) and for the female participants, 3.18 (SD = .46). By the time the participants graduate, the mean GPAs and standard deviations are roughly equal; the final GPA of the male participants is 3.33 (SD = .37), and that of the female participants is 3.32 (SD = .33). Perhaps the growing maturity of the male participants is a factor that can account, at least in part, for the patterns in achievement found here.

In conclusion, the results concerning gender partially confirmed the hypotheses of the fourth research question in that female gender significantly predicted HSGPA and EAP GPA. These results add to the literature, as this study is the first to examine the question of the predictive power of gender in Japan, and they might provide a basis for further study, particularly of the changes in male patterns of achievement that appear to take place from high school to university graduation.

Predicting HSGPA, EAP GPA, First-Year GPA, and Final GPA from Parents' Education

The results of this study only partially confirm the hypotheses of the fourth research question that the mother and father's education would predict HSGPA, EAP GPA, first-year GPA, and final GPA. The results indicate that the father's education is a significant predictor of HSGPA (but unexpectedly, in a negative direction). There is also an indication that the father's education is positively related to university achievement, although neither the father's nor the mother's education is a statistically significant predictor of EAP GPA, first-year GPA, or final GPA.

In terms of the relationship of parents' education with HSGPA, the results of this study conflict with previous studies. For example, Zwick and Green (2007) found that HSGPA correlated with fathers' education at .18, and with mothers' education, at .14 (p. 15). In this study, the correlation between HSGPA and fathers' education is -.15, and mothers' education, -.07.

In addition, the results from the hierarchical multiple regression analysis shown in Table 36 indicate that the father's education is a significant predictor of HSGPA, accounting for an additional 2% of the variance in HSGPA. However, contrary to expectations, the father's education is a negative predictor of HSGPA; as the father's education increases, the child's HSGPA decreases. Reasons for this negative relationship are not completely clear. However, these findings might be an indication that adolescents are rebelling in their high school years against the values, including the educational values, of their fathers. Clearly, further research is indicated.

In terms of the relationship of fathers' education to university GPAs, the results of this study are more similar to previous research results. For example, Zwick, Brown, and Sklar (2004) also examined the relationship of parents' education (represented by the number of years of education of the more highly educated parent) with first-year university achievement. They reported regression coefficients for parents' education for separate cohorts at seven campuses of the University of California, from 1996 to 1999; the regression coefficients they reported ranged from -.03 to .11. Out of 29 coefficients reported, nine were statistically significant. Results of the current study indicate that regression coefficients for fathers' education (the more highly educated parent in this sample) were .14 for first-year GPA, and .19 for final GPA. Although these coefficients did not quite reach statistical significance, they are similar in size to those reported by Zwick et al. In addition, the regression results reported by Zwick et al. indicated that adding parents' income and education to the regression model that included HSGPA and SAT test scores increased the proportion of variance explained (R^2) by .004 (p. 12). In the current study, income was not a variable studied. Rather, gender and parents' education were included in the second model. Table 40 shows that adding gender and parents' education in the second model increased the proportion of variance explained by .02. I repeated this regression analysis without the gender variable and found the same results for R^2. In short, the results of this study that concern the relationship of the fathers' education to first-year GPA and final GPA are similar to those found by Zwick et al.

The hierarchical multiple regression analysis results for EAP GPA, first-year GPA, and final GPA are shown in Tables 38, 40, and 42, respectively, and in every case, the father's education exerts a stronger influence than the mother's. As shown in Table 67 and 68, the father's education is a significant predictor of the child's finishing the EAP program on time. In addition, for first-year GPA and final GPA, the father's education is close to statistical significance, at $p = .09$ and .06, respectively. These results suggest that the father's education might be a significant predictor of GPA in another university or with a larger sample.

In a number of studies, the mother's education, rather than the father's, has been linked to children's educational success and attainment (e.g., Nora, Barlow, & Crisp, 2005; Travis & Kohli, 1995). The results of this study do not confirm these kinds of links to the mother's education. In addition, the education of the fathers was not shown to be a significant predictor of university GPAs. One possible reason for the lack of statistical significance might be that Japanese society as a whole tends to value education, thereby lessening the influence of parents' education. Nevertheless, there are indications in the results that the education of fathers is linked to the university success of their children: the fathers' education is a significant predictor of FOT; the results concerning fathers' education for first-year GPA and final GPA are close to statistical significance; and the overall regression results are similar to those of Zwick, Brown, and Sklar (2004). These three results suggest that the father's education might be a predictor of university success in some contexts in Japan.

The findings of this study also show, as Zwick (2007) pointed out, that female students tend to come from families that have lower levels of income and education (p. 24). Although it was not possible to determine the income levels of the participants of this study, it is possible to examine differences in parents' education for the male and female participants. Using the answers reported by students who matriculated from 2004 to 2008 (including 270 female and 109 male students), the mean educational level of the fathers was 14.2 for the female students and 14.7 for the male students, while that of mothers was 13.5 for the female students and 13.8 for the male students. The differences are not large, but the trend holds true in every case with this sample of participants. Although the implications of this trend are unclear, it is interesting that this trend exists in other countries and in this sample of Japanese participants.

In conclusion, the results concerning parents' education partially confirm the hypotheses of the fourth research question in that the father's education significantly predicts HSGPA (although negatively), and the father's education appears linked to his children's university success. Though the results are unexpected at times, the findings here add to the literature in that, as far as is known, this study is the first to examine the impact of parents' education on HSGPA and university GPA in Japan.

Research Question 5: Predicting EAP GPA, First-Year GPA, and Final GPA from Center Examination Scores

The fifth research question asked to what extent Center Examination scores predict EAP GPA, as well as first-year GPA, and final GPA. It was hypothesized that only the Center Examination English test scores would predict EAP GPA, first-year GPA, and final GPA. The hypotheses were partially confirmed in that the Center Examination English test scores from 2006 and after were significant predictors of EAP GPA; however, the English test scores did not predict first-year GPA or final GPA. Instead, the Math section test was a significant predictor of first-year GPA. No Center Examination section test score was a significant predictor of final GPA, but the Social Studies section test was close to statistical significance.

Predicting EAP GPA from Center Examination Scores

The results of this study show that the Center Examination English test is a significant predictor of EAP GPA, but only for data from 2006 and after. (When using the 2005 to 2008 data, the results indicated that the addition of the five Center Examination tests in the second model was not significant; in addition, none of the section tests was a significant predictor of EAP GPA.) When using the data from 2006 to 2008, which includes the listening component that was added to the English test in 2006, the results generally confirm the hypothesis of the fifth research question concerning the prediction of EAP GPA.

Using the 2006 to 2008 data to predict EAP GPA, the results show that the multiple correlation of HSGPA, ITP TOEFL scores, and all five Center Examination test scores is .57, with an R^2 of .32. In other words, the combination of these variables accounts for 32% of the variance in EAP GPA. The set of five Center Examination test scores accounts for 7% of that amount ($p < .05$), as shown in Table 47. It should be noted that the only Center Examination test score that is a statistically significant predictor of EAP GPA in the second model is the English test (though Science is close to statistical significance at $p = .06$). It is not possible to compare these results with past studies because this is the first such study of the Center Examination, to the best of my knowledge.

As can be seen in Table 47, the ITP TOEFL was not a significant predictor in the second model, though the Center Examination English test is a significant predictor of EAP GPA. This result might lead some to conclude that the Center English test is a more important predictor of EAP GPA than the ITP TOEFL. To explore the relative impact of ITP TOEFL and the Center Examination English test, I performed the regression twice. In the first case, I entered HSGPA in the first model, ITP TOEFL in the second model, and Center English test in the third model. This analysis showed that the Center Examination English test did not quite reach statistical significance in predicting EAP GPA, over and above the ITP TOEFL ($p = .06$). In the second case, I entered the Center Examination English test in the second model, and the ITP TOEFL in the third model. The results for this analysis showed that ITP TOEFL was able to significantly predict EAP GPA over and above the Center Examination English test ($p < .01$). Although both tests are able to predict EAP GPA, the ITP TOEFL adds to the prediction of EAP GPA after the addition of the Center Examination English test, but the same is not true for the Center Examination English test. Therefore, the results shown in Table 47 should not lead one to overlook the importance of the ITP TOEFL.

One unexpected result found in the prediction of EAP GPA is that the Center Examination Science test is close to statistical significance ($p = .06$) in its positive relationship with EAP GPA. In other words, performance in Science appears to be related to English language performance in the EAP program. One possible connection between Science and English can be found in the above discussion of high school grade factors. As shown in Table 20, the factor analysis of high school grades indicates that Science loads strongly onto the Numerical Ability factor; however, Science also loads onto the first factor, Language Ability, at .47. Therefore, performance in Science appears to rely on both numerical and language abilities. Perhaps performance on the Center Examination Science test also relies on both numerical and language abilities. If so, the language component of the Science test scores might be the underlying reason that the Center Examination Science test is a nearly significant predictor of EAP GPA.

In conclusion, the hypothesis of the fifth research question, that the Center Examination English test predicts EAP GPA, was confirmed for the participants who took this test after the addition of the listening component in 2006. In addition, the Science test might have a significant relationship with EAP GPA. These results add to the literature in two ways. First, the results contribute specific information about the English test, and show that the test improved in predictive validity with the addition of the listening component in 2006. Second, the results suggest a relationship between performance in English and Science courses, which might provide a basis for further study.

Predicting First-Year GPA from Center Examination Scores

The results of this study show that the Center Examination English test is not a significant predictor of first-year GPA (for either the 2006 to 2008 data, or the 2005 to 2008 data). Instead, the Math test is the only significant predictor of first-year GPA. These results therefore do not confirm the hypothesis of the fifth research question, which concerned predicting first-year GPA with Center Examination English test scores.

For first-year GPA, the results show that the multiple correlation of HSGPA, *hensachi* ratings, and all five Center Examination test scores is .46, with an R^2 of .21. In other words, the combination of these variables accounts for 21% of the variance in first-year GPA. The five Center Examination test scores account for 5% of that amount (p = .33), as shown in Table 49. As with the prediction of EAP GPA, it is not possible to compare these results with past studies because this is the first such study of the Center Examination, to the best of my knowledge.

In conclusion, the hypothesis of the fifth research question, that the English section of the Center Examination significantly predicts first-year GPA, was not confirmed, as the Math test is the only significant predictor of first-year GPA. Unlike the results for EAP GPA, no other Center Examination test was a significant predictor of first-year GPA. These results are similar to the results for factor scores in that the best predictor of first-year GPA is the Numerical Ability factor scores, as is shown in Table 26. As explained in the discussion of factor score results, the Center Examination Math test probably predicts first-year GPA because students complete many mathematics and science foundation courses in the first year of university study.

Predicting Final GPA from Center Examination Scores

The results of this study show that the Center Examination English test is not a significant predictor of final GPA. The only test that is close to statistical significance is the Social Studies test (p = .06). The results therefore do not confirm the hypothesis of the fifth research question which concerned predicting final GPA with Center Examination English test scores. However, the results are similar to the Bridgeman and Lewis (1994) finding that the History test was one of two subject tests of the Advanced Placement (AP) tests that correlated most with overall university GPA for freshmen. Perhaps abilities needed for success in history and social studies, such as reading comprehension and ability to memorize facts, are important for overall university success. Nevertheless, the Center Examination Social Studies test did not predict first-year GPA. In fact, it is difficult to make any firm conclusions about the importance of the Social Studies test because the results for final GPA are based on an inadequate sample of 38 participants. Therefore, conclusions are tentative, at best.

For final GPA, the multiple correlation of HSGPA, *hensachi* ratings, and all five Center Examination test scores is .54, with an R^2 of .29. In other words, the combination of these variables accounts for 29% of the variance in final GPA. The five Center Examination test scores account for 12% of that amount (p = .41), as shown in Table 51. As with the prediction of EAP GPA, it is not possible to compare these results with past studies because this is the first such study of the Center Examination.

