

IELTS and Written Syntactic Complexity as Predictors of GPA of Multilingual International Graduate Students

Nihat Polat 
Laura Mahalingappa 

University of Maryland, College Park

Rae Mancilla

University of Pittsburgh

Language proficiency policies act as critical gatekeepers for multilingual international students (MIS) who intend to pursue an education at English-medium universities. The current study revisits the question of the predictive validity of a standardized English proficiency exam—International English Language Testing System (IELTS)—in comparison with 10 indices of written syntactic complexity. This study addressed the following questions: (a) Do MIS' overall proficiency and writing sub-test scores (IELTS) predict their grade point averages (GPAs) in graduate school? (b) Which syntactic complexity indices in online written production (subordination, coordination, length of production, and degree of phrasal sophistication) predict MIS' GPAs in graduate school? and (c) Which syntactic complexity indices remain predictive of their GPAs when participants' English proficiency scores (overall and/or writing IELTS scores) are added into the model? Participants' IELTS scores, GPAs, and asynchronous discussion postings (224) were analyzed through three linear regressions models. Results confirmed that while IELTS composite and writing subscores do predict academic success of graduate MIS, writing scores are better predictors. After accounting for the contributions of overall and writing subscores, mean length of sentence and T-unit remained significant predictors of academic outcomes, explaining a considerable amount of variance in GPAs, beyond that of IELTS. Equity implications for the admissions of MIS are considered.

Keywords: *language proficiency tests, multilingual international students, L2 writing, grammatical complexity, asynchronous online discussions, GPA*

The recruitment and admission of international students globally has become an area of increasing focus for institutions of higher learning. Universities are not only incentivized to increase international student enrollment, ostensibly as a reaction to diversity, equity, and inclusion (DEI) initiatives, but also pragmatically to reap the economic benefits of increased revenues from tuition (housing, fees, etc.) (Sá & Sabzalieva, 2018). In fact, increasing international student enrollment by as much as 5% annually over the next decade is a part of strategic planning for some institutions (Laad & Sharma, 2021). In English-dominant countries (e.g., United States, Canada, Australia), international students already constitute a sizable portion of overall enrollments, whose numbers have substantially increased in the past two decades (Institute of International Education, 2022). For example, the number of multilingual international students (MIS) in European universities has now passed 83,000, and in the United States, the top destination for MIS, this number

passed one million for the first time in 2016 (Israel & Batalova, 2021). Despite some declines in the few years due to the COVID-19 pandemic, in 2022, approximately 385,097 of these MIS (total=948,519) were graduate students (Institute of International Education, 2022). Given these numbers, it is important that universities adopt admission policies that ensure equitable educational opportunities for MIS (Cadena et al., 2023).

Recently, however, issues related to diversity, equity, and inclusion efforts have highlighted inequities that persist in graduate school admissions processes. For instance, concerns related to racial and socioeconomic disparities and overreliance on standardized assessments have plagued U.S. institutions (Cadena et al., 2023; Posselt, 2016). For international students specifically, in addition to other admission criteria (e.g., GPA, personal statement), admission to an English-medium university requires that potential students take standardized language proficiency exams to prove their



readiness in academic English and their ability to participate in their programs of study. Thus, most universities require scores of standardized English proficiency exams such as the Test of English as a Foreign Language (TOEFL) and the International English Language Testing System (IELTS). However, particularly at the graduate level, these institutional language proficiency policies and required tests are often based on neoliberal, racialized language ideologies that have acted as critical gatekeepers for MIS who intend to pursue an education at U.S. universities (Piller & Bodis, 2024). Life-changing decisions for the population, from admission into a graduate program to graduate assistantships and other funding opportunities, depend on TOEFL/IELTS exam scores (Oliver et al., 2012; Pritasari et al., 2019), despite other possible indicators of students' language abilities (e.g., previous language study, work history). The widespread use of these tests rests on the assumption that they reflect students' readiness for academic study in English and leverage predictive validity for future graduate school achievement, although in reality this is not entirely the case (Ihlenfeldt & Rios, 2023). These concerns have initiated a call for more holistic criteria for making admissions decisions (Posselt, 2016).

Relying on the assumption that performance on language proficiency exams is somehow predictive of student success is problematic for several reasons. First, despite meeting threshold language proficiency requirements (e.g., commonly IELTS=6.5, TOEFL iBT=80 varying by school), some MIS still struggle with the linguistic demands of graduate school (Biber et al., 2017; Mahalingappa et al., 2021; Schoepp, 2018). While the TOEFL/IELTS exams may measure general academic English proficiency, other factors can impact students' experiences, such as institutional and faculty-readiness to support MIS in their classrooms. For instance, research has found that although many universities have DEI initiatives that attempt to provide a more inclusive environment for students and enact culturally responsive and sustaining pedagogies, language is typically excluded from these efforts (Wolfram, 2023). Unfortunately, issues around language have largely been ignored in higher education, such that there are few faculty and staff on campuses that are aware of how language works in the classroom and what language proficiency tests really mean in terms of students' ability to use discipline-specific language in their academic classes. Further, many institutions do not provide faculty with the requisite knowledge and skills required to support MIS; thus, faculty ultimately expect students to be ready to engage in all linguistic tasks in their classes based on their admission to the university (Ginther & Yan, 2017; Mahalingappa et al., 2021).

Second, research on the predictive validity of such proficiency tests has been generally inconclusive, such that some research has found a correlation between TOEFL/IELTS scores and academic success, while others have not

(Bridgeman et al., 2016; Dooley & Oliver, 2002; Ihlenfeldt & Rios, 2023; Wongtrirat, 2010). Evaluating the predictive validity of proficiency tests is not a new undertaking in assessment research. For decades, researchers have studied the relationship between academic performance and MIS' results on proficiency exams with few conclusive findings (Graham, 1987; Schoepp, 2018). While language proficiency test scores are not the sole admission criteria (e.g., statements, letters), minimum entry-level scores are often used as a prerequisite by admissions officers to even consider an application for review (Deygers & Malone, 2019). Researchers now challenge overreliance on language proficiency test scores—especially overall cut scores as a prerequisite—as predictors of academic performance, which is most often measured as grade point average (GPA). Instead, they have begun experimenting with new models that combine overall and subtest scores with common admissions-related criteria (Graduate Record Examination [GRE] Graduate Management Admission Test [GMAT], etc.) to determine the most effective formula for forecasting academic success across undergraduate and graduate academic disciplines (Cho & Bridgeman, 2012; Johnson & Tweedie, 2021; Oliver et al., 2012).¹ However, general waivers of standardized tests during pandemic-era admissions cycles have also made more visible inequities inherent in reliance on these types of standardized exams. The prohibitive costs associated with preparing for and taking exams as well as cultural biases have highlighted barriers to potential students from historically underrepresented groups and international students in particular (Cadena et al., 2023; Sullivan et al., 2022). These issues raise questions for university decision-makers about whether supplemental screening mechanisms, such as interviews or writing samples, and alternative measures of English proficiency that might predict students' academic success should be added to students' admissions profiles (Ginther & Yan, 2017; Posselt, 2016; Pritasari et al., 2019; Thorpe et al., 2017).

