Research Report

Exploring *GRE*® and *TOEFL*® Score Profiles of International Students Intending to Pursue a Graduate Degree in the United States

ETS RR–22-02

Katrina Roohr
Margarita Olivera-Aguilar
Jennifer Bochenek
Vinetha Belur

December 2022

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

RESEARCH REPORT

Exploring *GRE*® and *TOEFL*® Score Profiles of International Students Intending to Pursue a Graduate Degree in the United States

Katrina Roohr, Margarita Olivera-Aguilar, Jennifer Bochenek, & Vinetha Belur

ETS, Princeton, NJ

The United States continues to be a top destination for international students pursuing an advanced degree. Some information about the characteristics of international students applying to graduate programs in the United States is available, but little is known about how these characteristics are related to test taker performance on graduate admissions tests and how performance may be related to graduate program characteristics. The purpose of this study was to investigate different patterns of performance of international test takers from four cultural regions and two large countries (China and India) on both the *GRE*® test and the *TOEFL*® test and the relationship with demographic and graduate program characteristics. Using finite mixture modeling, we investigated the most common score profiles using GRE and TOEFL for international students intending to pursue a graduate program within the United States; evaluated the demographic and college-level factors related to the profiles; and evaluated whether the profiles were differentially associated with gender, intended field of study, and intended degree level. Results showed the following broad patterns of results: (a) Most countries and cultural regions, except for the Middle East, had three or four latent profiles representing low, medium, and high scores on the GRE and TOEFL sections; (b) two high-performing profiles were found in Confucian Asia, one with higher GRE Quantitative Reasoning scores and the other with higher scores on GRE Verbal and TOEFL; (c) regardless of profile, test takers from China performed highest on the GRE Quantitative Reasoning section as compared to other GRE and TOEFL section scores; (d) in general, there was a relationship with students in the lower performing profiles taking the TOEFL and GRE multiple times; (e) regardless of country or cultural region, men were represented more than women overall and across most of the profiles; and (f) test takers showed a preference for science-, technology-, engineering-, and mathematics-based fields and master's degrees, but this varied across country and cultural region. Implications for future research are discussed.

Keywords International students; graduate school; *GRE*® test; *TOEFL*® test; finite mixture modeling

doi:10.1002/ets2.12343

The United States continues to be a top destination for international students pursuing an advanced degree of study; 24% of international students pursue education in the United States, followed by 11% in the United Kingdom (Institute of International Education [IIE], 2018b). However, trends in the top destinations for international students have started to change. Between fall 2017 and fall 2018, graduate enrollment of international students in the United States decreased by 4%. This decline was driven by a decline in master's degree enrollment, while the number of doctoral applications actually increased by 1% (Okahana & Zhou, 2019b). Despite this recent decrease, the 5- and 10-year annual rates of change in applications remained strong (Okahana & Zhou, 2019a).

To better understand the changes in graduate school enrollment by international students, it is important to understand the sociodemographic breakdown of students as well as the types of programs and degrees they intend to pursue. Looking more closely across various countries and cultural regions can help provide a better picture of who is interested in pursuing a graduate-level degree. This information could be useful in informing recruitment practices and increasing diversity across graduate programs. For this study, we were particularly interested in looking more closely at international students intending to complete their graduate studies at U.S. higher education institutions.

Who Is Coming to the United States to Pursue a Graduate Degree?

As of fall 2018, institutions had received more than 2.2 million applications for admission to graduate programs. Among those applicants, approximately 534,000 students were first-time enrollees in graduate programs, pursuing the range of

Corresponding author: K. Roohr, E-mail: kroohr@ets.org

graduate education from postbaccalaureate certificates to doctoral degree programs. Among those students, approximately 20% were international students, a 2% decrease from 2015, and 80% were U.S. citizens and permanent residents. Forty-five percent of the international students enrolling for the first time were women (Okahana & Zhou, 2019a).

Among international student applications, 71% were by students from China (47%) and India (27%). Approximately two-thirds of international students submitted applications for master's and certificate programs, and 34% submitted applications for doctoral degree programs. India had the largest percentage of students (83%) submitting to master's degree programs. Applications from Iran (81%), South Korea (70%), and Europe (60%) were mainly for doctoral programs (Okahana & Zhou, 2019b).

Among the 84,000 first-time international graduate student enrollees, most students were from Asian countries, including China (40%) and India (27%), followed by South Korea (3%), Taiwan (3%), and Japan (1%). Most Indian (89%) and Chinese (80%) students were in master's and certificate programs. Iran had the greatest percentage of students enrolled in doctoral programs at 73% (Okahana & Zhou, 2019b).

Regarding fields of study, research has shown that international students are more likely to enroll in science, technology, engineering, and mathematics (STEM)-based fields and that these numbers are higher than U.S. citizen enrollment in the same fields. For instance, among first-time students enrolled in mathematics and computer science, 55.6% were international students. For engineering, 51% were international students (Okahana & Zhou, 2019a). This trend has been consistent since 2012; Posselt (2016) reported that overall, approximately 55% of international students enrolled in STEM-based graduate programs, compared to 17% of U.S. students, in 2012. Stephan et al. (2015) also noted that the most popular fields for foreign students include engineering and physical sciences. Specific to master's degrees, international students applying to graduate programs in the United States tend to be concentrated in three fields of study: (a) engineering, (b) mathematics and computer science, and (c) business. For doctoral programs, after engineering, the most popular majors tend to be physical and earth sciences, mathematics and computer sciences, biological and agricultural sciences, and social and behavioral sciences (Okahana & Zhou, 2018).

Factors Related to International Students' Decisions to Study Abroad

Previous research has not only looked at the composition of international students at the aggregate level but has also looked at some of the reasoning behind why international students wish to pursue a graduate degree abroad. Specifically, research has focused on the combination of "push and pull factors" that encourage students to pursue a degree overseas (e.g., Chen, 2007). *Push factors* are those factors in the student's home country that initiate a student's choice to study internationally, and *pull factors* are those factors that attract a student to study in a particular host country (Mazzarol & Soutar, 2002). Examples of push factors include limited availability of financial aid, limited access to high-quality education, and fewer opportunities to improve English-language skills. Examples of pull factors include educational quality, institutional rankings, better employment prospects, and poststudy opportunities (Bista & Dagley, 2015).

International students, as compared to domestic students, have more factors they need to consider (e.g., academic, financial, social) before enrolling in their ideal institutions (Lei & Chuang, 2010). In particular, international students coming from a country where English is not the main spoken language may have additional challenges if they intend on studying in English-speaking countries like the United States or if they want to enroll in a graduate program where English is the main language of instruction. Specifically, they have to take an exam to prove their English-language proficiency, such as the *TOEFL*® test, before being accepted to a U.S. higher education institution. The TOEFL consists of reading, speaking, listening, and writing sections and is intended to measure the academic English-language proficiency needed to be successful in higher education (ETS, 2019c).

Like domestic students, international students typically must take required entrance exams (e.g., the *GRE*® test, the GMAT) and receive adequate scores to be accepted to graduate programs. The GRE General Test consists of verbal reasoning, quantitative reasoning, and analytical writing scores and is intended to measure skills needed for success in a graduate or professional (i.e., business or law) school program. These types of assessments, along with English proficiency assessments, have been identified as a challenge that international students face when applying to institutions within the United States. For instance, Bista and Dagley (2015) interviewed international students at a small university in the southern United States and found that, following U.S. visa preparation, college entrance exams were identified as the biggest challenge by 48% of the surveyed students. These results were echoed in follow-up interviews with students who indicated that in some cases they searched for programs or institutions that did not require these exams for admission.

Although entrance exams are considered a challenge for international students, few studies have evaluated how the performance on these assessments could be related to the intended program or field of study. As a result, we decided to examine students' scores on the GRE and TOEFL by classifying students into various profiles based on their patterns of scores and to look at how those patterns relate to characteristics of graduate programs to which they opt to apply. For example, it could be that students with lower scores in particular subareas (e.g., GRE Verbal) tend to select non-STEM fields.

Classification of Students Into Profiles

The idea of classifying students into groups is not new. Some studies have classified students based on their motivational or noncognitive characteristics (Olivera-Aguilar et al., 2017; Pastor et al., 2007; Pastor & Barron, 2012), their previous academic performance measured with their grade point average (GPA; Sulak et al., 2017), or behavioral data on online courses (i.e., logging-in system, watching lecture videos, submitting assignments, and posting on discussion forums; Tseng et al., 2016). Other studies have included test scores as indicator variables to help define the profiles of students (Mattern et al., 2012; Schmitt et al., 2007). The advantage of classifying students into groups is that it becomes possible to evaluate whether the relationships between variables of interest vary by group. Although the studies by Mattern et al. (2012) and Schmitt et al. (2007) included test scores to define the clusters, these studies used standard cluster analysis, which has several disadvantages in comparison to finite mixture models.

Finite mixture modeling (FMM) assumes that the population consists of a mixture of unobserved groups, and the purpose of the analysis is to uncover the number and nature of such unobserved groups, also known as classes or profiles. An adequate solution consists of a sufficient number of profiles that reveal both separation (i.e., a distinctive pattern of responses) between the different profiles and relatively homogeneous responses within each profile (Collins & Lanza, 2010; Vermunt & Magidson, 2002). One advantage of FMM is that, because it is model based, several fit statistics can be used to evaluate the fit of the model; as a consequence, the selection of the best-fitting model is less arbitrary than in standard cluster analysis (Pastor et al., 2007; Vermunt & Magidson, 2002). Another advantage is that FMM facilitates the classification of individuals from a new sample to the profiles that were previously determined using an original sample from the same population. That is, when data on the FMM indicators are collected in a second sample, those individuals can be classified into latent profiles using the original model parameter estimates (Pastor et al., 2007).

It is usually of interest to examine the relationship between groups of individuals and external variables, such as covariates or outcomes. To examine these relationships in standard cluster analysis, individuals are classified into a single cluster, and external variables of interest are examined post hoc via techniques like analysis of variance (ANOVA). This approach assumes that each individual belongs to only one cluster, which increases the chances of misclassification and estimation error. In contrast, FMM computes the probability that a student belongs to each identified profile, and hence the classification of individuals into profiles is not deterministic. In FMM, external variables (such as covariates or outcomes) can be directly included in the model, accounting for any uncertainty in individual classification (Clark & Muthén, 2009; Pastor et al., 2007; Vermunt & Magidson, 2002).

For this study, we chose covariates that have been traditionally associated with academic performance, that is, mother's level of education, father's level of education, undergraduate overall GPA, and undergraduate major GPA (Robbins et al., 2004). Considering that students can take the TOEFL and GRE multiple times and that this may have a relationship with their scores, we also included the number of times applicants took the GRE and the TOEFL in the last 5 years. Finally, we included age as a covariate given the variability in the age of graduate students in the United States, where 47% of graduate students are aged 30 years or older (National Center for Education Statistics, 2019).

Study Objectives and Research Questions

Given that approximately one in five first-time graduate students in the United States is an international student, it is critical to evaluate this population and better understand the demographic composition of students coming from various countries as well as the types of programs and degrees they are interested in pursuing. Previous research has identified numerous push and pull factors that influence international student choice on where to apply to and attend graduate school and challenges students may have when applying (e.g., Chen, 2007; Mazzarol & Soutar, 2001). One challenge in applying to graduate school that has been identified, but that is understudied, is the influence of admissions tests on

application decisions (e.g., Bista & Dagley, 2015). Although this study does not look specifically at students' decision-making, this research provides the necessary first step in understanding test scores as a factor in international students' decisions when applying to graduate school. We sought to look at profiles of students based on their graduate admissions test performance and the ways this is associated with the pursuit of various fields of study and degree type. We were also interested in examining whether gender representation varied by profile, as some gender differences have been found in GRE and TOEFL scores (ETS, 2019a, 2019e). Specifically, the purpose of this study was to investigate different patterns of international test takers' performance on both the GRE and TOEFL and to evaluate the relationship between performance and graduate program characteristics. We examined patterns in students' profiles using a person-centered approach (i.e., FMM) instead of a variable-centered approach, as this allowed us a more holistic view of the heterogeneity in students' score patterns. We investigated two main research questions (RQs):

RQ1: What are the most common score profiles using GRE and TOEFL test scores for international students from various countries or cultural regions intending to pursue a graduate school program within the United States?