In conclusion, the hypothesis of the fifth research question, that the English section of the Center Examination significantly predicts final GPA, was not confirmed. No Center Examination test section significantly predicts final GPA (although Social Studies nearly reaches statistical significance at p = .06). Further study of final GPA is needed, especially because the sample size in that part of the study was not adequate.

Gender Differences in Center Examination Scores

Though it is not possible to compare results concerning how well the Center Examination tests predict university GPA with other studies, it is possible to compare the results with more general findings about high-stakes test performance. Although not specifically related to study hypotheses or research questions, findings that are of special interest concern the differential performance of female and male students on standardized tests. Zwick (2007) explained trends in the performance of female and male examinees on the ACT and SAT tests, which, like the Center Examination, are used as important criteria for university entrance. Zwick noted that male test takers tend to score better on the SAT math and verbal sections, and the ACT math and science sections and the composite score. Female test takers, on the other hand, tend to score better on the ACT English and reading tests (p. 24). The results here (from an additional independent-samples *t*-test) show somewhat different results for female and male examinees on the Center Examination. Scores on the Center Examination English, Japanese, and Science tests are not significantly different for female and male test takers. However, there are gender differences on the Social Studies and Mathematics tests. For Social Studies, male examinees (M = 83.84, SD = 8.64, N = 92) significantly outscore female examinees (M = 79.61, SD = 9.39, N = 172). The eta square index indicates that 5% of the variance in Social Studies scores is accounted for by gender. On the Mathematics test, male examinees (M = 79.43, SD = 12.04, N = 67) also significantly outscore female examinees (M = 74.92, SD = 13.21, N = 111). The eta square index indicates that 3% of the variance in Mathematics scores is accounted for by gender. In conclusion, although female and male test takers perform at about the same level on three of the Center Examination test sections, there are statistically significant differences in their performance on two test sections, with male examinees outperforming female examinees. Female test takers do not outperform male test takers on any Center Examination test section.

Results from the study by Bridgeman and Lewis (1994) provide an interesting backdrop to the above findings. In their study of essay and multiple-choice scores on Advanced Placement examinations in Biology, American History, European History, and English Language and Composition, they found that male test takers significantly outscored female test takers on a number of the multiple-choice tests studied, while female test takers tended to outperform male test takers on essay writing. In their conclusion, they suggested the need for further research into gender differences, including examining how differences might have developed on the "somewhat different constructs assessed by the two formats" (p. 49).

The results of this study also reveal gender differences in the multiple-choice Center Examination tests. For three of the tests, the performances of male and female test takers are not significantly different, but for two of the tests, male examinees outperform female examinees. These results are similar to those of Bridgeman and Lewis.

Though the results of this study of the Center Examination are similar to those of other studies that have shown that male examinees tend to outperform female examinees on multiple-choice tests, the results are still puzzling when one considers the high school performance of male and female students. As shown in Table 36, gender is a statistically significant predictor of HSGPA. In fact, female students perform significantly better than male students in high school in their Japanese, History, Civics, Science, Art, Foreign Language, and Home Economics classes. Though they do not outperform male students in Mathematics and Health Education, their mean averages in those subjects are still higher. Given this kind of high school performance, one would expect female examinees to outperform male examinees on most of the Center Examination tests because these tests are designed to measure achievement through high school, according to the agency that produces the Center Examination. Why then do female examinees score as they do?

One possible reason that female test takers are outperformed by male test takers on the Center Examination is that gender differences exist on the constructs assessed by different kinds of test formats, as suggested by Bridgeman and Lewis (1994). For example, female test takers might be more reluctant than male test takers to guess on multiple-choice test items, and this reluctance tends to lower their scores on such tests.

There are a number of other possible explanations for why female examinees perform at a lower level than expected on the Center Examination. One explanation is that the test content might be biased against female test takers. Zwick (2007) noted that American researchers realized as long ago as 1923 that test content can put female test takers at a disadvantage. For example, test questions that focus on subjects that are typically more familiar to male students, such as sports, the military, and mechanical topics, tend to bias the test in favor of male examinees (p. 25). Because American testing companies understand the need to review test items and eliminate content that might offend or be unfamiliar to female test takers (and other groups), test content has improved greatly since the early days of testing. Whether those in charge of constructing Center Examination test items, the majority of whom appear to be men, are aware of this potential problem is unclear. In my review of Center Examination materials, I have not seen any reference to the need for this kind of careful review of content; therefore, there is a possibility that test content might be biased against female examinees.

Another possible explanation for the apparent underperformance of female test takers on the Center Examination can be found in the “stereotype threat” theory proposed by Steele and Aronson (1995). According to this theory, members of groups that have been stereotyped in negative ways experience considerable anxiety when faced with a task (such as an achievement test) because of the risk of confirming the stereotype; the anxiety then leads to impaired performance on the task. In the case of gender, it seems likely that negative stereotypes about the abilities of women exist in Japan, a country that is dominated by men in practically every sphere of life. Therefore, it is also likely that young girls in Japan have been exposed to stereotypes regarding what women supposedly can and cannot do. When young women approach a test like the Center Examination, they might feel the kind of anxiety and stress that Steele and Aronson described, and, as a result, perform at an impaired level. Stereotype threat provides one explanation for the unexpected performance of female examinees on the Center Examination.

In conclusion, the results pertaining to the fifth research question show that the Center Examination English test predicts EAP GPA (for 2006 to 2008 data), and the Mathematics test predicts first-year GPA. Other good news for those who construct the Center Examination is that the predictive validity of the English test is improved, compared to the 2005 version of the test. In addition, the set of Center Examination scores adds to the total R^2 value for EAP GPA, first-year GPA, and final GPA, to varying degrees. However, most of the individual tests that comprise the Center Examination do not predict university performance in the EAP program, or in the regular university program. In addition, if the five Center Examination test scores are combined into one total score (as they sometimes are for university admissions purposes), the total score is also not a significant predictor of EAP GPA, first-year GPA, or final GPA. Because, as Zwick (2002) pointed out, the primary aim of an entrance examination is to predict university success, these results indicate that the Center Examination needs further study and revision.

To the best of my knowledge, these findings are the first to indicate the degree of predictive validity of each of the Center Examination tests. In addition, the findings are the first to relate ability factors, derived from high school grades, to the Center Examination tests. Finally, the results also are most likely the first to indicate that female test takers' performance on the Center Examination, and perhaps other important tests, might be negatively affected by factors such as gender differences in responding to different types of test formats, test content, or stereotype threat.

Research Question 6: Predicting EAP GPA, First-Year GPA, and Final GPA from University Entrance Examination Type and Test Scores

The sixth research question asked to what extent university entrance examination type and test scores predict EAP GPA, first-year GPA, and final GPA. Based on the unsystematic manner in which these tests are constructed and the widespread criticism of entrance examinations, it was hypothesized that only entrance examination English essay scores would predict EAP GPA, first-year GPA, and final GPA. The hypotheses were partially confirmed; as hypothesized, the entrance examination types did not predict GPA, and the entrance examination English essay scores were significant predictors of EAP GPA. However, contrary to expectations, the English essay did not predict first-year GPA or final GPA. In addition, the entrance examination type A and B English test was a statistically significant predictor of first-year GPA.

Predicting EAP GPA, First-Year GPA, and Final GPA from Type of Entrance Examination

The results of this study indicate that entrance examination type is not a significant predictor of EAP, first-year, or final GPA. The A, B, AO I, and AO II entrance examination types, when added as a set to HSGPA and either ITP TOEFL or *hensachi* rankings, did not significantly predict GPA at any level of university, as hypothesized.

Despite the fact that examination type does not predict GPA, a closer look at student averages for these examinations reveals that there are a number of significant differences between students who took these examinations. For example, the students who took the first examinations in February (types A and B) are students from high schools with significantly higher *hensachi* rankings. About half of these students also have significantly higher HSGPAs than the other students. However, the group of students who took examinations in February were average or significantly below average achievers in English, as reflected in their ITP TOEFL scores. Students who took the AO I examination in November (usually transfer students, working adults, or principal-recommended students) had significantly higher HSGPAs, but their high schools have lower *hensachi* rankings, and the students were significantly lower achievers in English, as reflected by their ITP TOEFL scores. Finally, the small number of students who took the AO II examination in August, which is principally offered for returnees, had significantly lower HSGPAs and *hensachi* rankings, but their English achievement was significantly higher than that of the other students.

Of these four groups of students, the only group that scored significantly higher than other students on any GPA (as determined by additional independent-samples *t*-tests) is the group who took the B examination. Their final GPA was significantly higher ($p < .05$) than students from all other groups, although the result should be viewed with caution because only 38 students took the B examination. This difference between students does not show itself in the regression analysis for final GPA because the HSGPA, ITP TOEFL scores, and *hensachi* rankings account for the majority of the variance in the GPAs. When I repeated the regression analysis for final GPA, and entered only the B type examination in the first model, and HSGPA and *hensachi* rankings in the second model, the B examination was not significant in the second model. Therefore, this extra analysis shows even more conclusively that entrance examination type is not an important variable in the prediction of university GPA.

Predicting EAP GPA, First-Year GPA, and Final GPA from Entrance Examination Tests

The results show that the A and B entrance examination English test is a significant predictor of first-year GPA, though not of EAP GPA or final GPA. This finding is unexpected but positive. Given the fact that entrance examinations in Japan are so frequently criticized, it is a pleasant surprise (especially for admissions officials) to find that one of the tests used in these examinations predicts university GPA, though to a limited degree.

The performance of female examinees on the A and B entrance examination English test is higher than that of the male examinees, though the difference is not statistically significant. These results are reasonable, given that the test is based on reading and writing, tasks on which female students often excel (Zwick, 2002, p. 144).

The type A and B entrance examination English test usually requires students to read an extended piece of writing in English, answer comprehension questions, translate short excerpts into Japanese, and write a longer reading response. Performance on this test predicts performance in the first year of regular university study. Therefore, the grades students earn in their first year of regular university study appear to be more dependent on these kinds of English skills, than are the EAP and final grades. Why is this the case? In the EAP program, reading is important, but grades are also based on listening, speaking, and writing classes. For the final GPA, perhaps grades are not based as much on reading comprehension alone. In the later years at university, students appear to take part in more discussions and presentations, so that grades might be less focused on reading comprehension.

Though such explanations are reasonable, explaining why the A and B entrance examination English test significantly predicts first-year GPA is difficult, if not impossible, because the test given on each administration lacks sufficient uniformity. Although most of the versions of the English test used for the A and B examinations included similar tasks, one or two of the versions I reviewed included only one task—a 200-word response to a reading. In other words, the response to that version of the English test was essentially an essay, resulting in an English test score that overlaps with the AO I and AO II essay scores. This kind of overlap is a problem because variables should represent scores of students on different types of tasks. In addition, this lack of uniformity in the A and B entrance examinations means that interpretation is nearly impossible. Because many of the tests include reading comprehension and translation questions, and an unspecified number of other tests include only a short essay, it is not possible to determine to what degree each of these tasks might contribute to the prediction of first-year GPA.

Lack of uniformity of entrance examinations might be a common problem because the professors who are asked to write test questions are rotated from year to year, and universities do not usually have a specially qualified testing committee. In addition, because the examinations are published after each administration, new versions have to be written every year. Given these conditions, it is understandable why test items are not typically evaluated or banked for future use. However, if university officials want to use their entrance examinations as vehicles for fairly and consistently choosing students who are most likely to succeed, they have to revise this approach. Part of that revision should include ensuring that particular tests have uniform content.

In conclusion, the A and B entrance examination English test is a significant predictor of first-year GPA, contrary to the hypothesis of the sixth research question. Though this finding is generally good news for admissions officials, interpretation is difficult, if not impossible, because of the lack of uniformity of the type A and B entrance examination English test.