Building on this trend around holistic admissions (Posselt et al., 2023) to reduce overreliance on proficiency tests as gatekeepers and considering the writing-intensive nature of graduate school (Norris & Ortega, 2009; Schoepp, 2018), the present study considers the potential of combining written syntactic complexity scores with standardized language proficiency exam scores in relation with MIS' graduate GPA. In additional language acquisition, written syntactic complexity is an important criterion for assessing MIS' level of linguistic performance (Norris & Ortega, 2009). It refers to “the range and sophistication of grammatical resources exhibited in language production” (Ortega, 2015, p.82) and has been tied to instructors' evaluations of MIS' writing quality (Yang et al., 2015). Because writing “is often the only means by which students' content knowledge is assessed in a large number of [graduate] disciplines” (Mazgutova & Kormos, 2015, p.3), syntactic complexity is

likely to contribute to students' course grades, and perhaps even more so for those in digital learning ecologies, where writing is the primary mode of communication.

The current study revisits the question of the predictive validity of standardized English proficiency exams in comparison with 10 indices of written syntactic complexity. It specifically reports on an analysis of the naturalistic English writing of MIS in an English as a second language (ESL) graduate program, examining how MIS' overall proficiency and writing subtest scores (IELTS) predict their GPAs, particularly which and how well syntactic complexity indices (subordination, coordination, length of production, and degree of phrasal sophistication) predict their GPAs relative to their standardized language test scores.

We chose the ESL graduate field (fully onsite program) as the focal population sample and examined written syntactic complexity measures (instead of accuracy or fluency) in connection with IELTS overall and writing scores for five reasons. First, a strong command of more syntactically-complex writing skills helps students produce assignments of higher quality. Since academic writing, particularly in graduate education with hybrid (online), is the most used language skill, it may have a high potential impact on students' GPA (Mazgutova & Kormos, 2015). Second, one of the goals of this study was to disentangle possible relationships between the writing subscore and indices of written syntactic complexity. Third, basing high-stakes admission decisions partially on a strictly defined (minimum) test score has limitations since it is not possible to identify a unilateral path between the decision and the test score (Chapelle, 2012). In fact, even if a predictive power is established, using a validated syntactic complexity analyzer as an additional L2 (second language) measurement tool can prove very useful for triangulation purposes, especially when freely accessible and practical online tools are available. Fourth, to our knowledge, no validity study has explored the ESL/applied linguistics graduate sub-specialty, which is a highly linguistically demanding field. Finally, the L2 Syntactic Complexity Analyzer (L2SCA), the computerized analytic tool (Lu, 2010; Lu & Ai, 2015) used in this study, allows for a multidimensional approach (10 indices under five dimensions), which affords multiple opportunities of comparison across numerous linguistic features (Figure 1).

Review of Literature

International Admissions and the Use of Standardized English Proficiency Tests

The admission of MIS into graduate programs in English-dominant countries usually involves additional criteria that exceed requirements for domestic applicants. MIS are required to submit standardized English language proficiency scores from universal commercial admissions exams (commonly TOEFL, or IELTS) in addition to general admissions

Measure	Abbreviation	Formula
<i>Length of production unit</i>		
Mean length of clause	MLC	# words/ # clauses
Mean length of sentence	MLS	# words/ # sentences
Mean length of T-unit	MLT	# words/# T-units
<i>Amount of subordination</i>		
Dependent clauses per clause	DC/C	# dependent clauses/# clauses
Dependent clauses per T-unit	DC/T	# dependent clauses/# T-units
<i>Amount of coordination</i>		
Coordinate phrases per clauses	CP/C	# coordinate phrases/# clauses
Coordinate phrases per T-unit	CP/T	# coordinate phrases/# T-units
T-units per sentence	T/S	# T-units/# sentences
<i>Degree of phrasal sophistication</i>		
Complex nominals per clause	CN/C	# complex nominals/# clauses
Complex nominals per T-unit	CN/T	# complex nominals/# T-units

FIGURE 1. Summary of automated measures of syntactic complexity.

testing, which may include the Scholastic Aptitude Test (SAT), the GRE, or the GMAT. While the TOEFL has a longer tradition of use within the United States, the IELTS has greater currency in other parts of the English-speaking world, including 100% of the universities in Australia and the United Kingdom (British Council, 2019; Ginther & Elder, 2014). IELTS has been considered “the world’s most popular English language proficiency test for higher education and global migration” (British Council, 2019, para.10).

TOEFL and IELTS are thematically the same in quantifying English competency across four language components (reading, writing, speaking, and listening), but some differences do exist, including the length of the exam, format, scoring, grading, cost, and acceptance rates (Wood, 2022). In general, TOEFL is more focused on academic English, while the IELTS assesses both academic and everyday English—both of which are needed for student success in graduate school. TOEFL is also slightly longer than the IELTS and is entirely computer-based, while IELTS speaking exams are conducted with an examiner, and the writing exam is handwritten (Bright, 2020). Further, task types and exam structure also differ, with TOEFL relying on multiple choice question formats and IELTS providing a variety of question types. Finally, scoring procedures vary, with TOEFL deploying a combination of artificial intelligence (AI) and human raters and IELTS only certified examiners. See Gagen (2019) for a detailed description of the IELTS exam.