RQ2: Are the profiles differentially associated with gender, intended field of study, or intended degree level?

The goal of this study was not to look at the relationships between GRE and TOEFL scores but rather to investigate how these two sets of scores work together to characterize student profiles across diverse countries and cultural regions. The assessments were developed for different purposes and target different skills and knowledge. For instance, while the GRE General Test measures the verbal reasoning, quantitative reasoning, and analytical writing skills needed for success in a graduate or professional (i.e., business or law) school program, the TOEFL measures the academic English-language proficiency needed to succeed in higher education. Both testing programs provide guidance on the appropriate use of test scores (see ETS, 2019b, 2019d). Haberman and Yao (2015) explored the extent to which these assessments complement one another and to which they are distinguishable using an augmentation approach with test repeaters. Their results showed that students' results on these assessments are complementary as opposed to redundant, suggesting that each assessment is providing some unique information regarding student performance across the various constructs. Specifically, results show that TOEFL Reading and Writing and GRE Analytical and Verbal were complementary, which was expected, as the tests involve reading and writing skills. The GRE Quantitative section was not redundant, because that section was not measured on TOEFL and augmentation had a limited effect. Other results show that TOEFL Speaking and Listening were somewhat affected by augmentation of other TOEFL sections but that the addition of GRE section scores resulted in little gain. These results ultimately indicate that although some constructs are closely linked, others are much less strongly related.

Method

Sample

For the purpose of this study, international students were defined as students who are neither U.S. citizens nor resident aliens, whose native language is not English, and who took the GRE outside of the United States between 2012 and 2017. Furthermore, the sample was restricted to those GRE test takers who also took the TOEFL in the same period of time. Prior to matching students who took the GRE and TOEFL, we removed duplicates and students who were missing any of the scores. When matching the TOEFL data to the GRE data file, 80% of the sample was matched based on full name, date of birth, and gender. The additional 20% was matched on email address, date of birth, and gender. This matching process ended up in a sample of 329,471 students. Students were also matched on their country of citizenship as reported on the GRE and the country from their address reported on the TOEFL background questionnaire. Additionally, we excluded from the analysis test takers who reported living 1 or more years in an English-speaking country and TOEFL test takers who reported taking the TOEFL for reasons other than attending a graduate program or who reported being interested in attending graduate school in countries other than the United States. These exclusions reduced the sample size to 195,985 test takers.

We focused on countries that had the largest number of GRE test takers from 2015 to 2016 (ETS, 2016). Students were classified into countries and cultural regions[1]; countries with fewer than 100 test takers and cultural regions with fewer

Table 1 Sample Size per Country and Cultural Region

Cultural region	Country	Country N	Total N
Southern Asia	India	108,371	108,371
Confucian Asia	China	41,805	41,805
Southern Asia	Iran	4,270	11,258
	Bangladesh	2,791	
	Pakistan	1,908	
	Nepal	1,781	
	Sri Lanka	508	
Confucian Asia	Taiwan	4,341	9,304
	Korea	3,649	
	Japan	861	
	Hong Kong	453	
Latin America	Brazil	2,787	8,444
	Mexico	1,873	
	Colombia	1,151	
	Chile	695	
	Ecuador	427	
	Peru	407	
	Venezuela	402	
	Argentina	361	
	Costa Rica	181	
	Dominican Republic	160	
Middle East	Turkey	2,336	4,839
	Saudi Arabia	804	
	Egypt	711	
	Lebanon	382	
	Jordan	333	
	Kuwait	127	
	Afghanistan	146	

than 4,000 test takers were not considered in the analysis to ensure adequate sample sizes for cross-validation. This reduced the sample by an additional 11,964 test takers, rendering a final sample consisting of 184,021 test takers distributed across 28 countries classified into 4 cultural regions (Table 1). Note that we decided to keep analyses for India and China separate given that their large sample sizes could overshadow the contribution of other countries in their cultural regions to our statistical models.

Statistical Procedures

FMM was used to select the number of underlying profiles and to examine the relationships between the profiles and external variables. For RQ1, we compared models with varying numbers of profiles and selected the best-fitting model per country and cultural region. To obtain more information about the profiles selected and to facilitate interpretation, we examined the relationship between the profiles and a set of covariates obtained from the GRE and TOEFL background information questionnaire that students filled out when registering for the assessment. For RQ2, we examined if the proportion of students in each profile differs in terms of gender, program type (e.g., STEM vs. non-STEM fields), and degree level (master's vs. doctoral).

Number of Profiles

For RQ1, the three GRE test scores (Verbal Reasoning [Verbal], Quantitative Reasoning [Quantitative], and Analytical Writing) and the four TOEFL scores (Reading, Listening, Speaking, and Writing) were standardized in the overall sample and used as indicators. For students who completed the GRE or TOEFL multiple times, we used their last scores for the purposes of analyses. We compared the fit of models, extracting from three to seven profiles. In each of these models, we also compared models with four different covariance structures (Bauer & Curran, 2004; Pastor et al., 2007). In Model A, the indicator variances were constrained to equality across classes, and the covariances among indicators were fixed at a zero value; this model represents the latent profile analysis (LPA) model that makes the assumption of local independence

(Bauer & Curran, 2004). In Model B, the indicator variances were freely estimated across profiles, and the covariances among indicators were fixed at zero. In Model C, the indicator variances were freely estimated across classes, and the indicator covariances were constrained to equality across classes. Finally, in Model D, the indicator variances and covariances were freely estimated across classes. To minimize the possibility of arriving at local maxima solutions, the models were estimated with 1,500 random sets of start values and 750 final-stage optimizations.

Model selection remains an important research topic (Marsh et al., 2009), and no single statistic has been shown to be preferable across every condition (Nylund et al., 2007; Tein et al., 2013). For this reason, we considered several sources of information when selecting the final model. First, we considered model fit indices, such as the Bayesian information criterion (BIC; Schwarz, 1978) and the adjusted Lo–Mendell–Rubin likelihood ratio test (aLMR; Lo et al., 2001). Both indices indicate relative fit of a model in comparison to other models; lower BIC values are evidence of better fit. The aLMR indicates the relative fit of a model with P profiles in comparison to a model with $P - 1$ profiles; a p-value greater than .05 indicates a better fit by the model with $P - 1$ profiles (Tofighi & Enders, 2008). Other sources of information considered in model selection were the number of replications where the model reached convergence, profile size, the average class assignment probabilities, and the interpretability of the profiles. Although there is no standard cutoff for the minimum percentage of the sample that should be assigned to a profile, we did not consider solutions in which one or more profiles consisted of less than 5% of the sample. Within each profile, the average class assignment probabilities for all individuals can be interpreted as the reliability of classifying students into that profile (Geiser et al., 2006). Reliability values of .90 are considered excellent, whereas values of .70 are considered adequate (Kline, 2012). Hence solutions in which the class assignment probabilities were less than .70 were not considered. Finally, the interpretability and uniqueness of the profiles were also considered when selecting the final model to retain (Pastor et al., 2007).

To examine whether the profiles were generalizable to other samples, we conducted cross-validation analyses. The sample for each country and cultural region was divided into two random samples; in the test sample, we compared several models to determine the total number of latent profiles; the final model was fit to the validation sample. We considered the overall characteristics of the obtained profiles to evaluate the extent to which the qualitative nature of the profiles was replicated in the validation sample. We also considered the number of starting values in the validation sample with convergence problems. In the cases in which the nature or the interpretation of the profiles was different in the validation sample or when convergence problems were found in half or more of the starting values in the validation sample, the second-best-fitting model was examined.

Analysis of Covariates

Once the number of latent profiles was determined in each country or cultural region and the results were replicated in the validation sample, the relationships between the profiles in the testing sample and a set of covariates were analyzed. This analysis helped to support the interpretation of the profiles as part of RQ1. The covariates were exogenous variables not used as indicators of the latent profiles. The purpose of this analysis was to provide more contextual information for the interpretation of the profiles in the selected model for each country or cultural region; hence, this analysis was only conducted once the final models were identified. Covariates included in this study were mother's and father's levels of education, undergraduate overall and major GPA, the number of times applicants took the GRE and TOEFL in the last 5 years, and age in years. As previously mentioned, these covariates were selected because they were associated with academic performance and because some of these variables could be related to performance on these measures.

The statistical effect of the covariates was estimated concurrently with the best-fitting model in each country and cultural region via the R3STEP option in *Mplus* 7.4 (Asparouhov & Muthén, 2014; Vermunt, 2010), in which the covariates are used to predict membership in the profiles via a multinomial logistic regression. In FMM, the classification of individuals into profiles is not deterministic; instead, probabilities of every student belonging to each profile are computed. Directly including the covariates in the model accounts for any inaccuracy of individual classification and provides unbiased estimates of regression coefficients (Clark & Muthén, 2009; Pastor et al., 2007; Vermunt & Magidson, 2002).

Profiles as Independent Variables

To address RQ2, we examined the extent to which the profiles were differentially associated with gender, the intended field of study (e.g., STEM vs. non-STEM), and eventual graduate degree objective (master's vs. doctoral). We used the

method proposed by Lanza et al. (2013) as implemented in *Mplus* 7.4 because it produces less biased estimates than the standard classify–analyze approach. In this procedure, the latent profiles were considered the independent variables, and the dependent variables were the estimated proportions of gender, field of study, and degree objective by profile. Differences in these proportions by profile were evaluated with the overall test of association via Wald's test and pairwise profile comparisons (Asparouhov & Muthén, 2014). To have a better understanding of the magnitude of the proportion differences by profile, we computed Cohen's *h*. However, one word of caution is that the effect sizes were computed using the formulas developed for observed groups; in contrast, in FMM, every individual has a likelihood of membership in all profiles. It may be possible that the formulas developed for observed groups do not translate to latent grouping. For this reason, we used the effect sizes in addition to the significance testing to understand overall trends rather than exact values.

Results

Number of Profiles

The model selected for China resulted in several starting values with no convergence in the validation sample. Additionally, the results for India in the validation sample showed a pattern with a qualitatively different interpretation from the testing sample. In these cases, the second-best-fitting models were selected in the testing sample and examined in the validation sample. The final models chosen with their model fit information for the testing and validation samples are shown in Table 2.[2] In most models, the aLMR resulted in significant *p*-values, even when the models with more profiles showed convergence problems or had average profile probabilities smaller than .70; hence it was not useful in identifying the best-fitting models. We relied on the rest of the criteria to select the final models. Figures 1–6 show the profiles obtained for each country and cultural region.

In most countries, the final model consisted of three or four latent profiles; all the profile assignment probabilities were above .70. The profiles obtained represented similar shapes, with only level differences. The results indicate that in China (Figure 1) and Southern Asia (Figure 2), there were three profiles representing low, medium, and high scores on GRE and TOEFL. Although the model with three profiles was the best-fitting model across these regions, the profiles have different patterns and are not directly comparable across regions. For example, whereas students in China had the higher scores on GRE Quantitative in every profile, GRE Quantitative scores were the lowest in most profiles in Southern Asia.