Predicting EAP GPA, First-Year GPA, and Final GPA from Entrance Examination English Essay and Interview Scores

The results of this study show that the entrance examination English essay scores are significant predictors of EAP GPA, as hypothesized, though not of first-year GPA or final GPA. The interview test scores are not significant predictors of GPA, as hypothesized. These results are partial confirmation of the hypothesis of the sixth research question concerning English essays.

Though not related to research questions or hypotheses of the current study, a notable finding is that the male test takers scored significantly higher than the female test takers on the English essay ($p < .05$). This finding is unexpected because the general trend noted by Zwick (2002, p. 144) and Bridgeman and Lewis (1994) is that female students outperform male students on reading and writing tasks. However, whether the finding of this study is trustworthy or not partly depends on the fact that only 41 male students took the essay test while 150 female students completed the task; a larger sample size of male students would inspire more confidence in these findings.

The finding that the English essay score predicts only EAP GPA, but not the first-year GPA and final GPA, is similar to Christopher's (1993) finding that scores on an English essay did not predict university GPA for the 55 students she studied. Though Christopher's sample size was inadequate, the results are reasonable. Perhaps an essay writing test is a better indicator of the kinds of abilities that are needed in the EAP program than it is for those needed in later work. The entrance examination English essay requires students to read quickly under timed circumstances, and then write an organized response quickly. In EAP classes, students are often graded based on timed tasks that involve reading and writing. Later university classes appear to involve more lecture listening, and perhaps more presentation and discussion.

Further support for the notion that the entrance examination English essay requires abilities that are more closely related to those needed in the EAP program than to abilities needed in later university work can be found in the similarities between this English essay and the ITP TOEFL. Like the English essay task, the ITP TOEFL is also a significant predictor of EAP GPA, but not of first-year GPA or final GPA. An examination of Table 61 reveals that the correlations and partial correlations of the ITP TOEFL and the English essay with EAP GPA are very similar. Table 62 also shows that the ITP TOEFL and the English essay predict EAP GPA to about the same degree. However, this similarity does not mean that the two overlap in the abilities they measure. In fact, the two correlate at .16 ($p < .05$), but no multicollinearity is present as shown by the VIF values of 1.07 and tolerance values of .94. Therefore, these two measures significantly account for variance in the EAP GPA in a manner that is similar, even though they are not highly correlated. Perhaps the fact that both are timed tests of English accounts for much of the similarity.

The interview test score is not a significant predictor of GPA, as hypothesized. The interview test score has a smaller partial correlation than the

English essay score with EAP GPA and first-year GPA, but this pattern changes with final GPA in that the partial correlation of the interview test score with final GPA is .11 higher than the partial correlation of the English essay test with final GPA. This pattern suggests that the abilities measured by the interview test are more related to abilities used in the later part of the university program than are those measured by the English essay. As suggested before, students might be asked to complete more presentation and discussion tasks in the regular university program, both of which require the same kinds of abilities used in the interview test. However, these speculations are tentative at best because the results for the interview test are not statistically significant.

In conclusion, the results of this study partially confirm the hypotheses of the sixth research question in that the entrance examination English essay scores are significant predictors of EAP GPA, though not of first-year GPA or final GPA. These results add to the literature because they are the first such results in Japan, to the best of my knowledge.

Research Question 7: Predicting Finishing the EAP Program on Time (FOT) from HSGPA, High School Grade Factor Scores, ITP TOEFL Total Scores, *Hensachi* Rankings, Gender, and Parents' Education

The seventh research question asked to what extent HSGPA, high school grade factor scores, ITP TOEFL total scores, *hensachi* rankings, gender, and parents' education predict finishing the EAP program on time (FOT). It was hypothesized that HSGPA, a portion of the high school grade factor scores, ITP TOEFL scores, gender, and parents' education would predict FOT. The hypotheses were mostly confirmed in that HSGPA, the Numerical Ability factor scores, the father's education, and the ITP TOEFL scores, in that order, significantly predict FOT. However, female gender and the mother's education did not predict FOT.

It is not possible to compare these results with previous studies, as none exists. However, it is possible to compare these results with those found in answers to the other research questions. In the regression analyses, a number of variables were shown to predict EAP GPA, which is plausibly related to FOT. The following variables were determined to be significant predictors of EAP GPA: HSGPA and ITP TOEFL total scores (Research Question 1), the Language Ability and Spatial Ability factor scores (Research Question 2), ITP TOEFL Section I and III scores (Research Question 3), female gender (Research Question 4), the Center Examination English test, for 2006 and after (Research Question 5), and the entrance examination English essay score (Research Question 6). There were also indications that a number of other variables, though not statistically significant predictors, such as the *hensachi* rankings and the Center Examination Science test for 2006 and after (Research Questions 1 and 5), held some importance in the prediction of EAP GPA as well.

Most of the statistically significant predictors of EAP GPA are clearly related to academic ability (e.g., HSGPA) and English proficiency (e.g., ITP TOEFL, the Center Examination English test, and the entrance examination

English essay). That these are predictors of EAP is reasonable because the EAP program requires both academic and English abilities. Though the Spatial Ability factor scores appeared to be surprising predictors at first glance, their relationship to EAP GPA is also reasonable upon closer examination, given that this ability is related to image and pattern manipulation, as explained in the discussion of the second research question.

The predictors of finishing the EAP program on time are similar to the predictors of EAP GPA, a result that is also reasonable. It makes sense that students with relatively strong academic ability and high English proficiency would finish the necessary courses in the EAP program without failure at any level. Therefore, the fact that HSGPA and ITP TOEFL are significant predictors is not surprising.

On the other hand, a number of results are unexpected, particularly concerning the high school grade factor scores, gender, and parents' education. Before suggesting reasons for the unexpected results, a brief discussion of the dependent variables, EAP GPA and FOT, is needed. As noted above, both academic and English abilities are generally required to earn high GPAs and to finish the EAP program without failure at any level. However, finishing the EAP program on time requires additional abilities. To complete the program on time, students must complete one to three semesters of English courses that require considerable work outside of class, and they must maintain at least a "C" average. In this sample, 640 of 718 students (89.14%) finished on time, but a small percentage of students failed one or more semesters of EAP. Although these students have the necessary intelligence and memory to succeed, they lacked other qualities, such as the ability to break down assignments into manageable tasks, manage time, plan, and persevere to goal completion, at least in the semester(s) they failed.

The high school grade factor scores that are the most important predictors of FOT, the Numerical Ability factor scores, might be related to these additional abilities that students need in order to finish the EAP program on time. The Numerical Ability factor scores were not significant predictors of EAP GPA, but they are the most important predictors of completing the EAP program on time, and with no failure at any level. These factor scores are most important for FOT perhaps because these Numerical Ability factor scores represent abilities that encompass more than simple numerical aptitude, and include abilities that are of importance in planning and completing academic work over the long haul.

Gardner (1983) offers support for the notion that numerical ability is related to the reasoning ability needed for managing tasks over time. He explained that logical-mathematical intelligence, which is closest in nature to the Numerical Ability factor discussed here, was underpinned by the ability to understand an extended chain of reasoning, as well as the relationship between the links in the chain (pp. 137-139). Logical-mathematical ability is needed for understanding how homework and test scores accumulate over a long semester and eventually convert to percentages that finally comprise a total score, which results in a passing or failing grade.

Bull, Johnston, and Roy (1999) also offer support for the idea that numerical ability is related to the ability to manage tasks over time. They explained that studies have shown that an aspect of mental functioning called the central executive is related to mathematical ability, as well as to general school achievement. This central executive controls planning and goal direction, as well as working memory. In their own study, Bull et al. demonstrated that "central-executive functioning plays some role in mathematical skill" (p. 439). In other words, mathematical ability is related to the region of the brain that also controls the ability to plan and aim for goals. In light of the work of Bull et al. and Gardner (1983), the finding that Numerical Ability factor scores predict FOT is reasonable, even though it was at first unexpected.

The results for FOT regarding gender were also unexpected. Although female gender predicts EAP GPA, it does not predict finishing the EAP program on time. As noted, being able to finish coursework without failing at any level is not the same as being able to earn a high GPA, even though academic ability and English proficiency are important to both. In the case of gender, a slightly greater percentage of female students finish EAP on time (90.22%, compared to 86.64% for male students), but the difference is not statistically significant. The qualities that are needed to finish the EAP program on time are possessed about equally by female and male students.

The last unexpected result concerns parents' education. Although the results partially confirm the hypothesis because the father's education predicts FOT, the mother's education is not a significant predictor of FOT. In discussing these results, it is useful to look back at the results to the fourth research question. Those results showed that neither parent's education was a significant predictor of EAP GPA, first-year GPA, or final university GPA. Nevertheless, the results revealed a pattern in which the father's education was consistently close to statistical significance in predicting all three GPAs, while the mother's education was not. In these answers to the seventh research question, it is clear that the father's education plays an important role in his children's academic success, though again, the mother's education was not a significant predictor.

There might be a number of reasons why the father's education plays a more important role than the mother's education in predicting their children's likelihood of finishing the EAP program on time, but sifting through the possibilities is difficult. However, one point demonstrated by the data is that the fathers in this sample are more educated than the mothers. For example, 52.99% of the fathers finished 16, 18, or 20 years of education, while 23.31% of the mothers did so. In addition, fathers tended to marry women who had completed less education than they had. Out of 200 fathers who completed 16 years of education and above, approximately 65% of them married women who had completed fewer years of education; therefore, fathers generally were the more educated of the two parents. As Tinto (1975) pointed out, students who persist in college tend to come from families in which parents are more educated and have higher expectations than parents of other students (p. 100). It is possible to extend Tinto's point and suggest that the parent with the greater level of education might also have greater influence over the children, in terms of the degree to which they persist and achieve in university.

In conclusion, the results pertaining to the seventh research question show that HSGPA, the Numerical Ability factor scores, the ITP TOEFL total scores, and the father's education are significant predictors of FOT. The results also indicate that the abilities, skills, and background factors that predict finishing the EAP program on time are similar, but not identical, to those that predict the EAP GPA. As far as I know, this is the first attempt to identify factors that predict finishing a foundation set of courses, like the EAP program, during the university career, either in Japan, or in other countries.

Research Question 8: Predicting Graduation on Time (GOT) and Graduation with Honors (GWH) from HSGPA, High School Grade Factor Scores, ITP TOEFL Total Scores, *Hensachi* Ratings, Gender, and Parents' Education

The eighth research question asked to what extent HSGPA, high school grade factor scores, ITP TOEFL total scores, *hensachi* rankings, gender, and parents' education predict graduating on time (GOT) and graduating with honors (GWH). It was hypothesized that HSGPA, one or two of the high school grade factor scores, ITP TOEFL scores, gender, and parents' education would predict GOT and GWH. The results are very different for GOT and GWH. For GOT, the results do not confirm the hypotheses regarding HSGPA, high school grade factor scores, gender, or parents' education. The only significant predictor of GOT was the ITP TOEFL total score. In contrast, the results mostly confirm the hypotheses concerning GWH; they indicate that HSGPA, the Language Ability Factor scores, ITP TOEFL total scores, and *hensachi* rankings are significant predictors of GWH.

Predicting GOT from ITP TOEFL Total Scores

As explained in the Methods section, I defined GOT as graduating in four years because this stricter definition tends to produce better results in analyses, according to Astin (1993). However, it should be noted that students who graduated in four years in this sample are not considered more academically successful than those who graduated in four and one-half or five years. A number of the students chose to extend their time in the university in order to take further coursework, or to time graduation according to entry into graduate school or jobs, for example. Therefore, the definition of GOT is somewhat arbitrary in that sense. Nevertheless, the group of students who graduated in four years was more academically successful in a number of ways. They efficiently completed a number of hurdles in their university careers, including the EAP program, TOEFL requirements for study abroad, one year of study abroad, and credit and GPA requirements for graduation, in only four years. In addition, these 185 students managed to complete all these requirements and still significantly outperform the 122 students who did not graduate on time; those who graduated on time had a mean final GPA of 3.38, compared to a mean final GPA of 3.15 for those who did not graduate on time ($p < .001$).