Although the intent of standardized language proficiency exams is to assess how well international applicants can read, write, listen, and speak in academic English (Ihlenfeldt & Rios, 2023), they are often used solely in admission decisions for MIS (Bista & Dagley, 2015), rather than using their results in providing formative language support systems throughout their education. Academic success can be influenced by a broad range of social and academic factors for both domestic and international students, including study habits, content knowledge, motivation, finances, community support, and the like (Ginther & Yan, 2017). Indeed, the relationship between academic English proficiency and academic success is complex; however, proficiency in academic English as a predictor of success in graduate education is unfairly applied to only MIS.

This gatekeeping practice has been challenged by critical scholars who have raised issues of testing bias and equity. For instance, these tests tend to privilege certain academic varieties of English that may impact the validity of standardized language assessments for marginalized student groups (Barkhuizen & Strauss, 2020). Critical language scholars have begun to critique and question the very notion of “academic English” and its use in educational contexts in general (Flores & Rosa, 2015) and specifically in higher education (Matsumoto, 2022). This critical language perspective sees English as a lingua franca that has a variety of legitimate forms (global Englishes) that are not captured by these standardized proficiency exams. For the IELTS listening subtest, Aryadoust (2023) uncovered underrepresented accent coverage and biases in topics covered that impacted testing fairness. To reduce biases in commercialized exams, and in light of the high demand to study internationally and the value of MIS to universities, testing scholars have recently called for admissions decision-makers to employ holistic review strategies using additional measures to gauge potential for academic success (Posselt, 2016; Posselt et al., 2023). These holistic strategies could include rubrics that consider students’ previous experience learning and using English (e.g., through English-medium schooling), additional writing samples, recommendations, academic and extracurricular activities, interviews, and other assessment of language proficiency.

In spite of inequities in standardized language proficiency exams, research on the admissions practices of U.S. institutions has shown that institutions continue to rely on these scores as a pivotal component of MIS’ admissions packages. A study by Arcuino (2013) found that international students with GPAs or GRE scores below the established threshold were accepted when they had high TOEFL or IELTS scores. In addition to admissions acceptance decisions, language proficiency scores can bear financial and curricular implications for MIS. They are used to determine whether the completion of preenrollment language support courses (e.g., English for academic purposes) is required prior to matriculating into the coursework for their discipline. MIS’ efforts

to test out of EAP coursework by achieving a high score on the TOEFL/IELTS exams has spawned a billion-dollar test prep industry (Johnson & Tweedie, 2021).

Predictive Validity of Standardized English Proficiency Tests

Predictive validity refers to how well an assessment tool indexes the outcome it intends to measure (Brown & Abeywickrama, 2019). Theoretically, an admission test has good predictive validity when applicants who perform well on the test (e.g., SAT) also perform well in real life (e.g., college education). This assumption is predicated on another assumption, namely, that the skills measured on the test are well aligned with skills required in real life (Polat, 2016). In L2 assessment, predictive validity studies are concerned with how well standardized proficiency tests (e.g., IELTS) forecast MIS’ academic success, given that “students who are below a certain threshold in these skills will not succeed in an English learning environment” (Ihlenfeldt & Rios, 2023, p. 277).

Prediction studies have typically referenced achievement in terms of GPA (Abunawas, 2014; Johnston & Tweedie, 2021), with the understanding that a multidimensional approach to benchmarking student achievement is preferable, as a single criterion cannot capture the fullness of academic performance (Johnson & Tweedie, 2021; O’Connor & Paunonen, 2007). Nonetheless, GPA retains its importance as a proxy for achievement and continues to be a central selection measure in graduate admissions, scholarship funding, and postgraduation employment.

Prediction studies have also encompassed limited academic disciplines, many including science, engineering, and business (Dooley & Oliver, 2002; Kerstjens & Nery, 2000; Pritasari et al., 2019), while also classifying MIS into broad categories (e.g., humanities/fine arts) that can mask within-group differences (Bridgeman et al., 2016). Within the “social sciences,” proficiency scores may vary in their ability to predict achievement for subdisciplines, like education, political science, or psychology. In the United States, validity studies are scarce and, to our knowledge, have not ventured into sub-specialties, like ESL. MIS who specialize in ESL instruction are a unique case for validity because their mastery of English must allow them “to be good language models in the classroom” (Richards, 2010, p.103) and to accomplish linguistically demanding tasks such as creating instructional resources in English. Therefore, IELTS scores should theoretically leverage more predictive value for this subgroup.

IELTS and TOEFL literature has collectively explained little about their predictive potential. In the United States, IELTS investigations in undergraduate and mixed undergraduate/postgraduate samples have ranged in their findings from no significant to weak correlations between overall proficiency scores and GPA (Cotton & Conrow, 1998) to

strong relationships between the two variables (Hill et al., 1999). At best, test scores have explained up to 29% of the variance in MIS' grades in mixed (Oliver et al., 2012; Thorpe et al., 2017) and undergraduate-only groups (Schoepp, 2018). Similar patterns across both graduate and undergraduate samples can be noted across IELTS/TOEFL studies, with results spanning from negative (Neal, 1998) to strong positive correlations between language proficiency and GPA (Arrigoni & Clark, 2015; Oliver et al., 2012). The earlier works however, reflect earlier versions of both examinations, limiting their applicability to the revised, web-based formats (IELTS/TOEFL introduced computerized versions in 2005). Nevertheless, a 2021 meta-analysis of 32 IELTS/TOEFL studies and 132 effect sizes supported the conclusion that both admissions assessments do predict academic success in both undergraduate and graduate populations, albeit with a weak positive correlation, with no significant differences in their predictive power (Ihlenfeldt & Rios, 2023). Since this meta-analysis found that the "overall correlation was low . . . practitioners are cautioned from using standardized English-language proficiency test scores in isolation in lieu of a holistic application review during the admissions process" (Ihlenfeldt & Rios, 2023, p. 276)

Recent research on predictive validity of the IELTS has found weak to moderate correlations between IELTS scores and MIS' GPAs (e.g., Thorpe et al., 2017; Wang-Taylor & Daller, 2014) both at graduate and undergraduate levels. For example, Wang-Taylor and Daller (2014) reported .30 ($p < .01$) correlations between overall IELTS scores and MIS' grades. More recently, Schoepp (2018) reported an IELTS score of 6.0 to be a benchmark predictor of undergraduate MIS' GPA. Others like Oliver et al. (2012) reported that all IELTS scores but writing subscores were significant predictors of MIS's GPAs, with reading being the highest ($r = .29$), followed by the overall score ($r = .28$) and speaking ($r = .16$). Conversely, Arrigoni and Clark (2015) found significant correlations between MIS' GPAs and IELTS reading (.32) and writing (.29) subscores for students in a rhetoric and composition program. These findings were also echoed by Gagen (2019) in a meta-analysis of IELTS scores and MIS' GPAs, noting the strongest correlation as the reading subscore ($r = .21$), followed by the overall composite score ($r = .23$).