The results for India (Figure 3) and Latin America (Figure 4) indicate the presence of four latent profiles. In India, Profiles 1–3 had relatively homogeneous scores in each GRE and TOEFL section, but students in Profile 4 had much lower scores in TOEFL Reading, Listening, and Writing in comparison to their GRE scores and much higher scores in

Table 2 Fit for Final Model in Testing and Validation Sample

Country/ cultural region	Model	Testing sample				Validation sample			
		AIC	BIC	aBIC	aLMR	AIC	BIC	aBIC	aLMR
India	Four profiles, free variance, covariance constrained to equality	785,187.9	785,900.0	785,645.7	6,877.137***	785,530.6	786,242.6	785,988.4	6,910.6***
China	Three profiles, free variance, zero covariance structure	268,071.9	268,421.6	268,281.8	15,766.1***	267,585.2	267,934.9	267,795.1	15,961.4***
Southern Asia	Three profiles, free variance, free covariance	84,136.3	84,846.3	84,506.3	1,479.7***	84,599.0	85,309.0	84,969.0	1,532.3***
Confucian Asia	Four profiles, free variance, covariance constrained to equality	66,545.1	67,060.7	66,806.5	632.6	66,363.9	66,879.5	66,625.3	756.2***
Latin America	Four profiles, free variance, free covariance	60,911.4	61,819.2	61,364.8	553.0***	61,294.7	62,202.5	61,748.1	598.4***
Middle East	Five profiles, free variance, covariance constrained to equality	38,744.8	39,294.9	38,993.1	200.0***	38,648.9	39,199.1	38,897.2	202.5

Note. AIC = Akaike information criterion; BIC = Bayesian information criterion; aBIC = adjusted Bayesian information criterion; aLMR = adjusted Lo–Mendell–Rubin likelihood ratio. ***p < .001.
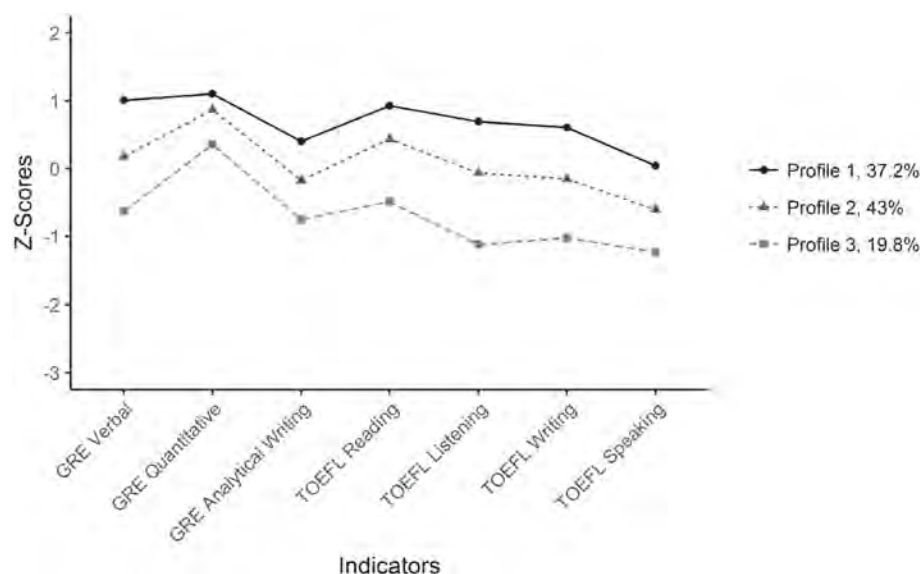
Figure 1 China latent profiles.
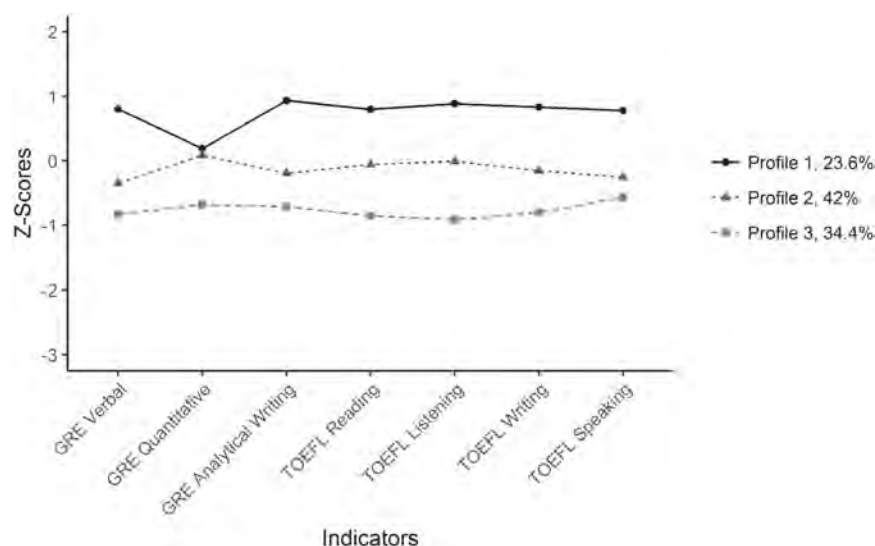
Figure 2 Southern Asia latent profiles.

TOEFL Speaking. For Latin America, the GRE Quantitative scores in each profile were lower than the scores in other sections.

For the Middle East, results show five profiles for test takers (Figure 5). Similar to other regions, four of the five profiles were defined by test scores between 1 standard deviation above the mean and 1 standard deviation below the mean. However, one difference with the profiles in other regions was that the lowest performing profile in the Middle East was defined by test scores between 2 and 3 standard deviations below the mean. Another interesting finding for the Middle East was that while three of the profiles only showed level differences corresponding to low (Profile 5), medium (Profile 4), and high (Profile 1) scores on GRE and TOEFL, the other two profiles showed shape differences. Profiles 2 and 3 showed similar scores on GRE Verbal, GRE Analytical Writing, TOEFL Reading, and TOEFL Writing but differed in the scores on GRE Quantitative and TOEFL Speaking. In comparison to Profile 3, Profile 2 showed high scores on GRE Quantitative and lower scores on TOEFL Speaking; it should be noted that the scores on GRE Quantitative were as high as those in Profile 1. The opposite pattern was found for Profile 3, with GRE Quantitative scores as low as those in Profile 4 and TOEFL Speaking scores as high as those in Profile 1.
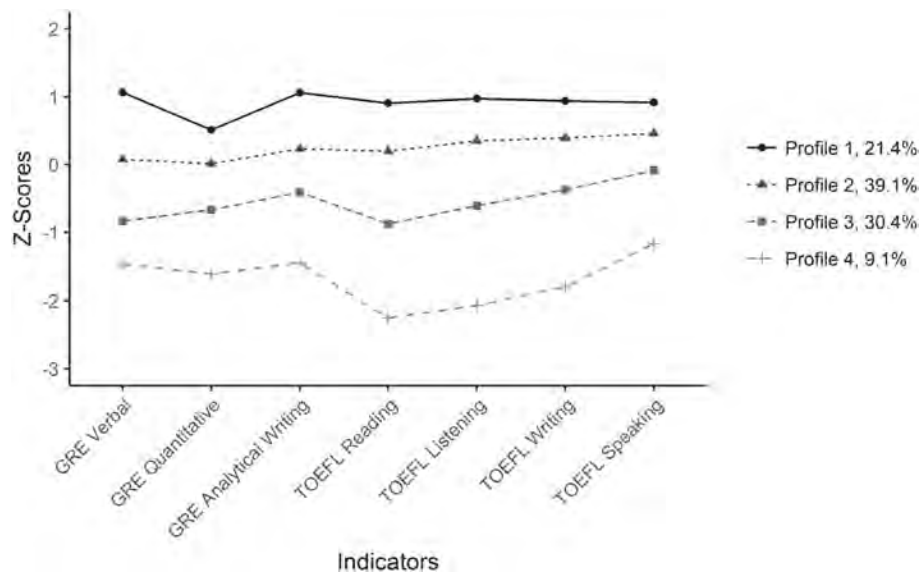
Figure 3 India latent profiles.
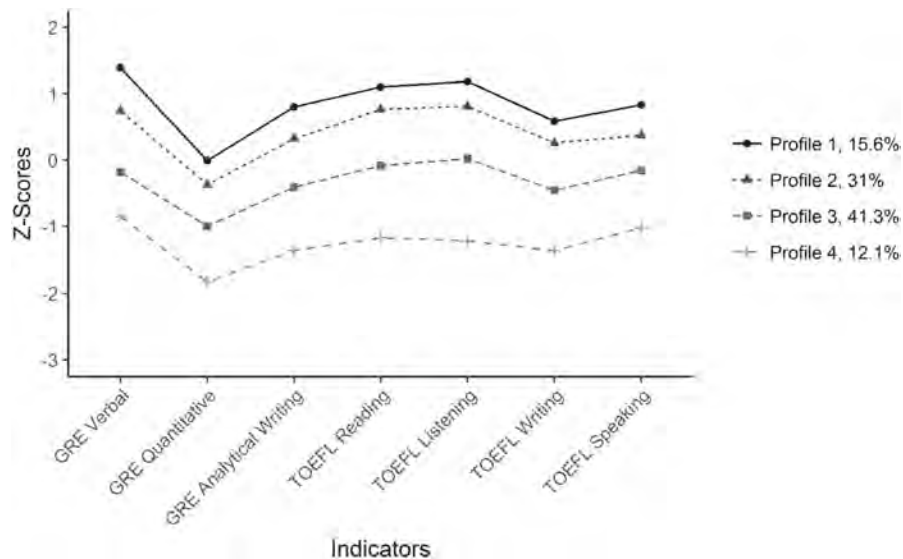
Figure 4 Latin America latent profiles.

Interesting patterns were also detected in Confucian Asia (Figure 6). While three of the four profiles found corresponded to low (Profile 4), medium (Profile 3), and high scores (Profile 2), an additional profile (Profile 1) showed a distinct shape. Students in Profile 1 had the highest scores in GRE Verbal, GRE Analytical Writing, and the four sections of TOEFL, and their GRE Quantitative scores were as low as for students in the medium profile (Profile 3). The profile sizes for every country/region are shown in Table 3. No profile size was smaller than 5%, which was our criterion for the minimum profile size.

Analysis of Covariates

The results in Table 4 indicate the effect of the covariates (measured in logits) in predicting profile membership to the lower performing profiles in comparison to the highest performing profile (Profile 1). Because of the large number of comparisons being conducted, although we indicate significant results at *p*-values of .05, .01 and .001 in Table 4, we only interpret results that were significant at a .01 level.
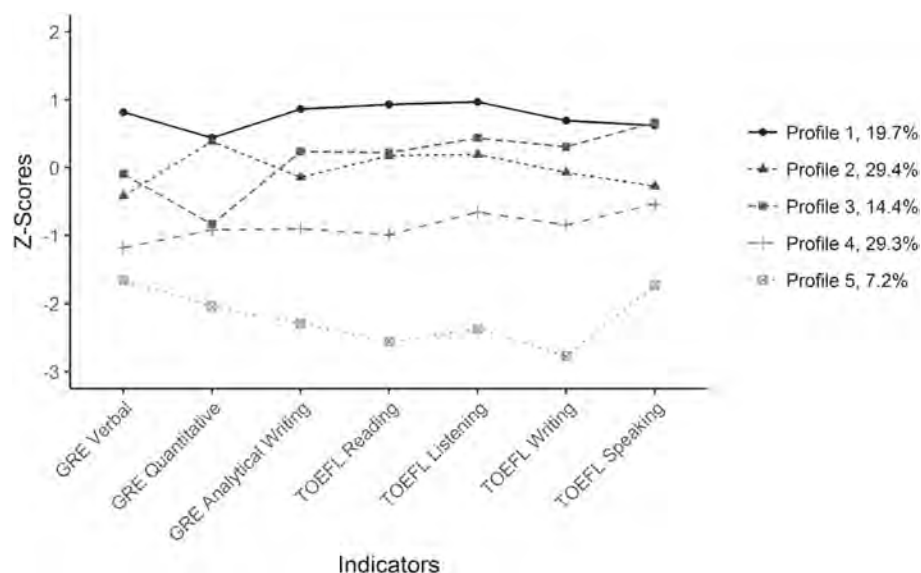
Figure 5 Middle East latent profiles.
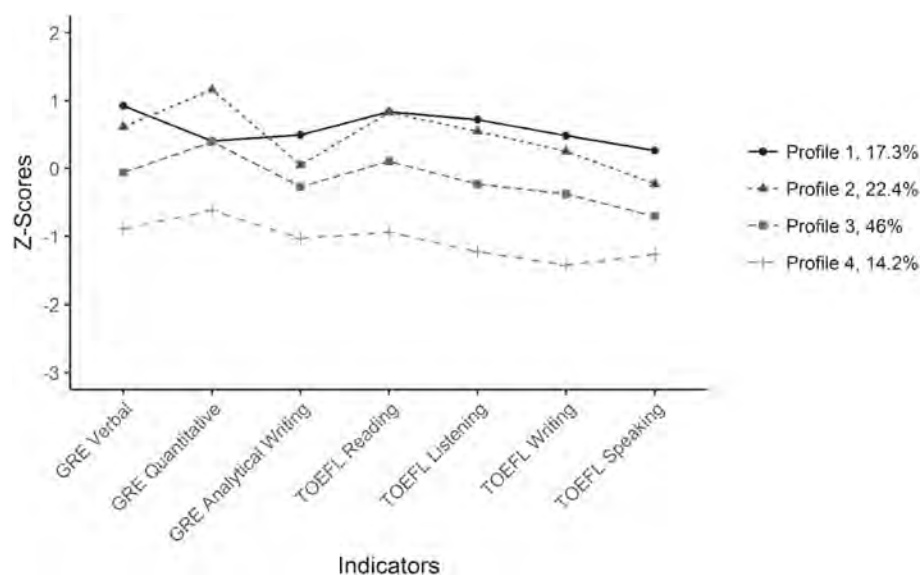
Figure 6 Confucian Asia latent profiles.