The results for GOT indicate that the only significant predictor was ITP TOEFL total score. This result is unsurprising. One expects students who have a relatively high level of English proficiency to advance quickly through all parts of an English-medium university program. In general, these students most likely begin the EAP program in the highest level, and therefore require only one semester of study in that program. Afterwards, they more quickly attain the required TOEFL score of 500 for entering the regular university program, as well as the required 550 for study abroad. Their English language skills would help them at each phase of the program and enable them to progress efficiently through the four years at the university. Even though students most likely reach the cut-off point of language proficiency by about their first year of regular university study, it appears that the advantage afforded them early on by their high level of English proficiency continues to persist. Perhaps the success they experience early in their college careers bestows on them a sense of confidence and an expectation that they will also succeed in later work. In fact, as noted earlier, they do succeed in their later years at university as well.

Other results concerning predictors of GOT are unexpected. The hypotheses that HSGPA, a number of high school grade factor scores, gender, and parents' education would also predict GOT, were not confirmed.

The hypothesis that HSGPA and one or more high school grade factor scores would predict GOT was primarily based on the work of Burton and Ramist (2001), who stated that HSGPA and SAT scores predicted "academic outcomes of college," including graduation itself. However, they also noted that these measures do not predict graduation as well as they predict university GPA and graduating with honors (p. 26). The results of this study show that academic measures do not predict GOT. An explanation for these results can be found from Zwick (2002), who made the point that completing a degree is influenced by many non-academic factors, including family responsibilities, health, and finances (p. 94). Likewise, whether a student decides to graduate in four years is affected by a number of such factors. Zwick said that, due to the influence of these kinds of factors, "any measure of academic performance" will probably not predict graduation well, especially for a sample within one school (p. 94). These results are based on a sample within one school, and the results showed that neither HSGPA nor *hensachi* rankings was a significant predictor of GOT.

The results also showed that gender is not a significant predictor of GOT. Although some researchers have noted that female students tend to perform better than male students in university (Astin & Oseguera, 2005; Perez, 2004), these results show that more male students graduate on time than female students (60.37%, compared to 53.89%), though these differences are not statistically significant. For this sample of students, the results of other research questions have shown that female students outperform male students in HSGPA and EAP GPA, but beginning in the first year of regular university study, performance differences between the sexes become insignificant. The differences in GOT are also insignificant.

The results show that parents' education also does not predict GOT. Though a number of researchers have reported that the level of parents' education predicts persistence and degree completion (Nora, Barlow, & Crisp, 2005; LeTendre, 1996), for this sample of students, the level of parents' education was not a statistically significant predictor of GOT. However, it should be noted that, for the first time, the mother's education was close to statistical significance in predicting GOT. For example, the *p* value of the mother's education in the logistic regression analysis using HSGPA was .09, and .07 in the analysis using high school grade factor scores (see Tables 69 and 70). In short, despite a lack of statistical significance, there is some indication that the mother's education might be more important in predicting GOT than it was in predicting other outcomes.

Explaining this result is difficult because many factors might account for this finding. However, the data indicate that, because the students included in each group for GOT and FOT differ, the average level of the mother's education also differs (although the father's average educational level remains nearly the same). For example, in this GOT group, the percent of mothers who had completed 16 or more years of schooling is higher. In the FOT group, 23.31% of the mothers had finished 16 or more years, but in the GOT group, 27.21% of the mothers had done so. Admittedly, this difference is small, but this group of more educated mothers might have affected their children's college performance more than mothers in the less highly educated group. At present, other reasons for this difference do not present themselves.

In conclusion, the results of the analysis only partially confirm the hypotheses of the eighth research question concerning graduating on time, and show that the ITP TOEFL total score is a significant predictor of GOT. These results add to the literature concerning the TOEFL test, as well as to the second language acquisition literature, in widening the range of what is known about the predictive validity of the ITP TOEFL for university performance in an English-medium university in Japan. The negative results for HSGPA, *hensachi* rankings, gender, and parental education also point out that there are aspects of university performance that are not predicted by the usual measures of academic achievement, or by variables that sometimes predict achievement, such as gender and parents' education.

Predicting GWH with HSGPA, the Language Ability Factor Scores, ITP TOEFL Total Scores, and *Hensachi* Rankings

As explained in the Methods section, GWH is defined as graduating with a GPA of from 3.6 to 4.0; therefore, GWH is a clear measure of academic performance. As hypothesized, the results show that HSGPA, one high school grade factor score (the Language Ability factor scores), and the ITP TOEFL total score, are significant predictors of GWH; unexpectedly, the *hensachi* rankings also are significant predictors of GWH. However, gender and parents' education are not significant predictors of GWH.

Most of the results confirm the hypotheses of the eighth research question concerning GWH, which were based on Burton and Ramist (2001), who asserted that measures of academic ability predict university GPA, as well as departmental and university honors (p. 26). Table 71 shows that HSGPA is the most powerful predictor of graduating with honors. For every unit change in HSGPA, students have a 4.42 greater chance of this outcome. When HSGPA is broken down into the three factor scores, the Language Ability factor scores are significant predictors of graduating with honors. This result is not surprising because the Language Ability factor scores are consistently important predictors of GPA; unlike other factor scores, they are significant predictors of all three levels of GPA.

The Language Ability factor scores are the most predictive of the factor scores, and, despite the name assigned to them, it is also possible that these scores represent a general intelligence factor as well. As noted in Table 20, the highest number of high school grades load onto this factor, making it the predominating factor. Jensen (1998) described a factor that predominates in this way as a "ubiquitous superfactor" or the *g* factor of general mental ability (p. 115). He explained that the *g* factor is the "largest measurable source of common factor variance" in school performance, both in the form of grades and achievement tests, and accounts for at least 50% of the variance in school achievement (p. 115). Considering the likelihood that the Language Ability factor score is the best representative of *g*, it is not surprising that it predicts GPA at all levels, as well as graduation with honors.

The *hensachi* rankings were not hypothesized to predict GWH, but the results indicated that these rankings are significant predictors. It is not possible to compare these results with those of past studies because *hensachi* rankings have not been previously studied. However, researchers such as Astin (1993) have noted that the variables that predict graduating with honors also predict undergraduate GPA. The results of other research questions show that the *hensachi* rankings predict university academic achievement in terms of first-year GPA and final GPA, in a manner similar to HSGPA, although HSGPA is always more predictive of achievement. Therefore, in light of these results, the finding that *hensachi* rankings also predict GWH is not surprising.

As hypothesized, ITP TOEFL total scores also are significant predictors of GWH, but only in the results shown in Table 72. In Table 71, the results for the first analysis using HSGPA showed that the ITP TOEFL total scores were not significant predictors, although the *p* value was close to significance, at .12. The results of other research questions have shown that the ITP TOEFL scores are consistently significant predictors of the EAP GPA, but also of other outcomes. For example, the results in Table 16 show that the ITP TOEFL scores are significant predictors of the first-year GPA in the second model, before the *hensachi* rankings are added. In addition, the ITP TOEFL total scores predict FOT and GOT. In short, the ITP TOEFL scores predict performance in an English program such as EAP, as expected, but they also predict general academic performance, to some degree. Therefore, it is reasonable that these scores play a role in predicting GWH, a clear measure of academic performance.

Female gender and parents' education were hypothesized to predict GWH, but the results indicated that they do not do so. As noted above, gender plays a significant role in predicting university academic performance, but only in terms of EAP GPA. By the time students begin regular university study, gender no longer appears to predict performance, and there is no indication that it is related to GWH. In addition, parents' education, and particularly the father's education, plays a role in predicting university achievement. However, in the case of GWH, parents' education is not a significant predictor.

In conclusion, the results pertaining to the eighth research question indicate that HSGPA, the Language Ability factor scores, the ITP TOEFL total scores, and the *hensachi* rankings are significant predictors of GWH. The measures that predict academic achievement in the first and final year of university are essentially the same as those that predict GWH.

As Burton and Ramist (2001) pointed out, there are few studies of cumulative GPA or of graduation with honors in the United States. As far as I know, this is the first time that factors predicting high academic achievement, in the form of graduating with honors, have been studied in Japan. These results contribute to what is known about academic success of Japanese students in

English-medium universities in Japan, and Japanese students studying in universities in other countries where English is spoken.

Research Question 9: Combining HSGPA, ITP TOEFL Total Scores, Hensachi Rankings, EAP GPA, First-Year GPA, and Final GPA in Path Models

The ninth research question was revised to ask how the cognitive variables, HSGPA, ITP TOEFL total scores, *hensachi* rankings, EAP GPA, first-year GPA, and final GPA interrelate and constitute parts of a path model for first-year GPA and for final GPA. It was hypothesized that HSGPA would predict all levels of university achievement, the ITP TOEFL total scores would predict EAP GPA, the *hensachi* rankings would predict first-year and final GPA, and all previous levels of university GPA would predict later GPAs.

The results indicate that HSGPA is a significant predictor of EAP GPA and first-year and final GPA; the ITP TOEFL scores are significant predictors of EAP GPA; the *hensachi* rankings are significant predictors of first-year GPA; and EAP GPA is a significant predictor of first-year GPA. All hypotheses were confirmed for the first-year GPA path model and most of the hypotheses were confirmed for the second path model that extends through the final GPA. In the final GPA path model, all hypotheses were confirmed except that the path from the *hensachi* rankings to final GPA was not part of the final path model. The final GPA path model using a path from the *hensachi* rankings to only first-year GPA, and a data-driven path from the ITP TOEFL scores to the first-year GPA produced a well-fitting model.

It should be noted that the path models were first tested on the 2005 to 2008 data because these data were used for most of the analyses. Using these data, the path model results for first-year GPA were like those presented in the Results chapter, and the results for the final GPA path model were as good as those of the first-year path model. However, the sample size for the final path model using the 2005 to 2008 data was only 165. Because a larger sample size (N = 295) was available using the 2004 to 2008 data, it was decided to use those data. With the larger sample from the 2004 to 2008 data, the final model, with only one path from the *hensachi* rankings to first-year GPA, and a data-driven path from ITP TOEFL to first-year GPA, displayed good fit, though slightly less so than the model produced with the 2005 to 2008 data, because the confidence intervals for RMSEA extended beyond .08.

There are many decisions to make when constructing a path model. As Boomsma (2000) pointed out, even models that show good fit to the data can be further evaluated in terms of how well the estimates match the expectations based on theory. In addition, the fact that one model fits the data well does not preclude the possibility of other models also fitting the data. In the case of this model, other possibilities would be a path from TOEFL to final GPA, from *hensachi* to final GPA, and from *hensachi* to EAP GPA. It would also be possible to delete paths, but doing so would contradict the theory tested thus far. As Boomsma explained, such tinkering with parameters is “like touching a spider’s web” in that one small change can affect many of the parameters in important ways (p. 475). In addition, she and other authorities who write about path analysis warn against efforts to improve fit without reference to theory (p. 475). It was decided to aim for and retain the simplest possible model that is based on theory.

As was often true in the discussion of other research questions, it is not possible to compare these results with those of past studies of Japanese students. However, the results of this study confirm the assertions of researchers in the United States and elsewhere who have said that HSGPA is consistently the most powerful predictor of university academic achievement (e.g., Astin, 1993; Burton & Ramist, 2001; Ramist, Lewis, & McCamley-Jenkins; Zwick, 2002).