Researchers have experimented with alternative models for predicting graduate MIS' academic achievement using hierarchical regression (e.g., Morris & Maxey, 2014). In a large-scale study, Cho and Bridgeman (2012) ran prediction models that included TOEFL iBT, GRE (verbal and quantitative), GMAT, and SAT scores to assess TOEFL scores' contribution to predicting GPA. The predictive power of TOEFL composite scores was fairly low and comparable across subgroups, ranging from .24 to .26 for business, humanities and arts, and social sciences, explaining around 6% to 7% of the variance in GPAs. Combined with GMAT

scores, TOEFL explained an additional 6% of variance (total = 14%). GRE verbal scores were only predictive with the TOEFL composite, accounting for an additional 9% of the variance for humanities and arts, 3% for science and engineering, and 5% for social sciences.

Syntactic Complexity and MIS' Academic Writing

Syntactic complexity is an important construct in L2 writing assessment and is used to benchmark MIS' overall language proficiency, linguistic development, and writing quality (Lu & Ai, 2015). Research into syntactic complexity development has documented various trajectories for different proficiency levels (Ortega, 2015; Polat et al., 2020). As MIS writers grow in proficiency, they learn to "manipulate a language's combinatorial properties" (Crossley & McNamara, 2014, p. 66), incorporating more sophisticated structures into their writings. Since high levels of syntactic complexity, especially complex noun phrases, are characteristics of academic writing (Biber et al., 2016; Kyle & Crossley, 2018), TOEFL and IELTS scoring rubrics consider "grammatical features" when assessing L2 writing proficiency.

Despite its effect being moderated by the writing genre (Biber et al., 2016), L2 proficiency is a powerful modulator of complexity. Several cross-sectional studies have investigated differences in the syntactic structures used by MIS writers of varying levels of proficiency (e.g., Polat et al., 2020). Ortega (2003) identified multiple T-unit metrics that could reliably indicate MIS' levels of proficiency, while Lu (2011) reported two indices of coordination and phrasal elaboration as well as three indices of length. In comparing native speakers (NS) and MIS, Ai and Lu (2013) concluded that MIS produce, on average, "shorter clauses, sentences, and T-units, a smaller amount of subordination, and a smaller proportion of complex nominals than NS university students" (p. 261). Together, these studies underscore the importance of considering L2 proficiency in analyzing MIS' written texts.

To account for variation in complexity across levels of L2 proficiency, Norris and Ortega (2009) proposed a measurement model using coordination to index beginning-level MIS writers, subordination for intermediate, and phrasal elaboration for advanced. While most early complexity investigations relied on limited complexity indices and T-unit analyses (Crossley & McNamara, 2014), recent work has employed computational tools (e.g., the Coh-Metrix) to facilitate the measurement of syntactic complexity as a multidimensional construct. Biber and Gray (2013), Biber et al. (2016), and Kyle and Crossley (2018), for instance, argue for using multivariate analyses and phrasal embedding (e.g., complex nominal) as more accurate measures of advanced writing than traditional holistic measures based on T-units or clause embedding. The L2SCA (see Lu & Ai, 2012), a computerized tool, gauges five aspects of complexity through 14 standardized indices. Several

studies have examined the L2SCA as a reliable means of capturing the full trajectory of complexity for college MIS writers of multiple L1 backgrounds (Ai & Lu, 2013; Lu & Ai, 2015; Yoon & Polio, 2017).

Like L2 proficiency, task, learner, and context-related factors also moderate levels of written syntactic complexity. Among these are time and planning conditions, topic and genre (Lu, 2010; Yoon & Polio, 2017), instructional setting (Ortega, 2015), cross-linguistic influences (Lu & Ai, 2015; Polat et al., 2020), and instructional modality (Ortega, 2015). Modality effects concern the impact of face-to-face (F2F) and computer-mediated platforms on written production. Asynchronous (time-delayed) writing can facilitate syntactically rich language, since MIS have time to compose and edit before publishing (Mancilla et al., 2017). For instance, Mancilla et al.'s (2017) corpus-comparison of NS' and MIS' online discussions showed that MIS' online writing is equivalent in length to NS' and relies more on coordination and phrasal sophistication than subordination. Few studies have explored how computer-mediated environments modulate written syntactic complexity. More work is needed to understand relationships among these variables (Ortega, 2015), especially since asynchronous discussions have become the main forums within blended and online courses even before the Covid 19 pandemic (Mancilla et al. 2017).

Finally, research has also examined relationships between syntactic sophistication and writing quality. Biber et al. (2017) found that TOEFL iBT scores on written independent tasks were moderate predictors of quality of language used in disciplinary writing, particularly in applied linguistics and engineering fields. In their study on human judgements of writing quality of offline essays, Bulté and Housen (2014) reported that, combined with lexical complexity, the mean length of noun phrase, proportion of simple sentences, and subclause ratio accounted for 45% of the variance in perceived writing quality. These findings partially coincide with Bulté and Housen's (2014), but only explained 9% of the variance in quality scores. Later, using the L2SCA, Yang et al. (2015) found that mean length of sentence and T-unit consistently correlated with writing scores across topics. Despite mixed findings, these studies demonstrate that L2 writing performance is correlated with several different syntactic features.