Table 3 Profile Size in Final Models in Testing Sample

	Profile 1		Profile 2		Profile 3		Profile 4		Profile 5	
Country/cultural region	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
India	11,605	21.4	21,213	39.1	16,456	30.4	4,912	9.1		
China	7,779	37.2	8,992	43	4,132	19.8				
Southern Asia	1,331	23.6	2,362	42	1,936	34.4				
Confucian Asia	806	17.3	1,044	22.4	2,139	46	663	14.2		
Latin America	659	15.6	1,307	31	1,746	41.3	510	12.1		
Middle East	476	19.7	712	29.4	349	14.4	708	29.3	175	7.2

Table 4 Multinomial Logistic Regression Coefficients for Predicting Membership to Profiles in Comparison to the High-Performing Profile

Covariate	India[a]			China[b]		Southern Asia[c]		Confucian Asia[d]			Latin America[e]			Middle East[f]			
	Profile 2	Profile 3	Profile 4	Profile 2	Profile 3	Profile 2	Profile 3	Profile 2	Profile 3	Profile 4	Profile 2	Profile 3	Profile 4	Profile 2	Profile 3	Profile 4	Profile 5
# GRE	0.42***	0.48***	0.29***	0.004	−0.12	0.08	−0.14	−0.21*	−0.28**	−0.69***	−0.11	−0.35***	−0.02***	−0.24	−0.23	−0.41*	−0.39
# TOEFL	0.21***	1.10***	1.17***	0.38***	0.47***	0.83***	0.91***	0.07	0.27***	0.30***	0.07	0.30***	0.42***	0.30*	0.24	0.74***	0.44
MotherEd	−0.12***	−0.25***	−0.36***	−0.06*	−0.08*	−0.12***	−0.16***	−0.022	−0.11*	0.02	−0.06	−0.12***	−0.21***	−0.11*	−0.03	−0.17**	−0.22
FatherEd	−0.09***	−0.14***	−0.23***	−0.06*	−0.10***	−0.08*	−0.08*	−0.04	−0.05	−0.17*	−0.01	−0.10*	−0.22***	−0.04	−0.01	−0.04	−0.2
MajorGPA	−0.10*	−0.26***	−0.29***	−0.13*	−0.25***	0.13	−0.07	−0.22	−0.36***	−0.40**	0.07	−0.04	−0.12	−0.11	0.02	0.07	−0.18
UGPA	−0.04	0.03	−0.03	−0.28***	−0.38***	−0.2*	−0.24*	0.21	−0.18	−0.32*	−0.32*	−0.37***	−0.21	−0.11	−0.28	−0.48*	−0.22
Age	0.08***	0.01	−0.03	0.35***	0.66***	0.12	0.48***	−0.44***	−0.02	0.25**	0.14*	0.24***	0.34***	0.11	0.78***	0.66***	0.91

Note. # GRE = number of times taking GRE; # TOEFL = number of times taking TOEFL; MotherEd = mother's level of education; FatherEd = father's level of education; MajorGPA = grade point average within major; UGPA = undergraduate grade point average. *p < .05. **p < .01. ***p < .001. [a]N = 32,673. [b]N = 8,780. [c]N = 2,949. [d]N = 3,480. [e]N = 3,088. [f]N = 1,235.

Results show that in general, test takers in lower performing profiles have taken the TOEFL significantly more times than test takers in the highest performing profile. The same pattern was observed for the number of times the GRE was taken in India. In Latin America and for two profiles in Confucian Asia, however, we found the opposite pattern such that test takers in the lower performing profiles tended to have taken the GRE significantly fewer times than test takers in the higher performing profiles, at $p < .01$.

In general, we found that a lower mother's level of education tended to be associated with membership in lower performing profiles. Similarly, lower father's level of education was associated with membership in lower performing profiles, but this relationship was not as generalized as the relationship between profile membership and mother's level of education. In some regions, lower overall undergraduate GPA and major GPA was associated with membership in lower performing profiles. Also, test takers in lower performing profiles tended to be older than test takers in the highest performing profile.

Given that Confucian Asia and the Middle East showed profiles with distinctive shapes, we focused on results from some of the profiles in these regions. In Confucian Asia, we focused on Profile 1, characterized by the highest scores in every indicator except for GRE Quantitative, and Profile 2, which had the highest GRE Quantitative scores. The covariates analysis indicated that, in comparison to Profile 1, test takers in Profile 2 were younger and took the GRE fewer times.

In the Middle East, we focused on Profile 1, which had the highest scores, except for GRE Quantitative and TOEFL Speaking; on Profile 2, which had GRE Quantitative scores as high as Profile 1 but lower scores than Profile 3 on the rest of the scales; and on Profile 3, with the second highest scores, except for GRE Quantitative. The analysis of covariates revealed that test takers classified in Profile 2 took the TOEFL more times and had lower mother's education than test takers in Profile 1, whereas test takers in Profile 3 were older than test takers in Profile 1.

It should be noted that the R3STEP automated procedure in *Mplus* uses listwise deletion; as a consequence, the analysis was conducted with a smaller sample size. Because of the large proportion of cases that were excluded from the analysis (between .25 and .58), we examined whether the exclusion of missing data introduced bias in the results by examining mean differences in the GRE and TOEFL test sections between test takers with and without missing data. The effect sizes as measured by Cohen's *d* were close to zero for most countries and cultural regions. Only the Middle East showed small effect sizes (Cohen's *d* between .16 and .30). Because of the small effect sizes, we decided that the bias introduced by excluding missing data from the analysis would be minimal.

Gender, Field of Study, and Degree Level Differences by Profiles

Gender

In terms of overall frequencies, there were larger proportions of men as compared to women across all countries and cultural regions, ranging from 55% in China to 69% in Southern Asia (Table 5). The estimated proportions of women and men in each profile and cultural region are shown in Table 5; the statistical tests comparing proportions between profiles are shown in Table 6. The results indicate that overall, there were significant differences by gender across profiles in India, China, and the four cultural regions (Table 6); however, the effect sizes in India and China were close to zero (Table 7).

Results in Confucian Asia show a larger proportion of men than women in every profile, except in Profile 1 (with the highest scores on GRE Verbal, GRE Academic Writing, and every TOEFL section), where 70% of the test takers were women. Similarly, in the Middle East, men were more prevalent than women in every profile, except for Profile 3, in which women represented 67% of the test takers. This profile had the highest scores on the TOEFL Speaking section and below average GRE Quantitative scores. In Latin America, the proportion of women was similar in Profiles 1, 2, and 3 but much larger in Profile 4 (small and medium effect sizes).

STEM Versus Non-STEM

In contrast to the comparisons by gender where there was a negligible amount of missing data (no more than three cases with missing data), the proportion of test takers with missing data when indicating their interest in STEM or non-STEM fields was between 18% and 24% (Table 8). For this reason, we included missing cases as another category to compare. It should also be noted that test takers also indicated their interest in business degrees, but because of the small proportion of these cases (between 4% and 7%), we did not include them in the analysis.

Table 5 Estimated Proportion of Women and Men in Each Profile per Country/Cultural Region

Gender by profile	India[a]	China[b]	Southern Asia[c]	Confucian Asia[d]	Latin America[e]	Middle East[f]
Full sample						
Women	0.32	0.45	0.31	0.39	0.37	0.34
Men	0.68	0.55	0.69	0.61	0.63	0.66
Profile 1						
Women	0.29	0.48	0.33	0.70	0.29	0.23
Men	0.71	0.52	0.67	0.30	0.71	0.77
Profile 2						
Women	0.32	0.42	0.23	0.20	0.32	0.18
Men	0.68	0.58	0.77	0.80	0.68	0.82
Profile 3						
Women	0.34	0.43	0.37	0.29	0.37	0.67
Men	0.66	0.57	0.62	0.71	0.62	0.33
Profile 4						
Women	0.29			0.49	0.54	0.36
Men	0.71			0.51	0.46	0.63
Profile 5						
Women						0.42
Men						0.58

[a] $N = 54{,}183$. [b] $N = 20{,}903$. [c] $N = 5{,}628$. [d] $N = 4{,}650$. [e] $N = 4{,}222$. [f] $N = 2{,}417$.

Table 6 Approximate Chi-Square Values for Proportion Comparison for Gender

Profile comparisons	India	China	Southern Asia	Confucian Asia	Latin America	Middle East
Overall test	51.07***	47.49***	35.29***	230.74***	57.86***	124.91***
1 versus 2	8.48**	37.73***	16.88***	179.17***	0.53	1.26
1 versus 3	41.17***	26.88***	3.46	57.12***	2.46	63.05***
1 versus 4	0.00			21.43***	20.17***	6.03*
2 versus 3	9.69**	0.15	33.34***	5.44*	0.61	97.98***
2 versus 4	8.91**			82.67***	26.58***	28.61***
3 versus 4	26.11***			20.00***	3.01	32.43***
1 versus 5						11.02**
2 versus 5						26.38***
3 versus 5						17.83***
4 versus 5						1.26

*$p < .05$. **$p < .01$. ***$p < .001$.

Table 7 Effect Sizes (Cohen's *h*) for Differences in Gender Proportions

Profile comparisons	India	China	Southern Asia	Confucian Asia	Latin America	Middle East
1 versus 2	0.05	−0.12	−0.22	1.06	0.06	−0.13
1 versus 3	0.10	0.11	0.09	0.84	0.18	0.92
1 versus 4	0.00			−0.43	0.51	0.29
2 versus 3	0.05	−0.01	−0.32	0.22	0.12	0.41
2 versus 4	0.05			0.62	0.45	−1.05
3 versus 4	0.10			0.41	0.34	0.42
1 versus 5						0.54
2 versus 5						−0.63
3 versus 5						−0.52
4 versus 5						0.12

Table 8 Estimated Proportion of Test Takers Pursuing STEM or Non-STEM Degrees in Each Profile per Country/Cultural Region

Field by profile	India	China	Southern Asia	Confucian Asia	Latin America	Middle East
Full sample						
Missing	0.23	0.24	0.18	0.18	0.21	0.21
Business	0.06	0.08	0.04	0.05	0.07	0.07
STEM	0.69	0.58	0.71	0.56	0.50	0.57
Non-STEM	0.01	0.11	0.07	0.21	0.22	0.16
Profile 1						
Missing	0.20	0.23	0.20	0.28	0.15	0.15
STEM	0.77	0.61	0.64	0.09	0.51	0.68
Non-STEM	0.04	0.17	0.16	0.63	0.34	0.17
Profile 2						
Missing	0.26	0.25	0.12	0.07	0.23	0.14
STEM	0.73	0.67	0.87	0.82	0.43	0.79
Non-STEM	0.01	0.08	0.01	0.11	0.34	0.08
Profile 3						
Missing	0.26	0.34	0.26	0.15	0.15	0.39
STEM	0.74	0.57	0.65	0.75	0.77	0.08
Non-STEM	0.01	0.09	0.09	0.10	0.08	0.53
Profile 4						
Missing	0.27			0.38	0.42	0.26
STEM	0.73			0.36	0.30	0.60
Non-STEM	0.00			0.26	0.28	0.14
Profile 5						
Missing						0.40
STEM						0.45
Non-STEM						0.15
Total sample used in analysis	50,778	19,335	5,394	4,398	3,943	2,261

Note. STEM = science, technology, engineering, mathematics.