These results also confirm the work of researchers who have found the TOEFL test to be a significant predictor of university academic achievement (e.g., Johnson, 1988; Light & Teh-yuan; Light, Xu, & Mossop, 1987). As expected, the ITP TOEFL total test scores predict EAP GPA; however, the best path model shows that these test scores also predict first-year GPA. Though the path from ITP TOEFL to first-year GPA was not originally hypothesized because the results of a number of previous studies and the regression analysis did not suggest it, the results of other research questions indicate that, despite a lack of statistical significance in predicting GPA after the EAP program, the TOEFL continues to play a role in university achievement. The results of other research questions show, for example, that the ITP TOEFL total scores significantly predict GOT and GWH. Therefore, in light of the results of other research questions, it is not surprising that the ITP TOEFL shows importance beyond EAP GPA in the path model.

These results also show that the *hensachi* rankings are valuable in predicting university success. In the results of other research questions, as well, they have been significant predictors of academic achievement, although not as valuable as HSGPA. These results are the first to show that these rankings have merit.

In conclusion, these results are the first to show the relationship among cognitive variables and EAP GPA, first-year GPA, and final GPA in an English-medium university in Japan. These results add to international educational research literature, as well as to second language acquisition literature by showing that (a) HSGPA is an important predictor of university success, just as it is in other countries; (b) English proficiency appears to impact academic performance even after students have reached a cut-off point that should allow them to operate academically in their L2 at approximately the same level that they operate in their L1; (c) *hensachi* rankings have importance in predicting academic achievement in Japanese universities; and (d) academic performance in an EAP program significantly predicts performance in the regular program in an English-medium university.

Implications for University Admissions in Japan

Information and results from the current study hold practical value for admissions officials in Japan. For example, regression could be made use of in Japan, just as it is in admissions departments in other countries. In the United States, for example, officials routinely use a regression equation that typically includes the HSGPA and test scores from the SAT or ACT, in order to predict the first-year university GPA of applicants, and then choose from among applicants on the basis of that predicted GPA (Zwick, 2002, pp. 80-81). Admissions officials in Japan could also make use of a regression equation, after some initial study of the predictive validity of such variables for their own students. The variables that appear to have the most potential so far for inclusion in such a regression formula would be the HSGPA and *hensachi* rankings. For English-medium universities, ITP TOEFL scores also should be included. If Japanese admissions officials were to use a regression formula in the above manner, at least as an initial step in the selection process, it would bring greater standardization and fairness to the endeavor.

Secondly, the current study revealed that most of the section tests of the Center Examination lacked predictive validity, a result that should concern those who work in university admissions offices. Clearly, more research is needed, but the results of this study indicate that only the English and mathematics section tests have any power in predicting university GPAs. If admissions directors across Japan worked more with the NCUEE, the organization producing the Center Examination, and called for, as well as participated in, the kind of open research that has been conducted on the SAT in the United States, the chances of improving the only national entrance examination in Japan would increase.

The current study also revealed problems with the university entrance examinations studied. Although results indicated that some types of reading and writing tasks used as part of the entrance examinations predicted later university success to some degree, the majority of the tasks lacked predictive validity. Clearly, universities should do more to ensure that these expensive examinations have a respectable degree of validity and reliability. Therefore, admissions directors should be as proactive as possible in constructing testing committees with faculty members who have testing expertise, evaluating and revising their own university entrance examinations, and looking for new and innovative testing options. A new direction for entrance examinations that was discussed in the review of literature is tests of practical intelligence (Sternberg, 2004). The university under study is currently experimenting with testing this kind of intelligence, and to date, the results seem promising.

Admissions directors should also pursue a variety of avenues for research. For example, if it could be determined that the individual *hensachi* rankings are sufficiently standardized, these rankings might hold greater promise as admissions criteria than the high school rankings, which, as has been discussed, are problematic in that they are not truly individual difference variables. Another variable that could have value for admissions is the class rank of students. If high schools could provide this information in transcripts, it could prove useful in predicting university success because class rank is a clear indicator of individual achievement.

Finally, admissions directors should evaluate and revise their admissions system and procedures regularly. Over time, schools, curricula, and students change; therefore, directors should adjust policies and procedures regularly, and strive to create a system that is up-to-date and fair to applicants. In addition, directors should continue to look for ways to offer university admission to students with potential for success, even if they have been unable to gain entrance in standard ways. Admissions directors and staff members have a very difficult task; they must create and adhere to an objective and standardized system that is based on numbers, but still retain the ability to see the potential of the people behind those numbers, as well as to aim for the overall good of the student body. As Zwick (2004) pointed out, admissions departments should consider more than past measures of achievement, and prioritize other goals as well, such as boosting diversity on campus.

CHAPTER 6

CONCLUSION

In this chapter, I re-examine the results of the study, discuss the limitations, and suggest possibilities for future research.

Summary of Study Findings

The three parts of this study produced a variety of results. The first part of the study was designed to determine the degree to which student application information predicts later university success. The cognitive variables, HSGPA, ITP TOEFL, the *hensachi* rankings, high school grade factor scores, ITP TOEFL section scores, and the non-cognitive variables, gender and parents' education were examined to determine how well each predicted EAP GPA, first-year GPA, and final GPA. The first four research questions were answered in the first part of the study.

The results of the first four research questions showed that all of the cognitive variables were significant predictors of GPA at times, confirming most of the hypotheses. For example, as hypothesized, HSGPA was a consistently significant predictor of all levels of university GPA; the ITP TOEFL total scores predicted EAP GPA; the *hensachi* rankings were significant predictors of the first-year GPA and final GPA; the high school grade factor scores and ITP TOEFL section scores varied in how each predicted GPA; female gender predicted HSGPA and EAP GPA; and the father's education predicted HSGPA (though the correlation was negative). However, a few hypotheses were not confirmed: The ITP TOEFL total scores were not significant predictors of first-year GPA or final GPA (though the result was close to significance for each); female gender did not predict first-year GPA or final GPA; and the parents' education did not predict any level of university GPA, though the father's education came close to statistical significance. In summary, the confirmation of many of the hypotheses for the first four research questions indicates that student application information contains valuable information for admissions officials at Japanese universities.

The second part of the study was designed to determine the degree to which Center Examination test scores and university entrance examination type and test scores predict EAP GPA and first-year GPA. The fifth and sixth research questions were answered in the second part of the study, and the hypotheses were partially confirmed. As hypothesized, the Center Examination English test score (from 2006 and after) was a significant predictor of EAP GPA, the type of entrance examination was not a significant predictor of GPA, and the entrance examination English essay was a significant predictor of EAP GPA. However, other hypotheses were not confirmed: The Center Examination English test score did not predict first-year GPA or final GPA, and the entrance examination English essay scores did not predict first-year GPA and final GPA. In addition, though not hypothesized to be of importance, the Center Examination Math test scores and the entrance examination English test scores were both significant predictors of first-year GPA.

In summary, two hypotheses were confirmed, and two of the tests that were not hypothesized to be significant were found to be predictors of GPA; these results indicate that the Center Examination and university entrance examinations already make use of a number of tests that predict university GPA. However, the unconfirmed hypotheses, the lack of uniformity in the entrance examination English test, and the lack of significance of the majority of these tests indicate that these tests do not consistently fulfill their task of predicting university achievement.

The third part of the study was designed to examine the “big picture” of how cognitive and non-cognitive variables interact over time in accounting for student success in the EAP and regular university program. The first aspect of the larger picture of student success was timely and exemplary completion of program requirements. The seventh and eighth research questions focused on this first aspect by determining the extent to which variables predict finishing the EAP program on time, graduating on time, and graduating with honors. The second aspect of the larger picture of academic success was the interaction of variables over time, from before the first year of university up to the graduation. The ninth research question focused on this second aspect.

The first aspect of the larger picture of academic success was timely and exemplary completion of program requirements, including FOT, GOT, and GWH. The results of the seventh research question confirmed most of the hypotheses for FOT: HSGPA, high school grade factor scores (Numerical Ability factor scores), ITP TOEFL scores, and the father’s education were significant predictors of FOT.

However, other hypotheses were not confirmed; gender and mother's education were not significant predictors of FOT. The results of the eighth research question did not confirm most of the hypotheses for GOT. Only ITP TOEFL scores were significant predictors; HSGPA, high school grade factor scores, gender, and parents' education were not significant predictors of GOT. On the other hand, the results for GWH confirmed most of the hypotheses: HSGPA, high school grade factor scores (Language Ability factor scores), and ITP TOEFL scores were significant predictors. In addition, the *hensachi* rankings were significant predictors of GWH, even though they were not hypothesized to be predictive. In summary, though FOT, GOT, and GWH each represent somewhat different aspects of timely and exemplary completion of program requirements, it is clear that English proficiency is a significant predictor of all three, and that the cognitive variables are significant predictors of FOT and GWH.

The ninth research question focused on the second aspect of the larger picture of academic success, the interaction of variables over time, from before the first year of university, up to graduation. The results for the ninth research question confirmed all hypotheses for the path model of academic performance up to the first-year GPA: HSGPA predicted all levels of university GPA, ITP TOEFL scores predicted EAP GPA, the *hensachi* rankings predicted first-year GPA, and the EAP GPA predicted first-year GPA. These hypotheses were also confirmed for the path model that extended up to the final GPA, except that the hypothesized path from the *hensachi* rankings to final GPA did not contribute to the model and was deleted, and an additional data-driven path from ITP TOEFL scores to first-year GPA was added in order to produce a well-fitting model. In summary, all levels of GPA were stable predictors of later GPAs; in addition, the *hensachi* rankings and the ITP TOEFL contributed most to the model up to the first-year GPA.

Limitations of the Study

Although I have spent many hours to ensure that this study was done as well as possible, there are a number of limitations regarding measurement of the variables, definitions, and generalizability that should be explained.

Many variables were studied: HSGPA and high school grade factor scores, EAP GPA, first-year GPA, final GPA, ITP TOEFL total and section scores, *hensachi* rankings, Center Examination scores, entrance examination scores, gender, parents' education, FOT, GOT, and GWH. The data were obtained from many sources, such as the university Student Records Office, the Admissions office, the EAP department, and student questionnaires. Because many people handled the data, it is unsurprising that I found many data-entry errors. As a result, a great deal of time was devoted to talking to people in each of the departments involved in order to correct those errors. However, despite the time and effort spent on making the data as correct as possible, there is still error involved in the measurement of variables that is outside my control. For example, a well-known source of unreliability and error is differences in grading standards and approaches. Because all the GPAs resulted from assessments of individual teachers in different high schools and university departments, these measures are not completely reliable, an issue that was discussed previously. Some researchers choose to adjust grades because of this issue, but such a solution is not without its problems. In short, using GPAs means that unreliability and error are issues that affect study results. An additional concern regarding the GPA is its measurement scale. For all practical purposes, GPA is measured on an interval scale. However, as Bond and Fox (2007) have pointed out, the distances between all points on this sort of scale are not truly equal. For example, the distance between grades of C- and C+ is not equal to the distance between an A- and A+ grade. In other words, it is more difficult to earn the grade at the high end of the scale than to earn the grade at the middle of the scale. Researchers understand that this kind of issue is a limitation of the GPA, but resolving the difficulty is usually not possible.

In the course of the study, it also became clear that the measurement of a number of variables was problematic. For example, the entrance examination English test was also not a reliable measure because the directions on the test were not consistent from one administration to another. In addition, having been an EAP faculty member at a number of universities, I am well aware of entrance examination English essay rating inconsistencies, another source of error. A final example is the parent education variable. Because students reported this information on questionnaires, there is a good chance of error there as well.

Another issue concerning the measurement of variables that was outside my control was what was not measured. Because there are many factors that impact academic achievement, it would have been useful to have included other variables. For example, it is likely that an achievement or aptitude test, personal *hensachi* scores, high school class rank, and family income would significantly predict GPAs, but those measures were not available. In addition, variables such as emotional intelligence, motivation, self-efficacy, and personality type must also impact achievement, but these variables were also not available. When using regression to predict an outcome, it is important to include pertinent predictors, but those variables are not always obtainable, as was true in this case.