Methodology

Current Study

Earlier additional language acquisition studies on syntactic complexity have focused on its value as an assessment criterion for measuring L2 learners' quality of writing or overall L2 level (Norris & Ortega 2009). Using a set of writing samples collected over seven years from asynchronous online discussions, this study examines the use of syntactic complexity and standardized proficiency test scores as

possible predictors of MIS' success in graduate school. It addresses these research questions:

1. Do MIS' overall proficiency and writing subtest scores (IELTS) predict their GPAs in graduate school?
2. Which syntactic complexity indices (if any) in online written production (subordination, coordination, length of production, and degree of phrasal sophistication) predict MIS' GPAs?
3. Which syntactic complexity indices remain predictive of MIS' GPAs after participants' English proficiency scores (overall and writing subscores) are inserted into the model?

Participants

The initial pool included 119 participants; however, due to incomplete data (e.g., GPA) from some participants, data from 7 participants were not included. Thus, participants were 112 (94 females, 18 males) MIS enrolled in a graduate program in ESL education at an urban university in the eastern United States. All participants had taken a graduate course on *additional language learning and teaching* between the years 2012 and 2018, in cohorts varying between 14 and 18 students. Most participants (mean age: 28) identified as Asian (76%), followed by White non-Hispanic (20%) and Black non-Hispanic (2%), while 2% marked "other" on the survey. Thirty-six percent of participants were from Mainland China, 40% from Saudi Arabia, and 24% from European nations.

Participants obtaining this degree (similar to a Master of Arts in Teaching English to Speakers of Other Languages [MA TESOL]) could become language instructors in intensive English programs in American universities or EFL teachers in international English preparatory programs. Besides additional admission requirements (e.g., personal statement, recommendation letters), all participants had an undergraduate GPA of at least 3.0 in addition to English proficiency test scores that met the university's (overall) minimum (IELTS=6.5) before they started the program. To fulfill basic graduation requirements, they completed 30 credit hours of graduate coursework. The program coursework was organized around content related to *language and linguistics, language acquisition, culture, L2 curriculum, instruction, assessment, and professionalism*.

Data Sources

Most participants submitted IELTS scores to fulfill the university's language proficiency admission requirement. Therefore, for convenience and uniformity purposes, only IELTS scores were used in this study. Note that most U.S. universities require 6.5 for master's programs, which is .5 higher than the 6.0 minimum that has been most recently reported as the strongest predictor of academic success in

English-medium universities (Schoepp, 2018). IELTS describes the three bands (scores) used in this study as:

- 6—Competent User: can use and understand fairly complex language, particularly in familiar situations.
- 7—Good User: generally handle complex language well and understand detailed reasoning.
- 8—Very Good User: handle complex and detailed argumentation well.

The overall IELTS scores used in our analyses varied between 6.5 and 8.5 (median=7), while the writing subscores varied between 5.5 and 8.5 (median=6.5). Due to the focus on written syntactic complexity, only overall (mean of reading, listening, writing, and speaking subscores) and writing subscores were used in the analyses.

The second data source included participants' 1st-year GPAs, which are averages of all grades from courses taught by multiple instructors in the program. We only used 1st-year GPA to control for possible effects of the program education (and various language socialization patterns) on participants' English proficiency and cumulative GPA (Ginther & Yan, 2017). First-year GPA is an often-used metric in research on the predictive validity of language proficiency assessments and student outcomes. GPAs were measured on a 4.0 scale and were obtained with participants' permission through the university's student academic services, which maintains all academic records. The GPAs used in our analyses varied between 2.45 and 4.00 (mean=3.1).

The dataset consisted of 224 writing samples (2 posts for each participant) gathered from asynchronous online discussions as part of a graduate course on L2 learning and teaching. To control for possible effects of English development during their graduate studies, samples were selected only from students' posts in the first 2 weeks of their first course in the program. The course was F2F and taught by the same professor to different student groups over 7 years. To capitalize on potential benefits of online learning, the course had an online component that required participation in weekly electronic reflections and two asynchronous online discussions. These discussions were solely for instructional purposes with no plans to use student data for research purposes. Therefore, these writing samples (Appendix A) represent formal yet natural utterances that are typical of online courses in graduate school. Upon deciding to use such data for research, these corpora were obtained through the university's Blackboard archives, in line with institutional review board (IRB) regulations. To ensure anonymity, these samples were de-identified and coded for analyses by the second author, who was not the course instructor. Each participant was given a pseudonym in an excel sheet, with five columns where their writing samples (2), test scores (2), and GPAs (1) were entered. To ensure accuracy across all data sources, the authors took turns to review and confirm the matches.

Students followed a set of specific guidelines (Appendix B) for discussion participation, which included the use of academic language. The only purpose of these online sessions was to learn the assigned academic content (not improve academic writing). Although they were required to post a minimum of three reflections on the content of weekly reading assignments as part of their participation in the online session, their reflections were not graded in terms of quality of language or content. Participants' first two reflections were included in the analyses to control for potential genre bias on levels of syntactic complexity. To qualify as a unit of analysis, a reflection had to comprise a minimum of five sentences (Lu, 2010; Yoon & Polio, 2017). Following a procedure similar to Stockwell and Harrington's (2003) for selecting L2 production units, participants' first and second posts were subjected to individual syntactic complexity analyses before a mean score was calculated for the combination of both posts. The total number of words (average of two posts for each participant) in the dataset was 15,344 ($M=137$; $SD=22$).

Syntactic Complexity Measures

Syntactic complexity indices were automatically generated by the L2SCA, a tool designed to analyze the written texts of university-level L2 users (Lu & Ai, 2012) like those in the present study. Due to its ability to apply a wide range of complexity measures at the sentential (global) and local (clausal) levels to large collections of L2 writing, the L2SCA is routinely employed in complexity research (e.g., Lu & Ai, 2015; Polat et al., 2020). The tool has been found to be reliable (Polio & Yoon, 2018) "in identifying relevant production units and syntactic structures from the essays as well as in computing the syntactic complexity indices" (Lu, 2010, p.491). For more information about reliability estimates for specific indices, please see Lu (2010). The L2SCA parses each sentence and computes the frequency of common sentential elements (e.g., T-units) to calculate 14 syntactic indices representing five domains of complexity —*length of production, amount of subordination, amount of coordination, degree of phrasal sophistication, and overall sentence complexity*. For better interpretability of the results, Lu's (2011) refined L2SCA model, which has the 10 most reliable indices across four domains of complexity, is used in this study (see Figure 1 for corresponding formulas and abbreviations from Ai & Lu, 2013).