Table 9 Approximate Chi-Square Values for Proportion Comparison by Field of Study

Profile comparisons	India	China	Southern Asia	Confucian Asia	Latin America	Middle East
Overall test	381.57***	276.25***	213.02***	1,004.99***	204.02***	164.76***
1 versus 2	184.89***	167.68***	124.43***	812.67***	6.26*	5.73
1 versus 3	292.22***	191.64***	24.38***	314.33***	89.41***	48.48***
1 versus 4	256.39***			92.89***	72.65***	11.80**
2 versus 3	28.65***	66.66***	97.39***	18.26***	102.97***	106.58***
2 versus 4	36.62***			111.04***	40.34***	30.33***
3 versus 4	3.01			96.74***	97.15***	46.74***
1 versus 5						23.64***
2 versus 5						45.01***
3 versus 5						24.88***
4 versus 5						8.92*

*$p < .05$. **$p < .01$. ***$p < .001$.

In general, test takers across all the countries and cultural regions had a preference for STEM-based fields, with overall proportions ranging from 50% in Latin America to 70% in India (Table 8). The results for the estimated proportions across profiles indicate that the majority of test takers (at least 57%) in each profile showed a clear preference for STEM fields in India, China, and Southern Asia (Table 8). Although the approximate chi-square results indicate significant differences in the proportion of test takers preferring STEM and non-STEM fields by profiles in India and China (Table 9), Cohen's *h* indicated only small effect sizes (Table 10).

In Southern Asia, the highest performing profile did not have the highest proportion of test takers interested in STEM fields. It was Profile 2, with test scores close to the mean, with the largest proportion of test takers interested in STEM (87%). Medium effect sizes were found when comparing Profile 2 with Profiles 1 and 3, whereas only small differences were found when comparing Profiles 1 and 3.

Table 10 Effect Sizes (Cohen's *h*) for the Comparison of Field of Study

Profile comparisons	India	China	Southern Asia	Confucian Asia	Latin America	Middle East
Missing data
1 versus 2	0.14	−0.06	−0.23	0.58	0.19	−0.04
1 versus 3	0.15	0.24	0.14	0.34	−0.02	0.55
1 versus 4	0.18			0.20	0.61	0.28
2 versus 3	0.00	0.18	−0.37	0.25	−0.21	−0.59
2 versus 4	−0.03			0.78	0.42	0.32
3 versus 4	−0.03			0.54	0.63	−0.27
1 versus 5						0.58
2 versus 5						0.62
3 versus 5						0.02
4 versus 5						0.29
STEM
1 versus 2	−0.09	−0.12	0.53	−1.67	−0.15	0.23
1 versus 3	−0.07	−0.07	0.01	−1.50	0.57	−1.37
1 versus 4	−0.10			0.69	−0.42	−0.18
2 versus 3	0.01	−0.20	0.52	−0.17	0.71	1.60
2 versus 4	0.02			−0.98	−0.28	−0.42
3 versus 4	0.03			−0.81	−0.99	1.18
1 versus 5						−0.47
2 versus 5						−0.71
3 versus 5						0.90
4 versus 5						−0.29
Non-STEM
1 versus 2	−0.16	0.26	−0.57	1.17	0.00	−0.28
1 versus 3	−0.23	−0.22	−0.20	1.19	−0.68	0.79
1 versus 4	−0.25			−0.76	−0.14	−0.07
2 versus 3	−0.08	0.05	−0.37	−0.02	−0.68	−1.07
2 versus 4	0.09			0.41	−0.13	0.20
3 versus 4	0.01			0.43	0.54	−0.86
1 versus 5						−0.06
2 versus 5						0.22
3 versus 5						−0.85
4 versus 5						0.01

Note. STEM = science, technology, engineering, mathematics.

Interesting results were found for Confucian Asia, the Middle East, and Latin America. In Confucian Asia, only 9% of test takers in Profile 1, with the highest scores, except for GRE Quantitative, were interested in STEM degrees. In contrast, the majority of test takers in Profile 2, which had the highest GRE Quantitative scores, and Profile 3, with similar GRE Quantitative scores to those of Profile 1 but lower scores in other test sections, were interested in STEM degrees. Although effect sizes when comparing the proportion in Profile 1 with those of Profile 2 and Profile 3 were large, the effect size comparing Profile 2 and Profile 3 for the non-STEM proportions was close to zero and very small for STEM fields. These results suggest a closer match between intended field of study in Profile 1 and actual scores.

A majority of test takers (68%) in the highest performing profile (Profile 1) in the Middle East reported being interested in pursuing a STEM field. This percentage increased to 79% in Profile 2, which had the same GRE Quantitative scores as test takers in Profile 1 but lower scores on every other test section. In Latin America, preference for STEM fields was the strongest in Profile 3 (77%), although Profile 3 was defined by below average test scores.

Regarding the proportion of missing data, there are some noteworthy findings in Confucian Asia, Latin America, and the Middle East. Profile 2 in Confucian Asia, which had the highest scores in all test sections, except for GRE Quantitative, had the smallest proportion of missing data across all profiles in all countries and cultural regions (7%), while the proportion of missing data was much higher (38%) in the lowest performing profile, Profile 4. In comparison to other profiles in Confucian Asia, Profiles 2 and 4 showed small and medium effect sizes. Similarly, Profile 4 in Latin America, which is the lowest performing profile, had the highest proportion of missing data (42%) in comparison to the other profiles with small and medium effect sizes. While the lowest performing profile in the Middle East, Profile 5, also had the

Table 11 Estimated Proportion of Test Takers Pursuing Master's or Doctoral Programs in Each Profile per Country/Cultural Region

Degree by profile	India[a]	China[b]	Southern Asia[c]	Confucian Asia[d]	Latin America[e]	Middle East[f]
Full sample						
Missing	0.29	0.58	0.43	0.25	0.2	0.45
Master's	0.62	0.23	0.17	0.35	0.41	0.21
Doctoral	0.10	0.19	0.40	0.40	0.40	0.34
Profile 1						
Missing	0.23	0.53	0.39	0.25	0.21	0.44
Master's	0.59	0.25	0.23	0.24	0.32	0.15
Doctoral	0.19	0.22	0.38	0.51	0.47	0.42
Profile 2						
Missing	0.27	0.61	0.47	0.25	0.20	0.43
Master's	0.64	0.22	0.11	0.28	0.41	0.17
Doctoral	0.10	0.18	0.42	0.47	0.39	0.40
Profile 3						
Missing	0.33	0.63	0.40	0.25	0.20	0.37
Master's	0.62	0.23	0.22	0.40	0.41	0.29
Doctoral	0.05	0.14	0.38	0.36	0.39	0.34
Profile 4						
Missing	0.42			0.26	0.18	0.47
Master's	0.56			0.44	0.48	0.26
Doctoral	0.03			0.30	0.34	0.28
Profile 5						
Missing						0.64
Master's						0.24
Doctoral						0.11

[a]$N = 54,186.$ [b]$N = 20,903.$ [c]$N = 5,629.$ [d]$N = 4,652.$ [e]$N = 4,222.$ [f]$N = 2,420.$

highest proportion of missing data (40%), Profile 3, with the second highest scores in every test section, except for GRE Quantitative, had the same proportion of missing data.

Master's Versus Doctoral Programs

As in the case for field of study, owing to the large proportion of missing data in degree level, ranging from 20% to 58% (Table 11), we included missing data as an additional category of comparison.

Interest in pursuing a master's versus a doctoral degree varied across country and cultural region. Overall, test takers in China and India showed a strong preference for pursuing master's degrees, whereas test takers in Southern Asia showed a strong preference for doctoral degrees (Table 11). Although there was also a strong preference for doctoral degrees in Confucian Asia, Latin America, and the Middle East, in general, the proportion of test takers seeking master's degrees increased in the lower performing profiles. Although most of the comparisons across profiles showed statistically significant results (Table 12), most comparisons resulted in small effect sizes (Table 13). Only some comparisons for doctoral degrees in India and the Middle East showed medium effect sizes, which correspond to the comparisons of the lowest performing profile with the largest proportion of missing data (Profile 4 and Profile 5 in India and the Middle East, respectively).

One interesting finding is that the missing data proportions for degree level are much higher than the ones observed in field of study, and China shows some of the highest proportions of missing data (53%–61%, depending on the profile). Furthermore, we see that in India, the missing data proportions increase in the lower performing profiles (small effect sizes; Table 13) and that the lowest performing profile in the Middle East, Profile 5, has the highest proportion of missing data (64%).

Discussion

In this study, we utilized FMM to investigate different patterns of international test takers' performance on both the GRE and TOEFL and their relationships with demographic and graduate program characteristics. We examined the score

Table 12 Approximate Chi-Square Values for Proportion Comparison by Degree Level

Profile comparisons	India	China	Southern Asia	Confucian Asia	Latin America	Middle East
Overall test	1,627.31***	170.35***	70.45***	50.94***	20.87**	114.83***
1 versus 2	331.28***	86.25***	54.66***	1.17	8.28*	0.28
1 versus 3	984.61***	141.14***	0.66	20.32***	11.22**	7.91*
1 versus 4	1,229.42***			31.03***	8.25*	25.26***
2 versus 3	560.42***	22.16***	42.17***	19.89***	0.07	5.70
2 versus 4	331.28***			27.97***	11.22**	17.26***
3 versus 4	118.33***			4.31	1.40	4.56
1 versus 5						64.39***
2 versus 5						61.51***
3 versus 5						27.98***
4 versus 5						25.85***

$*p < .05.$ $**p < .01.$ $***p < .001.$

Table 13 Effect Sizes (Cohen's *h*) for the Comparison of Degree Levels

Profile comparisons	India	China	Southern Asia	Confucian Asia	Latin America	Middle East
Missing data						
1 versus 2	0.09	−0.16	0.17	0.01	−0.01	−0.01
1 versus 3	0.23	0.22	0.03	0.02	−0.01	−0.13
1 versus 4	0.41			0.02	−0.06	0.06
2 versus 3	0.14	0.05	0.14	−0.01	0.00	0.13
2 versus 4	−0.32			0.03	−0.05	0.07
3 versus 4	0.41			0.04	−0.05	0.19
1 versus 5						0.42
2 versus 5						0.43
3 versus 5						0.55
4 versus 5						0.36
Master's						
1 versus 2	0.10	0.08	−0.33	−0.10	0.18	0.04
1 versus 3	0.06	−0.06	−0.03	−0.34	0.19	0.34
1 versus 4	−0.07			0.42	0.32	0.28
2 versus 3	−0.04	0.02	−0.29	0.24	0.01	−0.30
2 versus 4	0.17			0.32	0.15	0.23
3 versus 4	−0.07			0.08	0.13	−0.07
1 versus 5						0.24
2 versus 5						0.20
3 versus 5						−0.10
4 versus 5						−0.03
Doctoral						
1 versus 2	−0.27	0.11	0.08	0.08	−0.16	−0.02
1 versus 3	−0.43	−0.21	0.00	0.30	−0.18	−0.15
1 versus 4	−0.54			−0.42	−0.28	−0.30
2 versus 3	−0.17	−0.10	0.08	−0.22	−0.01	0.13
2 versus 4	0.28			−0.34	−0.11	−0.27
3 versus 4	−0.54			−0.12	−0.10	−0.15
1 versus 5						−0.71
2 versus 5						−0.69
3 versus 5						−0.56
4 versus 5						−0.42

profiles across two countries considered independently and four cultural regions comprising multiple countries and found the following broad patterns of results: (a) Most countries and cultural regions, except for the Middle East, had three or four latent profiles representing low, medium, and high scores on the GRE and TOEFL sections; (b) two high-performing profiles were found in Confucian Asia, one with higher GRE Quantitative scores and the other with higher scores on GRE Verbal and TOEFL; (c) regardless of profile, test takers from China performed highest on the GRE Quantitative section;

(d) there was a relationship with students in the lower performing profiles taking the TOEFL and GRE multiple times; (e) regardless of country or cultural region, men were represented more than women overall and across most of the profiles; and (f) international students showed a preference for STEM-based fields and master's degrees, but this varied across country and cultural region. These results are discussed in more detail.