An additional concern regarding measurement of the variables came from issues encountered when preparing the data for analysis. For example, although the data generally met all assumptions for the statistical analyses used, there was a concern with the skewness of some variables. This issue has been discussed at some length, and reasons were given why variables were not transformed. Nevertheless, there is some chance that the skew might have lowered correlations and regression coefficients. Another aspect of the data that was less than ideal was that the members of the 2004 cohort did not meet the same admission requirements that members of other cohorts had; in addition, they took only one ITP TOEFL test, while members of the other cohorts took the test twice. Because the second test was considered more reliable, it was necessary to use regression to impute scores for the second TOEFL score for these students. Because of a desire to use data that measured cohort members as uniformly as possible, I conducted most analyses using data from the 2005 to 2008 cohorts. However, at times, it was necessary to use data from all cohorts in order to have the largest possible sample. Perhaps differences between the 2004 cohort and the other cohorts partially accounted for the final GPA path model for the 2004 to 2008 group not being as well-fitting as the one produced for the 2005 to 2008 group.

The definitions used in the study also had a number of limitations. For example, academic success in this study is defined by GPA. Although GPA is certainly one good indicator of academic success, there are other indications, such as leadership in student government, involvement in study trips, publications in the school journal, and personal satisfaction with accomplishments, among other valid definitions. Including the variables, FOT, GOT, and GWH, was an attempt to extend the definition of success, but, as discussed earlier, the definition of GOT can be debated. For this sample of students, who have to first complete an EAP program, defining graduating on time as graduating within four years might have been too limited a definition, and might not have discriminated between more successful and less successful students. A final definition that requires further consideration is the label, Spatial Ability, for the third high school grade factor scores. Although there are legitimate reasons to use this label, as was discussed previously, the types of high school courses included under this label require a more diverse assortment of abilities than those included in the Language Ability and Numerical Ability factor scores; for example, Art also requires creativity, and Physical Education requires kinesthetic ability. A decision was made to label this set of factor scores Spatial Ability because this term is one apt description of the type of ability required by all three courses that loaded onto this factor, but this overall ability might have been better characterized in another way. In short, this definition might not be as precise as it should be.

Finally, there are limitations related to the generalizability of the study. As has been previously discussed, the results are most pertinent to Japanese university students attending an English-medium university in Japan. Study results might apply to other groups of students, but they might not. Similarly, the results of the study also do not necessarily apply to applicants who were not accepted into the university.

Zwick, Brown, and Sklar (2004) have pointed out another issue regarding this group of applicants who were not admitted to the university. This group of students had grades and test scores that were comparatively low; as a result, they were not admitted and consequently able to earn university GPAs. Because scores of these applicants were not included in the study analyses, range restriction and lowered correlations of variables with university GPAs resulted. If data had been available regarding this group of applicants, it would have been possible to statistically estimate the correlations for all applicants to the university. Such statistical corrections are not without problems; nevertheless, they are informative and might have allowed for clearer comparisons with other study results. Therefore, lack of these data resulted in another study limitation.

Suggestions for Future Research

There are many possibilities for future research. Because this area of study has been explored little in Japan, cognitive variables, including HSGPA and test scores, and non-cognitive variables offer many possible avenues for research. As noted, universities should conduct research on these variables for their own use, testing agencies such as the NCUEE should do further, more public research on the Center Examination, and researchers should continue to study these variables. In discussing specific areas for future research below, I suggest possible research questions for each.

HSGPA and High School Grade Factor Scores

The HSGPA and the factor scores should be studied further, and the university GPA should be examined in a similar fashion. As noted, the Spatial Ability factor scores particularly deserve further attention because this group of factor scores is not as uniform as the Language Ability and Numerical Ability factor scores. In other words, the high school courses that loaded onto the Spatial Ability factor required students to use a range of kinesthetic, creative, and spatial abilities. How to best represent this group of abilities should be examined further. Another reason that the Spatial Ability factor scores are worthy of further study is that the abilities represented by these scores appear to be clearly related to success in the EAP program, but not in the regular university program. A reason was suggested for this difference, but further study is needed. In addition, the relationship of the Language Ability factor scores to the *g* factor of intelligence is certainly worth investigating further. Finally, because the high school class rank of each student, if available, might prove to predict her university GPA, it should be further investigated. Possible research questions are: Can university grades be represented by three similar factors? What aspects of EAP performance are most correlated with the Spatial Ability factor scores? Do the Language Ability factor scores correlate more highly than the other factor scores with other measures of aptitude or ability, such as IQ tests or achievement tests? Do high schools maintain records of class rank for each student? If available, does class rank predict university GPA?

***Hensachi* Rankings**

The *hensachi* rankings should also be investigated further. Both admissions officials and researchers should continue to study this variable. Even though it was determined that the *hensachi* rankings did not moderate HSGPA, and that they operated independently in this study, these variables both are significant predictors of first-year GPA and final GPA. More study is needed to examine the role each plays in predicting GPAs. In addition, an interesting possibility for future research is the personal *hensachi* score earned by each student. Though these scores are now considered confidential and are not used in the university application process, it seems likely that these individual scores would be even better predictors of university achievement. Possible research questions include: How are *hensachi* rankings and HSGPA alike in predicting university GPA? How are *hensachi* rankings and HSGPA different in predicting university GPA? Do individual *hensachi* scores predict university GPA better than high school *hensachi* rankings? Are individual *hensachi* scores sufficiently standardized and reliable as measures of individual achievement?

Cut-Off Point of English Proficiency

More study should be devoted to a possible cut-off point of English proficiency using ITP TOEFL scores. Although this cut-off point was not a focus of close examination in the current study, it would be an interesting object of further study. The ITP TOEFL section scores are also potential objects of further study. For example, results pertaining to the Section III scores indicated that reading comprehension might continue to play an important role in university success, even after the cut-off point is reached. In addition, the Section II structure and written expression scores did not predict any level of university GPA, at least for this sample of Japanese university students. More study should be devoted to the predictive value of these scores. Possible research questions include: What is the cut-off point of English proficiency in an English-medium university in Japan? Does the cut-off point vary by gender? Are Section III reading comprehension scores the most important predictor of university GPA? Are Section II structure and written expression scores significant predictors of university GPA for students at any level of English proficiency?

Non-Cognitive Variables

Further study of the non-cognitive variables is also indicated. The results of this study suggest that these variables play a part in academic achievement, but the results were not definitive. More researchers should look at academic abilities and patterns of achievement for both female and male students. In order to understand these patterns in any depth, qualitative research, in the form of case studies and interviews, should be carried out. Similarly, the effect of parents' education should be investigated further, particularly the relationship of the fathers' education to the HSGPA and university GPA of their children. There are many other non-cognitive variables worth studying, as well, such as motivation, self-efficacy, emotional intelligence, personality, study habits and work ethic, class attendance habits, self-regulation, commitment, and the effect of entering university as a *ronin*; in addition, a number of variables that affect students' academic performance after matriculation are worth studying, such as participation in club activities and commuting distance to school (Dickinson & O'Connell, 1990; Grewal & Salovey, 2005; Gump, 2005; Pintrich, 1995; Rau & Durand, 2000; Sheard & Goldby, 2007). As noted, other types of mixed-methods and qualitative investigations that make use of interviews and case studies would help to clarify the role that these variables play. Only by talking with students directly will it be possible to understand their perceptions and experiences, and to determine in some depth why some are successful in university and others are not. Possible research questions are: How does the prediction of university GPA vary by gender in an English-medium university in Japan? How does motivation and academic performance of female and male students change during their university careers? What effect do mothers and fathers have on the academic performance of their children, according to students?

The Center Examination and University Entrance Examinations

There is also no doubt that the Center Examination and entrance examinations must be studied further. Although there are a number of positive results for these tests, they do not yet adequately fulfill their primary role of predicting university success for the sample of participants studied. The NCUEE could aid such research, as well, by examining and publishing results concerning validity and reliability of the Center Examination, as well as providing information about efforts made to avoid gender bias. Possible research questions include: Does the Center Examination Social Studies section test score predict university first-year or final GPA with a larger sample of participants? Do the Center Examination section tests correlate with other similar measures, such as tests and high school grades in the same subject areas?

Final Conclusions

This study makes a number of contributions to SLA and to international educational research. It is the first to examine the predictive validity of cognitive and non-cognitive variables for university achievement in Japan, and the results of this study and those of studies abroad were often parallel in that variables such as HSGPA, ITP TOEFL, a number of test scores, and non-cognitive variables (at times) were predictive of university success. In addition, the Japanese *hensachi* rankings were studied for the first time and found to predict university success. This study is the first to evaluate the predictive validity of entrance examinations and the Center Examination. Although the results were less than stellar, the news was not all bad for the testing agencies. A number of these tests predicted university GPA, at least at one point in the students' university career.

This study is also the first to examine a number of variables and their effects on the academic performance of Japanese university students. For example, high school grade factor scores and the FOT variable were created and studied for the first time. Using high school grade factor scores in the regression analyses shed light on how school abilities might be related to success in particular types of coursework, and at particular junctures in the four years at university, and the analysis with FOT indicated that completing an earlier program requirement reflects academic success and requires many of the same abilities that earning a GPA requires. Finally, the "big picture" of academic success in Japan as illustrated through analysis of GOT, GWH, and path models was also examined for the first time, and offered a wider look at the academic success of students in an English-medium university in Japan.

The results of this study have potential value for university admissions officers and administrators, as they consider student applications, as well as those who write university entrance examinations and produce the Center Examination. However, changes in admissions and testing procedures policies might come slowly in Japan. The tradition of Japanese universities creating and using their own entrance examinations has been long-standing, and might be slow to change. In addition, many universities soon might not be in a position to reject applicants. As Kikuchi (2006) pointed out, the fact that the population of students applying to enter universities in Japan is shrinking means that most students who want to enter university will find a place there, “if they are not particular about which institution they attend” (p. 78). Such a reality might mean that a number of universities will not need or use the results of this kind of research.

Nevertheless, many universities are still in a position to choose among applicants because many students continue to aim for particular universities. For example, the university under study rejects hundreds of applicants each year. Therefore, despite the changing demographics in Japan, there is still a need for universities to be aware of variables that predict academic success so that they might choose candidates who are most likely to take advantage of the opportunity to enroll in their institutions.

In conclusion, the results of this study have potential value for a number of audiences. It is hoped that these results are useful for researchers interested in student success, as well as for those involved in testing and admissions, so that decisions might be enhanced, and the most deserving of students receive the opportunity to enter the universities of their choosing. It is also hoped that this study provides an impetus to positive change in the university admissions process in Japan.

REFERENCES

Alderman, D. L. (1981). *Language proficiency as a moderator variable in testing academic aptitude* (TOEFL Research Report No. RR-10). Princeton, NJ: Educational Testing Service.

Al-Musawi, N. M., & Al-Ansari, S. H. (1999). Test of English as a Foreign Language and First Certificate of English tests as predictors of academic success for undergraduate students at the University of Bahrain. *System*, *27*, 389-399.

Amano, I., & Poole, G. S. (2005). The Japanese university in crisis. *Higher Education*, *50*, 685-711.

Ardila, A. (2001). Predictors of university academic performance in Colombia. *International Journal of Educational Research*, *35*, 411-417.

Armor, D. J. (1974). Theta reliability and factor scaling. *Sociological Methodology*, *5*, 17-50.

Astin, A. W. (1993). *What matters in college? Four critical years revisited*. San Francisco: Jossey-Bass.

Astin, A. W., & Oguera, L. (2005). Pre-college and institutional influences on Degree attainment. In S. Seidman (Ed.), *College student retention: Formula for student success* (pp. 245-276). Westport, CT: American Council on Education and Praeger.