Further, to ensure all 10 complexity indices did indeed constitute the four underlying structures of the L2SCA, we conducted an exploratory factor analysis (EFA) where we used the principal axis factoring (PAF) extraction method with a promax (oblique) rotation with Kaiser normalization (Preacher & Maccallum, 2003). To determine how many interpretable factors L2SCA measured (Appendix C), we considered all four common options: eigenvalues, the scree test, parallel analysis, and total variance explained (Mertler & Vannatta, 2010).

Further, considering methodological norms appropriate for our sample, as established in previous research, we only considered factors with pattern loadings greater than .50 as meaningful (Gorsuch, 1983; Guadagnoli & Velicer, 1988). In our analyses, the Barlett's Test of Sphericity ($p < .001$) and the Kaiser–Meyer–Olkin (KMO = .628) measures supported the factorability and sampling adequacy of our data (Mertler & Vannatta, 2010). Our results revealed that the four-factor structure conformed to L2SCA's presumed factor structure, with the four factors cumulatively accounting for 76% total variance. Results of our reliability tests for each factor revealed that all four factors ($F1 = .766$, $F2 = .751$, $F3 = .710$; $F4 = .770$) were within the commonly accepted ranges of .6 and .8. To ensure the reliability of our regression results, we also ran a variance inflation factor (VIF) test. Results revealed that all 10 measures of L2SCA were below 5 (range = 2.00–4.40), which is considered within an acceptable range for multicollinearity (Keith, 2014; Mertler & Vannatta, 2010). Based on these results, we included all 10 indices around the four measures in our regression analyses.

Data Analysis

First, we copied and pasted the two writing posts, for each participant separately, into the L2SCA software (Lu, 2010; Lu & Ai, 2012), which then generated a score for each syntactic complexity index that we used in three linear regression models using SPSS statistical software. The first two prediction models involved standard (simultaneous) multiple regression (the enter method) while the third was hierarchical. As a statistical procedure, standard multiple regression helps determine the value of an outcome variable based on the variability in the value of a set of independent variables. Hierarchical regression, a more complex analytic procedure, computes what one or more independent variables add to the prediction in variability of the dependent variable when they are inserted into the model (Keith, 2014).

This model included two sequential blocks. Into Block 1 we inserted the overall proficiency and writing subscores as the independent and the GPAs as the dependent variable, whereas in Block 2, the 10 complexity measures were added into the model as the primary variables of interest to determine their predictive power beyond the language proficiency scores. We also considered demographic variables, "country" and "gender," as part of our analyses. Although we excluded gender due to statistical power, we used "country" as a control variable to validate the relationship between our dependent and independent variables. In reporting the practical significance of our findings, instead of r -squared, we use adjusted r -squared, which is a more conservative measure adjusted for sample size and the number of predictors in the model (Keith, 2014).

TABLE 1

Descriptive Statistics for the GPA, Overall Proficiency, Writing Subscores, and Syntactic Complexity Measures

	<i>M</i>	<i>SD</i>	<i>N</i>
GPA	3.10	0.607	112
OverallS	7.01	0.575	112
WritingS	6.54	0.916	112
MLS	21.9	4.20	112
MLT	19.8	4.15	112
MLC	10.6	2.41	112
DCC	0.42	0.119	112
DCT	0.85	0.390	112
TS	1.13	0.146	112
CPT	0.56	0.292	112
CPC	0.31	0.178	112
CNT	2.44	0.631	112
CNC	1.32	0.309	112

Note. OverallS=overall score; WritingS=writing score; MLS=mean length of sentence; MLT=mean length of T-unit; MLC=mean length of clause; DCC=dependent clause per clause; DCT=dependent clause per T-unit; TS=T-unit per sentence; CPT=coordinate phrases per T-unit; CPC=coordinate phrases per clause; CNT=complex nominal per T-unit; CNC=complex nominal per clause.

Results

Overall Proficiency and Writing Subscores as Predictors of MIS' GPA

A standard multiple regression model was constructed to determine if overall proficiency and writing subscores (independent variables) predicted MIS' GPAs (dependent variable) in their graduate studies. The groups' GPA average was 3.10, ranging between 2.45 and 4.00 ($SD = 0.61$) with a letter-grade distribution between C– and A. The group overall IELTS mean was 7.01 ($SD = 0.58$) and writing subscores mean was 6.54 ($SD = 0.92$). For more descriptive statistics, see Table 1. Preliminary analyses revealed no violations of normality, linearity, homoscedasticity, tolerance, and VIF assumptions (Keith, 2014).

The prediction model was statistically significant, $F(2,109) = 52.46$, $p < .01$, suggesting that 48% of the variance (adjusted $R^2 = .48$) in students' GPAs was explained by the linear combination of their overall proficiency and writing subscores. To determine which of the two proficiency measures more strongly predicted students' GPAs, we examined standardized coefficients. The pairwise comparisons between the two language proficiency measures and the GPA indicated that while the MIS' levels of overall proficiency scores were moderately correlated with their GPAs ($\beta = .55$, $p < .01$), their writing subscores were stronger predictors ($\beta = .70$, $p < .01$).

We also ran post-hoc regression tests to determine which cutoff scores were stronger predictors of GPA. Considering sampling adequacy (only two bands had over 30 students: 36 in 6.5 and 48 in 7.0), we were able to run these tests only for 6.5 and 7.0 bands. Results revealed that for the overall scores, band 6.5 ($\beta = .77, p = .01$) was a much stronger predictor of GPA than band 7.0 ($\beta = .20, p < .05$). Findings were similar for the writing sub-skills, with band 6.5 ($\beta = .81, p < .05$) being even a stronger predictor of GPA than band 7.0 ($\beta = .23, p < .05$).

Aspects of Online Written Syntactic Complexity as Predictors of MIS' GPA

To address this research question, students' GPAs as the dependent and the 10 syntactic complexity measures as the independent variables were inserted into a multiple regression model. The predictor variables included three indices of *length of production unit* (mean length of clause, mean length of sentence, and mean length of T-unit), two indices of *amount of subordination* (dependent clauses per clause, dependent clauses per T-unit), three indices of *amount of coordination* (coordinate phrases per clause, coordinate phrases per T-unit, T-unit per sentence), and two indices of *degree of phrasal sophistication* (complex nominals per clause, complex nominals per T-unit). Once again, assumptions of normality, linearity, homoscedasticity, tolerance, and VIF were all satisfied (Keith, 2014).