Five Profiles in the Middle East

The Middle East (along with North Africa) accounts for 9% of international students in the United States—the second highest region after Asia (IIE, 2018a); however, changes in visa requirements may impact students' willingness or ability to study abroad in the United States moving forward (IIE, 2018a). The Middle East cultural region consisted of test takers from seven countries, with roughly half the sample from Turkey. Two of the profiles identified (Profiles 4 and 5) had scores below the mean, whereas Profile 1 had the highest scores. Profiles 2 and 3 had scores close to the mean on every scale, with a few exceptions. Profile 2 had GRE Quantitative scores comparable to the high-performing Profile 1. Profile 3 had below average GRE Quantitative scores, comparable to the lower performing Profile 4, but high TOEFL Speaking scores, comparable to the high-performing Profile 1. When looking at the outcomes related to these two profiles, we found that 82% of test takers in Profile 2 were male and were interested in pursuing doctoral (40%) STEM-based degrees (79%). Alternatively, for Profile 3, 67% were female, with the majority of test takers interested in pursuing non-STEM degrees (53%), and approximately one-third were interested in a master's degree (29%), were interested in a doctoral degree (34%), or did not indicate their degree of interest (37% missing).

Despite the overall high performance of test takers in Middle East Profile 3, there was less certainty about pursuing a master's versus a doctoral degree. Given the higher proportion of women in this profile, these results also suggest that there may be differences in the norms and expectations between men and women in the Middle East with regard to graduate education, specifically around field of study and pursuing higher-level degrees (i.e., doctoral vs. master's). For instance, despite the focus on gender equality for women in Turkey, there still appear to be gender disparities. Findik (2016) and Çobanoğlu (2018) both found the largest gender disparities in graduate programs and that the disparities in master's degree and doctoral degree programs are roughly similar. Both studies also found that women are overrepresented in health and welfare programs and education but underrepresented in engineering, manufacturing, and construction and in social science, business, and law (Çobanoğlu, 2018; Findik, 2016).

In Saudi Arabia, one of the main reasons for the gender differences could be that despite some lifting of restrictions, it is still socially unacceptable for women to pursue certain careers in Saudi Arabia, and many Saudi women are restricted to careers in sectors like health care and education (Alsubaie & Jones, 2017). As a result, many families will send their daughters abroad to create opportunities in various fields. This could be one of the main reasons for the higher proportion of women in Middle East Profile 3. Additionally, Buckner (2013) found that in Egypt, gender is an important predictor of university enrollment and that women are much more likely to be enrolled in humanities, arts, and education, rather than in fields like engineering. Similarly, Tjomsland (2009) noted gender segregation in the Arab-Muslim world and that the proportion of women in higher education has historically been very low in most Arab countries. That said, some countries, such as Egypt and Tunisia, now offer free higher education, which could help to increase enrollment and could be one step toward increasing equality among men and women (Krafft & Alawode, 2018).

Two High-Performing Profiles for Confucian Asia

The cultural region of Confucian Asia was made up of four countries, with most test takers from Taiwan and the Republic of Korea. Results from Confucian Asia (Figure 6) are unique in that there were two high-performing profiles (Profiles 1 and 2). Profiles 3 and 4 corresponded to medium and low performance, respectively. Specifically, test takers in Profile 1 showed the highest scores on all test sections, except for GRE Quantitative and TOEFL Reading, where their scores were comparable to those of test takers in Profiles 3 and 2, respectively. Test takers in Profile 1 were mostly female (70%) and interested in applying to non-STEM fields (63%). Alternatively, test takers in Profile 2 were mostly male (80%) and interested in applying to STEM fields (82%). In both Taiwan and the Republic of Korea (the two largest represented countries in the Confucian Asia cultural region), there is a gender imbalance that has been found in STEM-based fields. For instance, in Taiwan, more than two-thirds of all doctoral test takers pursuing STEM-based fields are men (Marginson et al., 2013). Additionally, some researchers have argued that sociocultural norms and attitudes related to the Confucian tradition in

South Korea may influence the fields of study that women pursue; that is, women are more likely to pursue fields related to social science rather than natural science (United Nations Educational, Scientific, and Cultural Organization, 2015). This could partially explain the differences between the two profiles in terms of gender and intended field of study.

Compared to test takers in Profile 2, test takers in Profile 1 also had significantly lower undergraduate GPAs and were older in age. Despite these differences, both Profiles 1 and 2 had 51% and 47% of test takers interested in applying to doctoral programs, respectively. It is not surprising that Profile 2 had the highest GRE Quantitative scores and thus had test takers more interested in attending STEM-based programs. These results for Confucian Asia suggest that there are inherent differences in these test-score profiles based on the type of program and field of study that test takers are interested in pursuing within the United States.

High GRE Quantitative Reasoning Scores for China

As expected, regardless of test taker profile, test takers from China performed highest in GRE Quantitative; however, Southern Asia showed the opposite trend, with GRE Quantitative scores being the lowest across most profiles of test takers. These results support the common perception among faculty that test takers from China tend to have higher quantitative reasoning scores (Posselt, 2016).

Across profiles, there was a statistically significant higher proportion of men in Profiles 2 (58%) and 3 (57%) as compared to Profile 1 (52%); however, the effect was small. Despite statistically significant differences in the proportion of test takers interested in pursuing STEM-based degrees across profiles, all profiles showed more than 57% of test takers interested in STEM-based fields. The interest in STEM-based degrees is not surprising; in China, a science and engineering path in secondary school is considered the best route to getting into top universities, and more than 50% of test takers enroll in STEM-based bachelor's degree programs (Marginson et al., 2013). Additionally, China leads the world with the highest number of STEM graduates, with a total of 4.7 million in 2016 (McCarthy, 2017). This focus on STEM-based degrees likely also explains why test takers had higher GRE Quantitative scores overall, regardless of test taker profile. In relation to program type, many students were less certain about pursuing either a master's or doctoral degree, with more than 50% of the data missing for all three profiles. These results could suggest that Chinese test takers are more certain about their field of study than about the degree level they wish to attain.

The Effect of Taking Admissions Tests Multiple Times

One notable finding is the number of times that test takers took either the GRE or TOEFL across the various profiles within each country or cultural region. Consistent trends were found regarding the number of times test takers took the TOEFL, with test takers in the lower performing profiles across countries and cultural regions taking the TOEFL significantly more times than test takers in the highest performing profile. These results imply that students with high and low scores have different behaviors. That is, students with high scores may get higher scores the first time they take the assessment and thus do not feel the need to take the assessment again, whereas students with lower scores may feel the need to take the assessment more times to try to improve their performance.

For the GRE, results in India show the same pattern as for the TOEFL, with test takers in lower performing profiles taking the GRE more times as compared to the highest performing profile. Opposite trends were found with the GRE for Latin America and Confucian Asia: Test takers in lower performing profiles took the GRE significantly fewer times than test takers in the highest performing profile. Results for China, Southern Asia, and the Middle East found no statistically significant differences in the number of times test takers took the GRE across profiles.

It is interesting that varying trends were found across the two testing programs. One reason for this difference could be related to the fact that many programs have hard cut-scores for TOEFL as a minimum standard for admission (e.g., American Exam Services, 2014). As a result, test takers in lower performing profiles may need to take the assessment again to hit a certain cut-score for a particular program even to be considered. For the GRE, however, a lower score may direct students to less selective programs but may not prevent students from getting into a program, so there may be less of a desire to maximize scores when in a lower profile. However, high-performing students may be trying to get into a highly selective program and may perceive that additional points on their score would help them get into the program. Those higher performing students could also be seeking scholarships, fellowships, or other financial support that could depend on their GRE performance.

Disproportionate Gender and Performance

According to Okahana and Zhou (2019a), the majority of first-time graduate students in all degree levels were women in fall 2018, with 59.7% at the master's degree level and 54.4% at the doctoral level. However, when looking specifically at first-time graduate enrollment for international students, only 44.6% were women (Okahana & Zhou, 2019a). Our overall results are consistent with these findings, with men representing 55%–69% of test takers across the countries and cultural regions (Table 5). Our results also show a higher proportion of men across the profiles and fewer international women taking these assessments. Except for Confucian Asia, where the top-performing profile had 70% women, women tended to be underrepresented in the higher performing profiles. In Latin America, the proportion of women increased with the lower performing profiles.

In general, regardless of country, few profiles had a significantly higher number of women as compared to men. The largest differences in the proportion of men and women across profiles were found in Confucian Asia, Latin America, and the Middle East, with smaller effects in India, China, and Southern Asia. Among those profiles with more women, women typically performed highest on GRE Verbal and lowest on GRE Quantitative, which is expected based on trends in test scores from the past 5 years (ETS, 2019a). Even though women are highly represented in graduate education, cultural values and expectations still play a role in access to higher education for women (Tjomsland, 2009). Additionally, STEM-based fields tend to be the most sought-after fields for international students and to be dominated by men, which could be another reason for the lower proportion of women in our findings.

Intended Field of Study and Degree

As compared to students in the United States (17.2%), international students are more likely to enroll in graduate programs in STEM-based fields (54.7%; Posselt, 2016). Stephan et al. (2015) found that the most popular fields for foreign students include engineering and physical sciences. These findings were echoed in our results. In general, in every country and cultural region, students showed a strong preference for STEM fields regardless of score profile. There were two exceptions to this trend in the Middle East and in Confucian Asia: Only 8% of test takers in Profile 3 in the Middle East and 9% of test takers in Profile 1 in Confucian Asia indicated that they were interested in pursuing a STEM-based field. Interestingly, in the Middle East, this profile consisted of average or above average scores in GRE and TOEFL sections, except for GRE Quantitative, with scores 1 standard deviation below the mean. A similar but less pronounced pattern was found in Profile 1 in Confucian Asia, with GRE Quantitative having the lowest scores. These findings indicate that for students in these profiles, there may be a closer match between students' scores and their fields of interest. Because of the requirements of STEM programs, it is possible that students with low GRE Quantitative scores would show less of a preference for such degrees. One caveat is that it is not possible to know from our results if students in these profiles are not interested in STEM fields because they do not have higher GRE Quantitative scores or if students with no interest in pursuing a degree in STEM fields have not prioritized their learning of the quantitative skills captured by the GRE.

Regarding degree objective (master's vs. doctoral), results show large amounts of missing data ranging from 20% to 58% across countries and cultural regions for this particular variable. For China, Southern Asia, and the Middle East, the amount of missing data was particularly high. The amount of missing data was higher than the amount of missing data for intended field of study. This suggests that in general, students may have a better sense of the field of study in which they are interested but are less certain about the level of degree they would like to attain. That said, among the data that were available, results show preferences for doctoral degrees within the top profiles across most countries and cultural regions. These results are not surprising considering that the Organisation for Economic Co-operation and Development (2018) reported that the proportion of international students in doctoral programs in the United States is much larger than in master's programs. Only in India did test takers show a great preference for pursuing a master's degree, regardless of scores, with all four profiles showing a proportion of 56% or higher interested in pursuing a master's degree.

Limitations

One of the limitations of this study is that we had access only to background information from students' test registration forms, which likely resulted in the larger amounts of missing data for intended field of study and degree type. We also did not have access to the types of programs in which students ultimately enrolled but rather only the fields and types of

degrees they were interested in pursuing. It is possible that after receiving their scores, students could have changed their minds about the decision to go abroad or about the type of degree they intended to pursue.

Another limitation is subjectivity in selecting the number of profiles. Model selection in FMM remains an important research topic (Marsh et al., 2009), with different indices sometimes pointing to different solutions. This indicates that different solutions may result in an adequate model fit, and in these cases, selecting the final model might become more subjective. In other words, given that two or more solutions may show similar fit to the data, selecting one is subjective. Future studies could explore how much the findings by country or cultural region differ by selecting another model with similar fit.

Sample sizes within individual countries limited our ability to conduct analyses across individual countries. Instead, we had to group countries together into cultural regions. It is possible that results could vary across individual countries. Future research would also benefit from conducting a formal comparison across the various countries and cultural regions and conducting invariance studies in the FMM solution. Although the profiles might seem different, it could be the case that formal comparisons reveal the differences not to be statistically differentiable or meaningful.