Ayers, J. B., & Peters, R. M. (1977). Predictive validity of the test of English as a foreign language for Asian graduate students in engineering, chemistry, or mathematics. *Educational and Psychological Measurement*, *37*, 461-463.

Ayers, J. B., & Quattlebaum, R. F. (1992). TOEFL performance and success in a masters program in engineering. *Educational and Psychological Measurement*, 52, 973-975.

Bachman, L. F. (1991). What does language testing have to offer? *TESOL Quarterly*, *25*(4), 671-704.

Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge: Cambridge University Press.

Bayliss, D., & Raymond, P. M. (2004). The link between academic success and L2 proficiency in the context of two professional programs. *The Canadian Modern Language Review*, *61*(1), 29-51.

Bacon, D. R., & Bean, B. (2006). GPA in research studies: An invaluable but neglected opportunity. *Journal of Marketing Education*, *28*(1), 35-42.

Bedeian, A. G., Day, D. V., & Kelloway, E. K. (1997). Correcting for measurement error attenuation in structural equation models: Some important reminders. *Educational and Psychological Measurement*, *57*(5), 785-799.

Bentler, P. M. (2005). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software (www.mvsoft.com).

Betts, J. R., & Morell, D. (1999). The determinants of undergraduate grade point average. *The Journal of Human Resources*, *34*(2), 268-293.

Bond, T. G. & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

Boomsma, A. (2000). Reporting analyses of covariance structures. *Structural Equation Modeling*, *7*(3), 461-483.

Bridgeman, B. & Lewis, C. (1994). The relationship of essay and multiple-choice scores with grades in college courses. *Journal of Educational Measurement, 31*(1), 37-50.

Brown, J. D. (1988). *Understanding research in second language learning*. Cambridge: Cambridge University Press.

Brown, J. D. (1995). English language entrance examinations in Japan: Myths and facts. *The Language Teacher*, *19*(10), 21-26.

Brown, J. D., & Yamashita, S. O. (1995a). English language tests at Japanese universities: What do we know about them? *JALT Journal*, *17*(1), 7-30.

Brown, J. D., & Yamashita, S. O. (1995b). English language entrance examinations at Japanese universities: 1993 and 1994. In J. D. Brown & S. O. Yamashita (Eds.), *Language testing in Japan* (pp. 86-106). Tokyo: Japan Association for Language Teaching.

Brown, J. D. (2004). What do we mean by *bias*, *Englishes, Englishes in testing*, and *English language proficiency*? *World Englishes*, *23*(2), 317-319.

Bull, R., Johnston, R. S., & Roy, J. A. (1999). Exploring the roles of the visual-spatial sketch pad and central executive in children's arithmetical skills: Views from cognition and developmental neuropsychology. *Developmental Neuropsychology*, *15*(3), 421-442.

Burg, S. S. (2008, March). An investigation of dimensionality across grade levels and effects on vertical linking for elementary grade mathematics achievement tests. Paper presented at the meeting of the National Council on Measurement in Education, New York, NY.

Burton, N. W., & Ramist, L. (2001). *Predicting success in college: SAT studies of classes graduating since 1980.* (College Board Research Report No. 2001-2). New York, NY: College Entrance Examination Board.

Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd ed.). Mahwah, NJ: Erlbaum.

Camara, W. (1998). *High school grading practices*. (College Board Research Note No. RN-04). New York, NY: College Entrance Examination Board.

Christopher, V. L. (1993). Direct and indirect placement test scores as measures of language proficiency and predictors of academic success for ESL students. Unpublished master's thesis, University of British Columbia, Vancouver, British Columbia, Canada.

Clark, E. L. (1964). Reliability of grade-point averages. *The Journal of Educational Research*, *57*(8), 428-430.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation for the behavioral sciences* (3rd ed.). Mahwah: NJ: Erlbaum.

Cornwell, S., Simon-Maeda, A., & Churchill, E. (2007). Selected research on second-language teaching and acquisition published in Japan in the years 2000-2006. *Language Teaching, 40*, 119-134.

Cummins, J. (1980). The cross-lingual dimensions of language proficiency: Implications for bilingual education and the optimal age issue. *TESOL Quarterly*, *14*(2), 175-187.

Cummins, J. (1982). Tests, achievement, and bilingual students. *FOCUS*, *9*, 1-7.

Cutrona, C. E., Cole, V., Colangelo, N., Assouline, S. G., & Russell, D. W. (1994). Perceived parental social support and academic achievement: An attachment theory perspective. *Journal of Personality and Social Psychology*, *66*(2), 369-378.

Dickinson, D. J., & O'Connell, D. Q. (1990). Effect of quality and quantity of study on student grades. *Journal of Educational Research*, *83*(4), 227-231.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Erlbaum.

Educational Testing Service. (2006). *Institutional Testing Program: Manual for Supervisors*. Princeton, NJ: Educational Testing Service.

Elliott, R., & Strenta, A. C. (1988). Effects of improving the reliability of the GPA on prediction generally and on comparative predictions for gender and race particularly. *Journal of Educational Measurement*, *25(*4), 333-347.

Ellis, R. (1985). *Understanding second language acquisition*. Oxford: Oxford University Press.

Ferketich, S. (1990). Focus on psychometrics: Internal consistency estimates of reliability. *Research in Nursing and Health*, *13*, 437-440.

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage.

Firebaugh, G. (1978). A rule for inferring individual-level relationships from aggregate data. *American Sociological Review*, *43*(4), 557-572.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Gardner, H. (1990). The theory of multiple intelligences. In N. Entwistle (Ed.), *Handbook of educational ideas and practices* (pp. 930-938). London: Routledge.

Gardner, H., & Moran, S. (2006). The science of multiple intelligences theory: A response to Lynn Waterhouse. *Educational Psychologist, 41*(4), 227-232.

Graham, J. G. (1987). English language proficiency and the prediction of academic success. *TESOL Quarterly, 21*(3), 505-521.

Green, S. B., & Salkind, N. J. (2005). *Using SPSS for Windows and Macintosh: Analyzing and understanding data*. Upper Saddle River, NJ: Pearson Prentice Hall.

Grewal, D., & Salovey, P. (2005). Feeling smart: The science of emotional intelligence. *American Scientist, 93(4)*, 330-339.

Guest, M. (2008). A comparative analysis of the Japanese university entrance *Senta Shiken* based on a 25-year gap. *JALT Journal*, *30*(1), 85-104.

Gump, S. E. (2005). The cost of cutting class: Attendance as a predictor of student success. *College Teaching, 53*(1), 21-26.

Hale, G. A., Stansfield, C. W., & Duran, R. P. (1984). *Summaries of studies involving the TOEFL test, 1963-1982* (TOEFL Research Report No. RR-16). Princeton, NJ: Educational Testing Service.

Hashimoto, K. (1998). *Mechanisms to implement and maintain policy agendas for teaching English as a foreign language in Japan.* Paper presented at the annual meeting of the Association for Active Educational Researchers, Adelaide, Australia.

Hayashi, A. (2005). Japanese education system and National Center Test. Retrieved November 6, 2009 at http://rche.kyushu-u.ac.jp/~hayashi/haifu/China05/3DNCTest_China2_Teiji4.pdf.

Hecht, S. A. (2001). Latent variable modeling and cross-national differences in college-level academic performance. *International Journal of Educational Research*, *35*, 419-425.

Hoffman, J. L., & Lowitzki, K. E. (2005). Predicting college success with high school grades and test scores: Limitations for minority students. *The Review Of Higher Education*, *28*(4), 455-474.

Holloway, S. D., Fuller, B., Hess, R. D., Azuma, H., Kashiwagi, K.,& Gorman, K. (1990). The family's influence on achievement in Japan and the United States. *Comparative Education Review*, *34*(2), 196-208.

Ito, A. (2005). A validation study on the English language test in a Japanese nationwide university entrance examination. *The Asian EFL Journal Quarterly*, *7*(2), 90-116.

Jensen, A. (1998). The *g* factor and the design of education. In R. J. Sternberg & W. M. Williams (Eds.), *Intelligence, instruction, and assessment: Theory into practice* (pp. 111-131). New York: Routledge.

Johnson, P. (1988). English language proficiency and academic performance of undergraduate international students. *TESOL Quarterly, 22*(1), 164-168.

Jürges, H. & Schneider, K. (2004). International differences in student achievement: An economic perspective. *German Economic Review*, *5*(3), 357-380.

Kamiya, S. (2009, January 20). College entrance exams: “Exam hell” now not so hot. *The Japan Times Online*. Retrieved March 15, 2010, from http://search.japantimes.co.jp/print/nn20090120il.html

Kikuchi, K. (2006). Revisiting English entrance examinations at Japanese universities after a decade. *JALT Journal*, *28*(1), 77-96.

Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.

Kline, R. B. (1991). Latent variable path analysis in clinical research: A beginner’s tour guide. *Journal of Clinical Psychology*, *47*(4), 471-484.

Kobayashi, H., & Rinnert, C. (2002). High school student perceptions of first language literacy instruction: Implications for second language writing. *Journal of Second Language Writing, 11*, 91-116.

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). Validity of the SAT for predicting first-year college grade point average. *College Board Report No. 2008-5*. New York: College Entrance Examination Board.

Lakin, J. (2010, March 23). HLM 6.0: Short course. Retrieved July 15, 2010, from University of Iowa, Statistics Outreach Center at: http://www.education.uiowa.edu/StatOutreach/shortcourses/documents/0323-HLM-6-Handout.pdf

LeTendre, G. K. (1996). Constructed aspirations: Decision-making processes in Japanese educational selection. *Sociology of Education*, *69*, 193-216.

Li, H., & Wainer, H. (1998). *Toward a coherent view of reliability in test theory*. Princeton, NJ: Educational Testing Service.

Light, R. L., Ming, X., & Mossop, J. (1987). English proficiency and academic performance of international students. *TESOL Quarterly*, *21*(2), 251-261.

Light, R. L., & Wan, T. (1991). Soviet students at U.S. colleges: Social perceptions, language proficiency, and academic success. *TESOL Quarterly*, *25*(1), 179-185.

Linn, R. L. (1966). Grade adjustments for prediction of academic performance: A review. *Journal of Educational Measurement*, 3(4), 313-329.

Linn, R. I. (1990). Admission testing: Recommended uses, validity, differential prediction and coaching. *Applied Measurement in Education*, *3*(4), 297-318.

Mattern, K. D., Patterson, B. F., Shaw, E. J., Kobrin, J. L., & Barbuti, S. M. (2008). Differential validity and prediction of the SAT. *College Board Report No. 2008-4*. New York, NY: College Entrance Examination Board.

Mayer, J. D., Caruso, D. R., & Salovey, P. (1999). Emotional intelligence meets traditional standards for an intelligence. *Intelligence*, *27*, 267-298.

McVeigh, B. (1997). *Life in a Japanese women's junior college: Learning to be ladylike*. London: Routledge.

McVeigh, B. (2002). *Japanese higher education as myth*. Armonk, NY: M. E. Sharpe.

Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*, *30*, 955-966.

Mori, R. (2002). Entrance examinations and remedial education in Japanese higher education. *Higher Education, 43*, 27-42

National Center for University Entrance Examinations. (n.d.). Retrieved June 9, 2008, from http://www.dnc.ac.jp/

Neubauer, A. C. & Fink, A. (2005). Basic information processing and the psychophysiology of intelligence. In R. J. Sternberg & J. E. Pretz (Eds.), *Cognition & intelligence: Identifying the mechanisms of the mind* (pp. 68-87). Cambridge: Cambridge University Press.

Newfields, T. (2003). An interview with J. D. Brown. *Shiken: JALT Testing & Evaluation SIG Newsletter*, *7*(1), 10-13.