The adjusted *r*-square values (Table 2) indicated that these 10 variables explained approximately 25% of the variance in these graduate students' GPAs. Findings indicated that this combination of the syntactic complexity indices accounted for a significant amount of variability in MIS' GPAs, $F(2,109) = 4.66, p < .01$. As presented in Table 2, results revealed that of the 10 indices only *mean length of sentence* ($\beta = .39, p < .05$) and *mean length of T-unit* ($\beta = .34, p < .05$) were significant predictors of students' GPAs. Given this finding and the size of intercorrelations amongst the syntactic complexity measures (Table 3), the unique variability accounted for by each of these variables seemed rather low.

Syntactic Complexity Indices Remaining Predictive of GPA When Other Scores Are Added into the Model

A hierarchical multiple regression model was constructed to examine which indices still remained predictive of MIS' GPAs when the L2 proficiency scores were included into the model. This model included two sequential blocks. In Block 1, the two language proficiency scores were entered as the independent and the GPAs as the dependent variable; whereas in Block 2, the 10 complexity measures and one demographic variable (country) were added into the model as the primary variables to determine their predictive power independent of the test scores.

TABLE 2

Linear Contribution of the Syntactic Complexity Measures to GPA

Predictors of GPA	Beta	<i>t</i>	R^2	R^2 Adj.	95% confidence intervals	
					Lower	Upper
			.32**	.25**		
MLS	.394*	2.40			.231	.532
MLT	.336*	2.23			.210	.472
MLC	-.117	-0.390				
DCC	-.090	-0.454				
DCT	.026	0.083				
TS	-.028	-0.240				
CPT	-.192	-1.37				
CPC	-.177	-1.31				
CNT	-.111	-1.64				
CNC	.282	1.27				

Note. MLS=mean length of sentence; MLT=mean length of T-unit; MLC=mean length of clause; DCC=dependent clause per clause; DCT=dependent clause per T-unit; TS=T-unit per sentence; CPT=coordinate phrases per T-unit; CPC=coordinate phrases per clause; CNT=complex nominal per T-unit; CNC=complex nominal per clause.
* $p < .05$. ** $p < .01$.

Results indicated that both of these tests were statistically significant (Block 1: $F[2,109] = 52.46, p < .01$; Block 2: $F[13,109] = 11.35, p < .01$). While the linear combination of the two proficiency scores in Block 1 accounted for 48% of variance in students' GPAs, with the addition of the syntactic complexity measures (Block 2), this amount increased to 53% (Table 4). Undoubtedly, the most important result of this hierarchical regression model involves the examination of R^2 change values to determine the unique contributions of the syntactic complexity measures to the prediction of GPA. Results indicated that the composite of the 10 syntactic complexity indices made a statistically significant contribution to the model's predictive capacity, explaining an additional 5% of variance in GPAs (R^2 change = .05; $F[13,99] = 11.45, p < .05$). As for the unique contribution of each predictor, findings revealed that two complexity indices (MLS: $\beta = .31, p < .05$; and MLT: $\beta = .32, p < .05$) and the writing subscore ($\beta = .64, p < .01$) remained significant predictors of GPA (Table 4), while the overall score did not. Results also suggested that the demographic factor "country" did not have a confounding effect on any of the relationships between our dependent and independent variables.

Discussion

Findings confirmed that the English composite and writing subscores explained a little under 50% of MIS' GPAs, while writing scores were stronger predictors. Data also suggested that syntactic complexity predicted (25% variance)

TABLE 3
Intercorrelations Among the Syntactic Complexity Measures

	MLS	MLT	MLC	DCC	DCT	TS	CPT	CPC	CNT	CNC
MLS										
MLT	.68**									
MLC	.19*	.24**								
DCC	.42**	.37**	-.37**							
DCT	.46**	.47**	-.41**	.69**						
TS	.21*	-.25**	-.10	-.12	-.18					
CPT	.15	.18*	.23*	-.07	-.04	-.21*				
CPC	.18*	.22*	.48**	-.24**	-.24**	-.16	.69**			
CNT	.55**	.68**	.22*	.27**	.41**	-.26**	.28**	.26**		
CNC	.21*	.28**	.67**	-.30**	-.34**	-.10	.24**	.45**	.63**	
Cntry	.14	.14	.10	-.13	-.08	-.03	.18	.15	.17	.16

Note. MLS=mean length of sentence; MLT=mean length of T-unit; MLC=mean length of clause; DCC=dependent clause per clause; DCT=dependent clause per T-unit; TS=T-unit per sentence; CPT = coordinate phrases per T-unit; CPC = coordinate phrases per clause; CNT=complex nominal per T-unit; CNC=complex nominal per clause.
 * $p < .05$. ** $p < .01$.

TABLE 4
Hierarchical Linear Contributions of the Syntactic Complexity Measures and Language Proficiency Variables to GPA

Predictors of GPA	Beta	<i>t</i>	R^2	R^2 adj.	R^2 change	95% confidence intervals	
						Lower	Upper
Model 1			.49**	.48**			
Model 2			.58**	.53**	.05*		
OverallS	.223	184					
WritingS	.464	3.83**				.306	.599
MLS	.310	212*				.191	.427
MLT	.327	2.30*				.206	.443
MLC	.023	0.193					
DCC	-.060	-0.383					
DCT	.113	0.450					
TS	.050	0.526					
CPT	.137	1.63					
CPC	-.016	-0.149					
CNT	.154	0.645					
CNC	.150	0.626					

Note. OverallS=overall score; WritingS=writing score; MLS=mean length of sentence; MLT=mean length of T-unit; MLC=mean length of clause; DCC=dependent clause per clause; DCT=dependent clause per T-unit; TS=T-unit per sentence; CPT=coordinate phrases per T-unit; CPC=coordinate phrases per clause; CNT=complex nominal per T-unit; CNC=complex nominal per clause.
 * $p < .05$. ** $p < .01$.

students' GPAs. Only two measures of complexity independently predicted GPA, both relating to length of production. After accounting for students' global and written language proficiency, the complexity measures remained predictive of academic outcomes, explaining a considerable amount of variance in GPA, even beyond that of proficiency exams. The best predictive model explained over half the variance in GPA, primarily due to the writing scores and the two complexity indices (MLS and MLT). Below, we discuss our

results, cautiously interpreting their potential contributions to the field considering relevant research.