Conclusions and Implications for Future Research

This study provided important insights about the relationships between international student test performance and demographic (gender, parental education) and graduate degree-related (field of study, degree level) factors. The study was exploratory in nature but provided the necessary first step in understanding the role of test scores as a factor in international students' decisions when applying to graduate school, especially because international students have identified admissions tests as a challenge when applying to graduate school (Bista & Dagley, 2015). Student profiles across countries and cultural regions provided us with a snapshot of how test taker performance on both the GRE and TOEFL may be related to intended field of study or degree. Results show different profile characteristics across countries and cultural regions, demonstrating the importance of not assuming that international students compose a homogeneous group.

Results from this study can be used to inform future research, such as focusing on longitudinal trends and looking at students who ultimately enrolled in a graduate program and whether their profiles were also differentially associated with being admitted to various other programs. Future research would also benefit from interviewing or surveying students about their perceptions about the role of admissions assessments in the decision-making process. It would be interesting to evaluate how student perceptions relate to their profile classification. For instance, do students have misconceptions about how high scores need to be to get into various programs, which may prevent them even from considering higher tier programs? Do students in lower performing profiles make different application decisions than students in higher performing profiles? Are there any trends across country or cultural region? Having a better understanding of factors impacting international students' decisions to pursue a graduate degree within the United States and the relationship with test-score performance could help recruitment and admissions officers achieve a more diverse and talented pool of applicants within their programs. Additionally, utilizing the profiles, along with other background information and academic interests, may be useful in helping to better match international students to graduate programs that may best fit their interests and needs. Matching international students to graduate programs may help to ensure that students are successful in ultimately completing those programs and achieving professional career success.

Notes

1 Cultural regions are the societal clusters that were used in the GLOBE research program. These clusters were created using empirical studies and other factors, such as common language, geography, religion, and historical accounts. The GLOBE research program identified 10 clusters, including Anglo, Latin Europe, Nordic Europe, Germanic Europe, Eastern Europe, Latin America, Middle East, Sub-Saharan Africa, Southern Asia, and Confucian Asia (House et al., 2004).

2 The tables with the fit of all the models compared per country and cultural region results are available in the Appendix, Tables A1 – A6.

References

Alsubaie, A., & Jones, K. (2017). An overview of the current state of women's leadership in higher education in Saudi Arabia and a proposal for future research directions. *Administrative Sciences*, *7*(36), 1–15. https://doi.org/10.3390/admsci7040036

American Exam Services. (2014). 16 ways to improve your TOEFL score. http://www.americanexamservices.com/about-the-toefl/16-ways-to-improve-your-toefl-score

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: 3-step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(3), 329–341, https://doi.org/10.1080/10705511.2014.915181

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods*, *9*(1), 3–29. https://doi.org/10.1037/1082-989X.9.1.3

Bista, K., & Dagley, A. (2015). Higher education preparation and decision-making trends among international students. *College and University*, *90*(3), 2–8.

Buckner, E. (2013). Access to higher education in Egypt: Examining trends by university sector. *Comparative Education Review*, *57*(3), 527–552. https://doi.org/10.1086/670665

Chen, L. H. (2007). East-Asian students' choice of Canadian graduate schools. *International Journal of Educational Advancement*, *7*, 271–306. https://doi.org/10.1057/palgrave.ijea.2150071

Clark, S. L., & Muthén, B. (2009). Relating latent class analysis results to variables not included in the analysis. http://www.statmodel.com/download/relatinglca.pdf

Çobanoğlu, F. (2018). Gender equality in Turkish higher education. *International Journal of Higher Education*, *7*(5), 97–113. https://doi.org/10.5430/ijhe.v7n5p97

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. Wiley. https://doi.org/10.1002/9780470567333

ETS. (2016). *A snapshot of the individuals who took the GRE general test: July 2013–June 2016*. https://www.ets.org/s/gre/pdf/snapshot_test_taker_data_2016.pdf

ETS. (2019a). *A snapshot of the individuals who took the GRE general test: July 2014–June 2019*. https://www.ets.org/s/gre/pdf/snapshot_test_taker_data_2019.pdf

ETS. (2019b). *GRE guidelines for the use of scores*. https://www.ets.org/gre/institutions/admissions/using_scores/guidelines?WT.ac=40361_owt19_180820

ETS. (2019c). *Structure of the TOEFL iBT test and revised TOEFL paper-delivered test*. https://www.ets.org/toefl/institutions/about/content/content_structure/

ETS. (2019d). *TOEFL iBT score use*. https://www.ets.org/toefl/institutions/scores/use/

ETS. (2019e). *TOEFL iBT test and score data summary 2019*. https://www.ets.org/s/toefl/pdf/94227_unlweb.pdf

Findik, L. Y. (2016). Gender disparity in Turkish higher education. *Universal Journal of Educational Research*, *4*(1), 2499–2504. https://doi.org/10.13189/ujer.2016.041029

Geiser, C., Lehmann, W., & Eid, M. (2006). Separating "rotators" from "non-rotators" in the mental rotations test: A multigroup latent class analysis. *Multivariate Behavioral Research*, *41*(3), 261–293. https://doi.org/10.1207/s15327906mbr4103_2

Haberman, S., & Yao, L. (2015). Repeater analysis for combining information from different assessments. *Journal of Educational Measurement*, *52*(2), 223–251. https://doi.org/10.1111/jedm.12075

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (Eds.). (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Sage.

Institute of International Education. (2018a). *A world on the move: Trends in global student mobility* (Issue 2). https://www.iie.org/-/media/Files/Corporate/Publications/World_On_Move_Final.ashx

Institute of International Education. (2018b). *Opendoors* 2018 *"fact facts."* https://www.iie.org/-/media/Files/Corporate/Open-Doors/Fast-Facts/Fast-Facts-2018.ashx

Kline, R. B. (2012) Assumptions in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 111–125). Guilford Press. https://doi.org/10.1080/10705511.2013.769397

Krafft, C., & Alawode, H. (2018). Inequality of opportunity in higher education in the Middle East and North Africa. *International Journal of Educational Development*, *62*, 234–244. https://doi.org/10.1016/j.ijedudev.2018.05.005

Lanza, S. T., Tan, X., & Bray, B. C. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling*, *20*(1), 1–26. https://doi.org/10.1080/10705511.2013.742377

Lei, S. A., & Chuang, N. (2010). Demographic factors influencing selection of an ideal graduate institution: A literature review with recommendations for implementation. *College Student Journal*, *44*(1), 84–96.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*(3), 767–778. https://doi.org/10.1093/biomet/88.3.767

Marginson, S., Tytler, R., Freeman, B., & Roberts, K. (2013). *STEM—Country comparisons: International comparisons of science, technology, engineering and mathematics (STEM) education* [Final report]. Australian Council of Learned Academies.

Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling*, *16*(2), 191–225. https://doi.org/10.1080/10705510902751010

Mattern, K. D., Marini, J., & Shaw, E. J. (2012). *Examining multiple profles of non-returners to develop targeted college retention interventions* [Paper presentation]. American Educational Research Association Annual Meeting, Vancouver, BC, Canada.

Mazzarol, T., & Soutar, G. N. (2001). *Push–pull factors influencing international student destination choice* (Discussion Paper No. DP 0105). Centre for Entrepreneurial Management and Innovation.

Mazzarol, T., & Soutar, G. N. (2002). The push–pull factors influencing international student selection of education destination. *International Journal of Educational Management*, *16*(2), 82–90. https://doi.org/10.1108/09513540210418403

McCarthy, N. (2017). The countries with the most STEM graduates. https://www.statista.com/chart/7913/the-countries-with-the-most-stem-graduates/

National Center for Education Statistics. (2019, January). *Profile and financial estimates of graduate students: 2015–16* (Report No. 2019-469). U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/pubs2019/2019469.pdf

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, *14*(4), 535–569. https://doi.org/10.1080/10705510701575396

Okahana, H., & Zhou, E. (2018). *International graduate applications and enrollment: Fall 2017*. Council of Graduate Schools.

Okahana, H., & Zhou, E. (2019a). *Graduate enrollment and degrees: 2008 to 2018*. Council of Graduate Schools. https://cgsnet.org/publication-pdf/5877/CGS_GED18_Report_web.pdf

Okahana, H., & Zhou, E. (2019b). *International graduate applications and enrollment: Fall 2018*. Council of Graduate Schools. https://www.cgsnet.org//ckfinder/userfiles/files/Intl_Survey_Report_Fall2018.pdf

Olivera-Aguilar, M., Rikoon, S. H., & Robbins, S. B. (2017). Using latent profile analysis to identify noncognitive skill profiles among college students. *Journal of Higher Education*, *88*(2), 234–257. https://doi.org/10.1080/00221546.2016.1244413

Organisation for Economic Co-operation and Development. (2018). *Education at a glance 2018: OECD indicators*. https://doi.org/10.1787/eag-2018-en

Pastor, D. A., & Barron, K. E. (2012). *Capturing the relationships between achievement goals and outcome: Latent profile analysis versus multiple regression* [Paper presentation]. American Educational Research Association Annual Meeting, Vancouver, BC, Canada.

Pastor, D. A., Barron, K. E., Miller, B. J., & Davis, S. L. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemporary Educational Psychology*, *32*(1), 8–47. https://doi.org/10.1016/j.cedpsych.2006.10.003

Posselt, J. R. (2016). International students and ambiguities of holistic review. In J. R. Posselt (Ed.), *Inside graduate admissions: Merit, diversity, and faculty gatekeeping* (pp. 133–153). Harvard University Press. https://doi.org/10.4159/9780674915640-007

Robbins, S. B., Lauver, K., Le, H., Davis, D., & Langley, R. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, *130*(2), 261–288. https://doi.org/10.1037/0033-2909.130.2.261

Schmitt, N., Oswald, F. L., Kim, B. H., Imus, A., Drzakowski, S., Friede, A., & Shivpuri, S. (2007). The use of background and ability profiles to predict college student outcomes. *Journal of Applied Psychology*, *92*, 165–180. https://doi.org/10.1037/0021-9010.92.1.165

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Stephan, P., Scellato, G., & Franzoni, C. (2015). International competition for PhDs and postdoctoral scholars: What does (and does not) matter. In W. R. Kerr, J. Lerner, & S. Stern (Eds.), *Innovation policy and the economy* (Vol. *15*, pp. 73–113). University of Chicago Press. https://doi.org/10.1086/680060

Sulak, T. N., Massey, J., & Thomson, D. (2017). Using latent profile analysis to harness the heterogeneity of nonretained college students. *Journal of College Student Retention: Theory and Practice*, *18*(4), 457–473. https://doi.org/10.1177/1521025115611385

Tein, J. Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(4), 640–657. https://doi.org/10.1080/10705511.2013.824781

Tjomsland, M. (2009). Women in higher education: A concern for development? *Gender, Technology, and Development*, *13*(3), 407–427. https://doi.org/10.1177/097185241001300305

Tofighi, D., & Enders, C. K. (2008). Identifying the correct number of classes in growth mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 317–341). Information Age.