Newfields, T. (2006, December). Suggested answers for assessment literacy self-study quiz. S*hiken: JALT Testing and Evaluation SIG Newsletter*, *10*(2). Retrieved March 15, 2010, from http://jalt.org/test/SSAl.htm

Nora, A., Barlow, E., & Crisp, G. (2005). Student persistence and degree attainment beyond the first year in college. In S. Seidman (Ed.), *College student retention: Formula for student success* (pp. 129-153). Westport, CT: American Council on Education and Praeger.

Nurm, J., & Aunola, K. (2001). How does academic achievement come about: cross-cultural and methodological notes. *International Journal of Educational Research*, *35*, 403-409.

Ono, H. (2001). Who goes to college? Features of institutional tracking in Japanese higher education. *American Journal of Education*, *109*(2), 161-195.

Otsu, T. (2004) The unified test by the NCUEE—Its role in Japanese university admissions. Retrieved November 6, 2009 at http://www.rd.dnc.ac.jp/~otsu/doc/Tianjin2004Slide.pdf

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). Singapore: Wadsworth.

Perez, C. (2004). Reassessing college admissions: Examining tests and admitting alternatives. In R. Zwick (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (pp. 345-354). New York: RoutledgeFalmer.

Pike, G. R., & Saupe, J. L. (2002). Does high school matter? An analysis of three methods of predicting first-year grades. *Research in Higher Education*, 43(2), 187-207.

Pintrich, P. R. (1995). Understanding self-regulated learning. In P. R. Pintrich (Ed.), *Understanding self-regulated learning*. San Francisco: Jossey-Bass.

Poole, G. (2003). Assessing Japan's institutional entrance requirements. *Asian EFL Journal*, *5*(1), 1-11.

Porter, S. R. (2005). Introduction to multilevel modeling. Retrieved July15, 2010, from the Iowa State University Air Statistics Institute at: http://www.airweb.org/institutedata-display.asp?id=44

Ramist, L., Lewis, C., & McCamley-Jenkins, L. (1994). *Student group differences in predicting college grades: Sex, language, and ethnic groups* (College Board Report 93-1). New York: College Entrance Examination Board.

Ransdell, S. (2001). Predicting college success: The importance of ability and non-cognitive variables. *International Journal of Educational Research, 35*, 357-364.

Rau, W., & Durand, A. (2000). The academic ethic and college grades: Does hard work help students to "make the grade"? *Sociology of Education*, *73*, 19-38.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2004). *HLM6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Reason, R. D. (2003). Student variables that predict retention: Recent research and new developments. *NASPA Journal*, *40*(4), 172-191.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*(3), 351-357.

Rosenfeld, M., Oltman, P. K., & Sheppard, K. (2004). *Investigating the validity of TOEFL: A feasibility study using content and criterion-related strategies* (TOEFL Research Report 71). Princeton, NJ: Educational Testing Service.

Saville-Troike, M. (1984). What really matters in second language learning for academic achievement? *TESOL Quarterly*, *18*(2), 199-219.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350-353.

Schoenfeldt, L. F., & Brush, D. H. (1975). Patterns of college grades across curricular areas: Some implications for GPA as a criterion. *American Educational Research Journal*, *12*(3), 313-321.

Schreiber, J. B. & Griffin, B. W. (2004). Review of multilevel modeling and multilevel studies in *The Journal of Educational Research* (1992-2002). *The Journal of Educational Research*, *98*(1), 24-33.

Selvin, H. C. (1958). Durkheim's suicide and problems of empirical research. *The American Journal of Sociology*, *63*(6), 607-619.

Sheard, M., & Goldby, J. (2007). Hardiness and undergraduate academic study: The moderating role of commitment. *Personality and Individual Differences, 43*, 579-588.

Steele, C. M. & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797-811.

Sternberg, R. J. (2004). Theory-based admissions testing for a new millennium. *Educational Psychologist*, *39*(3), 185-198.

Sternberg, R. J. (2007). Finding students who are wise, practical, and creative. *The Chronicle of Higher Education*, 53, B11.

Sternberg, R. J., & Hedlund, J. (2002). Practical intelligence, g, and work psychology. *Human Performance, 15*(1/2), 143-160.

Stricker, J. S., Rock, D. A., Burton, N. W., Muraki, E., & Jirele, T. J. (1994). Adjusting college grade point average criteria for variations in grading standards: A comparison of methods. *Journal of Applied Psychology, 79*(2), 178-183.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (4th ed.). Upper Saddle River, NJ: Pearson Education.

Thorndike, E. L. (1939). On the fallacy of imputing the correlations found for groups to the individuals or smaller groups composing them. *The American Journal of Psychology*, *52*(1), 122-124.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, *45*(1), 89-125.

Travis, R., & Kohli, V. (1995). The birth order factor: Ordinal position, social strata, and educational achievement. *The Journal of Social Psychology, 135*(4), 499-507.

Vinke, A. A., & Jochems, W. M. G. (1993). English proficiency and academic success in international postgraduate education. *Higher Education*, *26*, 275-285.

Wakabayashi, T. (2002). Bilingualism as a future investment: The case of Japanese high school students at an international school in Japan. *Bilingual Research Journal*, *26*, 631-658.

Ward, W. (1997). Language proficiency and academic success: Relationships between proficiency in two languages and achievement among Mexican American students. *Bilingual Research Journal, 21*(4), 334-347.

Werts, C., Linn, R. L., & Jöreskog, K. G. (1978). Reliability of college grades from longitudinal data. *Educational and Psychological Measurement*, *38*(1), 89-95.

Westrick, P. (2005). Score reliability and placement testing. *JALT Journal*, *27*(1), 71-93.

Williams, M., & Burden, R. L. (1997). *Psychology for language teachers: A social constructivist approach*. Cambridge: Cambridge University Press.

Wilson, K. M. (1983). *A review of research on the prediction of academic performance after the freshman year*. (College Board Report No. 83-2). New York: College Entrance Examination Board.

Young, J. W. (1990a). Adjusting the cumulative GPA using item response theory. *Journal of Educational Measurement*, *27*(2), 175-186.

Young, J. W. (1990b). Are validity coefficients understated due to correctable defects in the GPA? *Research in Higher Education*, *31*(4), 319-325.

Young, J. W. (1993). Grade adjustment methods. *Review of Educational Research*, *63*(2), 151-165.

Young, J. W. (2004). Differential validity and prediction: Race and sex differences in college admissions testing. In R. Zwick (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (pp. 289-301). New York: RoutledgeFalmer.

Yu, C. H., Popp, S. O., DiGangi, S., & Jannasch-Pennell, A. (2007, October). Assessing unidimensionality: A comparison of Rasch modeling, parallel analysis, and TETRAD. *Practical Assessment, Research & Evaluation*, *12*(14). Retrieved February 1, 2010, from http://pareonline.net/genpare.asp?wh=0&abt=12

Zwick, R. (2002). *Fair game? The use of standardized admissions tests in higher education*. New York: RoutledgeFalmer.

Zwick, R. (2004). Is the SAT a "wealth test"? The link between educational achievement and socioeconomic status. In R. Zwick (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (pp. 289-301). New York, NY: RoutledgeFalmer.

Zwick, R. (2007, February). *College admission testing*. (Paper commissioned by the National Association for College Admission Counseling). Paper retrieved September 10, 2010, from http://education.ucsb.edu/rzwick/FinalNACAC.pdf

Zwick, R., Brown, T., & Sklar, J. C. (2004, July). California and the SAT: A reanalysis of University of California admissions data. Retrieved December 28, 2010, from the Center for Studies in Higher Education, UC Berkeley at: http://escholarship.org/uc/item/9hf3t044?query=Zwick

Zwick, R., & Green, J. G. (2007). New perspectives on the correlation of scholastic assessment test scores, high school grades, and socioeconomic factors. *Journal of Educational Measurement, 44*(1), 1-23.

APPENDICES

APPENDIX A

STUDENT QUESTIONNAIRE (JAPANESE VERSION)

教育研究の為のアンケートご協力のお願い

学生の皆さん

私は現在、本校で「学生達の学業での成功」に関する教育研究を行っています。
当アンケートでの回答内容は極秘扱いされ、この研究目的以外に使用されることはありません。また、私の研究目的はグループとしてのパターン解析です。すべての回答は集計されグループ情報として扱われ、また個人情報が公になる事もありません。
本日、皆さんにお願いするアンケートはあくまでも任意ですが、これは本校の将来にとっても大変有益な情報源となります。
以上を理解して頂いた上で、皆さんのご協力を宜しくお願いします。

2008 年 11 月 22 日
Kristy King Takagi

まず下記の欄に、学生番号と氏名を忘れずに記入して下さい。

学生番号___　　　　氏名__________________________

尚、不明の場合はそのまま空白にしておいて下さい。

１.　　生年月日 １９ __ 年 __ 月 __ 日

２.　　兄弟姉妹（数字を記入。いない場合は無記入）
兄　人　/ 姉　人 /　弟　人　/ 妹　人

３.　　血液型 (〇で囲む) A / B / O / AB

４.　　**資格、習い事**と、その**級、段、習得年数**を記入して下さい。
（英検 1 級、空手 2 段、ピアノ１２年、フラダンス 3 年等）

____

５.　　自分自身の、様々な事柄（学業、趣味、習い事、仕事等）で「成功する能力」を評価して下さい。（最高を 100、最低を 0 として）

____

6．　　実家で**兄弟以外**の同居人（父、母、祖父母、叔父等）の最終学歴　(〇で囲む)

父（　中学　高校　専門学校　短大　大学　修士　博士　）
母（　中学　高校　専門学校　短大　大学　修士　博士　）
祖父（　中学　高校　専門学校　短大　大学　修士　博士　）
祖母（　中学　高校　専門学校　短大　大学　修士　博士　）
その他（明記）__（　中学　高校　専門学校　短大　大学　修士　博士　）

7．　　あなたの周り（父母、兄弟、友達、先生等）の人たちで、あなたの学業での業績に、影響を与えた人は誰ですか。

____

ご協力ありがとうございました。

APPENDIX B

STUDENT QUESTIONNAIRE (ENGLISH VERSION)

A Request for your Cooperation with a Questionnaire to be used for Educational Research

To All Students:

I am now conducting educational research at your university about academic success of students. All answers to this questionnaire will be kept strictly confidential and used only for the purpose of this research. In addition, the objective of this research is to analyze group patterns of academic success; therefore, all responses to the questionnaire will be gathered together and treated as group information. No individual information will be made public. The questionnaire that I am asking you to complete today is completely voluntary. However, any information you provide could be very valuable and of great use to the university. I hope that your understanding of the above will permit you to cooperate by answering the questions below.

November 22, 2008
Kristy King Takagi

First, would you remember to write in your student number and name?

Student Number__________________ Signature____________________________

If any of the questions below are unclear to you, you may leave the answers blank.

1. Birthdate_____________

2. Number of Siblings (Please enter the numbers. If you are an only child, leave this blank.)
3.

Older brothers____ Older sisters____ Younger brothers____ Younger sisters_____

3. Blood Type (Circle one) A B O AB

4. If you have a certification, or have made extended study of a subject, write the level and years of study in the blank below. (For example, level 1 of *Eiken*, 2nd level of karate, 12 years of piano study, 3 years of hula dance, etc.)
   ____

5. Rate your ability to succeed in general (as in schoolwork, hobbies, training, work). (Rate yourself with one number, from a low of 0 to a high of 100.)

____

6. How many years of education did family members you live with, other than your siblings, complete? (Circle one for each member.)

Father	(9 12 (technical)14 (junior college)14 16 18 21)
Mother	(9 12 (technical)14 (junior college)14 16 18 21)
Grandfather	(9 12 (technical)14 (junior college)14 16 18 21)
Grandmother	(9 12 (technical)14 (junior college)14 16 18 21)
Other (______)	(9 12 (technical)14 (junior college)14 16 18 21)

7. Of those around you (such as parents, siblings, friends, teachers, etc.) what person influenced your study in school the most?

____

Thank you very much for your cooperation and help.