Tseng, S.-F., Tsao, Y.-W., Yu, L.-C., Chan, C.-L., & Lai, K. R. (2016). Who will pass? Analyzing learning behaviors in MOOCs. *Research and Practice in Technology Enhanced Learning*, *11*(8), 1–11. https://doi.org/10.1186/s41039-016-0033-5

United Nations Educational, Scientific, and Cultural Organization. (2015). *A complex formula: Girls and women in science, technology, engineering and mathematics in Asia.*

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, *18*(4), 450–469. https://doi.org/10.1093/pan/mpq025

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89–106). Cambridge University Press. https://doi.org/10.1017/CBO9780511499531.004

Appendix

FMM Results by Country/Cultural Region

Table A1 India Finite Mixture Models A–D

No. profiles	Model	Covariance structure	AIC	BIC	aBIC	aLMR
3	A	Fixed variance, zero covariance	865,999.8	866,266.8	866,171.4	64,260.87
3	B	Free variance, zero covariance	839,452.9	839,844.5	839,704.6	67,587.04
3	C	Free variance, fixed covariance	792,077.1	792,655.7	792,449.1	13,681.20***
3	D	Free variance, free covariance	783,188.6	784,141.0	783,800.9	12,875.74***
4	A	Fixed variance, zero covariance	837,557.2	837,895.4	837,774.6	28,135.95***
4	B	Free variance, zero covariance	809,738.1	810,263.2	810,075.7	29,563.98***
4[a]	C	Free variance, fixed covariance	785,187.9	785,900.0	785,645.7	6,877.14***
4	D	Free variance, free covariance	777,625.0	778,897.7	778,443.2	5,621.34***
5	A	Fixed variance, zero covariance	825,947.7	826,357.1	826,210.9	11,493.68***
5	B	Free variance, zero covariance	798,264.3	798,922.9	798,687.7	11,433.84***
5	C	Free variance, fixed covariance	781,754.6	782,600.1	782,298.2	3,442.32***
5	D	Free variance, free covariance	775,057.0	776,650.1	776,081.3	2,633.26***
6	A	Fixed variance, zero covariance	818,951.5	819,432.1	819,260.4	6,932.72***
6	B	Free variance, zero covariance	818,951.5	819,432.1	819,260.4	6,932.72***
6	C	free variance, fixed covariance	778,927.3	779,906.3	779,556.7	2,839.95***
6	D	Free variance, free covariance	773,107.7	775,021.3	774,338.0	−2,692.30
7	A	Fixed variance, zero covariance	814,893.9	815,445.7	815,248.7	4,027.39***
7	B	Free variance, zero covariance	788,327.2	789,252.8	788,922.3	4,668.58***
7	C	Free variance, fixed covariance				
7	D	Free variance, free covariance	771,406.5	773,640.4	772,842.8	1,520.14

Note. If no results, model did not converge. AIC = Akaike information criterion; BIC = Bayesian information criterion; aBIC = adjusted Bayesian information criterion; aLMR = adjusted Lo–Mendell–Rubin likelihood ratio. *$p < .05$. **$p < .01$. ***$p < .001$. [a] Denotes final model.

Table A2 China Finite Mixture Models A – D

No. profiles	Model	Covariance structure	AIC	BIC	aBIC	aLMR
3	A	Fixed variance, zero covariance	283,168.1	283,406.5	283,311.1	12,881.03[***]
3[a]	B	Free variance, zero covariance	268,071.9	268,421.6	268,281.8	15,766.15[***]
3	C	Free variance, fixed covariance	257,359.2	257,875.8	257,669.2	5,772.86[***]
3	D	Free variance, free covariance				
4	A	Fixed variance, zero covariance	277,048.6	277,350.6	277,229.8	6,059.34[***]
4	B	Free variance, zero covariance	262,201.4	262,670.3	262,482.8	5,861.19[**]
4	C	Free variance, fixed covariance				
4	D	Free variance, free covariance				
5	A	Fixed variance, zero covariance	273,547.0	273,912.6	273,766.4	3,473.93[***]
5	B	Free variance, zero covariance				
5	C	Free variance, fixed covariance				
5	D	Free variance, free covariance				
6	A	Fixed variance, zero covariance	271,557.0	271,986.2	271,814.5	1,981.12[***]
6	B	Free variance, zero covariance				
6	C	Free variance, fixed covariance				
6	D	Free variance, free covariance				
7	A	Fixed variance, zero covariance	270,228.1	270,720.9	270,523.9	1,328.14
7	B	Free variance, zero covariance				
7	C	Free variance, fixed covariance				
7	D	Free variance, free covariance				

Note. If no results, model did not converge. AIC = Akaike information criterion; BIC = Bayesian information criterion; aBIC = adjusted Bayesian information criterion; aLMR = adjusted Lo – Mendell – Rubin likelihood ratio. *$p < .05$. **$p < .01$. ***$p < .001$. [a] Denotes final model.

Table A3 Southern Asia Finite Mixture Models A – D

No. profiles	Model	Covariance structure	AIC	BIC	aBIC	aLMR
3	A	Fixed variance, zero covariance	91,983.45	92,182.52	92,087.19	5,971.48[***]
3	B	Free variance, zero covariance	89,461.48	89,753.45	89,613.63	6,517.67[***]
3	C	Free variance, fixed covariance	85,260.17	85,691.49	85,484.94	1,240.85[***]
3[a]	D	Free variance, free covariance	84,136.25	84,846.27	84,506.26	1,479.72[***]
4	A	Fixed variance, zero covariance	89,467.25	89,719.40	89,598.65	2,496.08[***]
4	B	Free variance, zero covariance	87,293.55	87,685.06	87,497.57	2,181.09[**]
4	C	Free variance, fixed covariance	84,695.93	85,226.79	84,972.57	589.68[***]
4	D	Free variance, free covariance	83,729.79	84,678.69	84,224.28	476.93
5	A	Fixed variance, zero covariance	88,428.67	88,733.91	88,587.74	1,039.53
5	B	Free variance, zero covariance				
5	C	Free variance, fixed covariance	84,340.76	84,971.15	84,669.27	382.22
5	D	Free variance, free covariance				
6	A	Fixed variance, zero covariance	87,871.32	88,229.65	88,058.06	565.19[***]
6	B	Free variance, zero covariance				
6	C	Free variance, fixed covariance				
6	D	Free variance, free covariance				
7	A	Fixed variance, zero covariance	87,386.74	87,798.15	87,601.13	493.45[***]
7	B	Free variance, zero covariance	84,652.54	85,342.65	85,012.17	2,000.52[***]
7	C	Free variance, fixed covariance				
7	D	Free variance, free covariance				

Note. If no results, model did not converge. AIC = Akaike information criterion; BIC = Bayesian information criterion; aBIC = adjusted Bayesian information criterion; aLMR = adjusted Lo – Mendell – Rubin likelihood ratio. *$p < .05$. **$p < .01$. ***$p < .001$. [a] Denotes final model.

Table A4 Confucian Asia Finite Mixture Models A–D

No. profiles	Model	Covariance structure	AIC	BIC	aBIC	aLMR
3	A	Fixed variance, zero covariance	72,424.33	72,617.69	72,522.36	3,089.17***
3	B	Free variance, zero covariance	70,075.52	70,359.10	70,219.29	3,777.84***
3	C	Free variance, fixed covariance	67,152.70	67,571.63	67,365.08	862.32*
3	D	Free variance, free covariance				
4	A	Fixed variance, zero covariance	70,753.70	70,998.61	70,877.86	1,662.03***
4	B	Free variance, zero covariance	68,448.19	68,828.45	68,640.97	1,644.34*
4[a]	C	Free variance, fixed covariance	66,545.07	67,060.67	66,806.46	632.64
4	D	Free variance, free covariance				
5	A	Fixed variance, zero covariance	70,020.61	70,317.09	70,170.92	738.16
5	B	Free variance, zero covariance	67,510.78	67,987.72	67,752.57	959.83
5	C	Free variance, fixed covariance				
5	D	Free variance, free covariance				
6	A	Fixed variance, zero covariance	69,663.78	70,011.82	69,840.22	367.39
6	B	Free variance, zero covariance				
6	C	Free variance, fixed covariance				
6	D	Free variance, free covariance				
7	A	Fixed variance, zero covariance	69,294.66	69,694.00	69,497.25	379.50
7	B	Free variance, zero covariance				
7	C	Free variance, fixed covariance				
7	D	Free variance, free covariance				

Note. If no results, model did not converge. AIC = Akaike information criterion; BIC = Bayesian information criterion; aBIC = adjusted Bayesian information criterion; aLMR = adjusted Lo–Mendell–Rubin likelihood ratio. *$p < .05$. **$p < .01$. ***$p < .001$. [a] Denotes final model.

Table A5 Latin America Finite Mixture Models A–D

No. profiles	Model	Covariance structure	AIC	BIC	aBIC	aLMR
3	A	Fixed variance, zero covariance	68,835.11	69,025.55	68,930.22	3,879.77*
3	B	Free variance, zero covariance	65,732.89	66,012.20	65,872.39	4,305.14***
3	C	Free variance, fixed covariance	62,288.15	62,700.77	62,494.23	1,229.82
3	D	Free variance, free covariance	61,394.33	62,073.57	61,733.57	1,025.02***
4	A	Fixed variance, zero covariance	66,601.68	66,842.91	66,722.16	2,216.24***
4	B	Free variance, zero covariance	63,853.65	64,228.19	64,040.71	1,894.11***
4	C	Free variance, fixed covariance	61,466.21	61,974.05	61,719.85	845.19
4[a]	D	Free variance, free covariance	60,911.44	61,819.21	61,364.82	553.05***
5	A	Fixed variance, zero covariance	65,739.86	66,031.87	65,885.7	864.87***
5	B	Free variance, zero covariance	63,026.23	63,495.98	63,260.84	850.63
5	C	Free variance, fixed covariance	61,386.96	61,990.02	61,688.15	108.39
5	D	Free variance, free covariance	60,590.94	61,727.24	61,158.45	391.20
6	A	Fixed variance, zero covariance	65,309.05	65,651.85	65,480.26	440.21*
6	B	Free variance, zero covariance	62,446.17	63,011.15	62,728.34	605.22*
6	C	Free variance, fixed covariance	61,002.77	61,701.06	61,351.52	410.90**
6	D	Free variance, free covariance				
7	A	Fixed variance, zero covariance	64,917.38	65,310.96	65,113.95	401.66*
7	B	Free variance, zero covariance	62,063.43	62,723.63	62,393.16	409.47
7	C	Free variance, fixed covariance				
7	D	Free variance, free covariance				

Note. If no results, model did not converge. AIC = Akaike information criterion; BIC = Bayesian information criterion; aBIC = adjusted Bayesian information criterion; aLMR = adjusted Lo–Mendell–Rubin likelihood ratio. *$p < .05$. **$p < .01$. ***$p < .001$. [a] Denotes final model.

Table A6 Middle East Finite Mixture Models A – D

No. profiles	Model	Covariance structure	AIC	BIC	aBIC	aLMR
3	A	Fixed variance, zero covariance	42,806.47	42,980.22	42,884.90	2,778.02[***]
3	B	Free variance, zero covariance	41,663.84	41,918.67	41,778.87	2,795.41[***]
3	C	Free variance, fixed covariance	39,221.60	39,598.05	39,391.53	644.71[***]
3	D	Free variance, free covariance	38,766.35	39,386.04	39,046.08	623.64[**]
4	A	Fixed variance, zero covariance	41,512.43	41,732.51	41,611.78	1,289.35[***]
4	B	Free variance, zero covariance	40,324.20	40,665.90	40,478.44	1,358.02[***]
4	C	Free variance, fixed covariance	38,916.49	39,379.82	39,125.64	332.26[***]
4	D	Free variance, free covariance	38,504.56	39,332.75	38,878.41	332.60[**]
5	A	Fixed variance, zero covariance	41,083.52	41,349.93	41,203.77	437.89
5	B	Free variance, zero covariance	39,784.67	40,213.24	39,978.12	564.70[*]
5[a]	C	Free variance, fixed covariance	38,744.76	39,294.95	38,993.11	200.03[*]
5	D	Free variance, free covariance				
6	A	Fixed variance, zero covariance	40,780.15	41,092.90	40,921.33	314.32[***]
6	B	Free variance, zero covariance	39,387.10	39,902.54	39,619.77	423.94[**]
6	C	Free variance, fixed covariance	38,604.14	39,241.21	38,891.71	169.17
6	D	Free variance, free covariance				
7	A	Fixed variance, zero covariance	40,554.25	40,913.33	40,716.34	238.08
7	B	Free variance, zero covariance	39,138.08	39,740.39	39,409.96	276.65
7	C	Free variance, fixed covariance	38,468.06	39,192.00	38,794.85	164.67
7	D	Free variance, free covariance				

Note. If no results, model did not converge. AIC = Akaike information criterion; BIC = Bayesian information criterion; aBIC = adjusted Bayesian information criterion; aLMR = adjusted Lo – Mendell – Rubin likelihood ratio. [*]$p < .05$. [**]$p < .01$. [***]$p < .001$. [a]Denotes final model.