Standardized English Proficiency Scores and Academic Achievement

Our study provides evidence that MIS' scores on English proficiency examinations partially predicted their future success in graduate studies. Both the composite and writing

subscores yielded positive correlations with GPAs. This pattern is consistent with recent validity research that has shown TOEFL (Bridgeman et al., 2016; Cho & Bridgeman, 2012) and IELTS composite scores (Thorpe et al., 2017; Wang-Taylor & Daller, 2014) are fairly predictive of GPA for similar populations. Further, our findings also help determine that, both at the overall and writing subskill levels, band 6.5 may be a stronger predictor of GPA than band 7.0 in language intensive graduate programs. It is possible that 6.5 is the threshold point after which the test items and writing ratings (human raters) become less predictive, or this result is due to the unique characteristics of the sample population, and thus should not be generalized to other settings. Similar to previous research, we analyzed scores from the web-based exams, which may explain why our findings bear some similarities to them, yet digress from early research reporting null or negative relationships among these same variables (Dooley & Oliver, 2002; Kerstjens & Nery, 2000).

Regarding the magnitude of the relationship between overall language proficiency and academic performance, our findings displayed a significantly stronger correlation than previously documented in validity research. We observed moderate correlations between composite scores and GPAs ($r = .55$), greatly exceeding effect sizes reported for some other graduate-level populations (e.g., $r = .18$ in Cho & Bridgeman, 2012; $r = .28$ in Oliver et al., 2012; and $r = .30$ in Wang-Taylor & Daller, 2014). Combined with writing subscores, these two predictors explained almost half the variance in students' GPAs (48%), which is substantially more than the maximum variance reported in some previous TOEFL or IELTS investigations (e.g., 29% in Hill et al., 1999). This difference is likely due to the homogeneity of the sample, which represented one academic discipline, contrary to heterogeneous samples used in other studies. Pooling effects have been used to explain increases in correlation coefficients that occur when heterogeneous groups sampled together are segregated (Hassler & Thadewald, 2003). Indeed, such effects were documented by Bridgeman et al. (2015), where some correlations doubled when examined by discipline. It is also plausible that standardized exam scores were better predictors of academic performance for this specific sample of MIS given their chosen career path into ESL education and the nature of their coursework (e.g., many writing tasks).

Moreover, note that because over 70% of our participants were from China and Saudi Arabia, some of these results could have been mitigated by the particularities of the testing systems and English language education in these countries, and/or sociocultural differences reported to influence MIS's performance on such tests (Thorpe et al., 2017). The question of population validity (Shulman, 1970) as it pertains to sociocultural differences or graduate students in ESL/applied linguistics programs is one meriting further attention, especially considering (Arrigoni & Clark, 2015)

negligible findings between IELTS scores and graduate educators' GPAs. Either way, as research on graduate MIS' (particularly from different regions) academic success in applied linguistics programs is markedly meager, this result can now be used as a basis for future cross-disciplinary comparisons about the power of standardized tests and syntactic complexity indices in predicting academic achievement.

Finally, our data revealed that the writing subtest is a moderate predictor of graduate MIS' GPA. This was somewhat unexpected because it not only contradicts the nonsignificant relationship (Oliver et al., 2010) or relatively weak correlations between writing and GPAs reported in some validity studies (e.g., Bridgeman et al., 2015; Kerstjens & Nery, 2000), but also surpasses the moderate correlation reported by Arrigoni and Clark (2015). Generally, the positive predictive relationship between writing and achievement aligns with the differing demands of curricula, where graduate work includes intensive, discipline-specific forms of writing, like theses. Further, given previous research, education courses rely more heavily on writing skills than other disciplines (Thorpe et al., 2017). This result should be interpreted with caution, though, since, besides being calculated separately, the writing subscores are also used in the calculation of composite scores. We can perhaps speculate that writing emerged as an independent, stronger predictor of achievement vis-à-vis students' participation in a hybrid learning ecology, where greater emphasis is placed on written output.

Written Syntactic Complexity and Academic Achievement

Complexity research has mostly sought to describe L2 learners' performance, rather than predict it. Turning to the most closely related research on modeling syntactic complexity indices to predict writing quality, it is possible to observe some patterns. For instance, a fairly comparable amount of variance was accounted for by Crossley and McNamara (2014), who explained 20.8% of the variability in quality ratings for their examination set of essays and 31.6% for their full sample using three of seven syntactic indices. Yang et al. (2015) reported three of 14 indices that explained 9% of the variance in ratings of essay quality. Despite differences in the number of indices entered into these models and ours, it is promising to observe the potential of complexity scores to predict aspects of MIS' academic performance beyond writing quality.

As for contributions of individual aspects of complexity, the data revealed few salient predictors. Of the 10 indices entered into the statistical model, only 2 were modestly correlated with academic outcomes, both of which referenced students' length of production: mean length of sentence (MLS, $r = .39$) and mean length of T-unit (MLT, $r = .34$). These findings corroborate earlier findings by Lu (2011), who reported a significant relationship between these two

indices and a linear progression in L2 proficiency. These metrics are known to index global, rather than local levels of sentential complexity (Norris & Ortega, 2009), and lend support to the existing research on the syntactic indices that reliably predict perceptions of L2 writing quality. For example, Bulté and Housen (2014) documented strong correlations between high ratings of text quality and MLS ($r = .41$) and MLT ($r = .40$), along with measures of clausal elaboration and embedding.

Using a similar model to ours, Yang et al. (2015) observed consistent correlations between MLS and MLT and the values attributed to MIS' writing samples, concluding that "either of the two global measures can work well as a generic syntactic complexity measure" (p. 63). Although their final prediction model was driven by clausal-level indices, this is unsurprising given the impact of topic and genre on the syntactic structures deployed by writers (Crossley & McNamara, 2014). Indeed, some of the differences between our findings and results of some other studies cited here could be attributed to the nature of task or genre that involve more or less complex uses of certain syntactic structures (Yoon & Polio, 2017).

Moreover, our findings regarding the two global measures of length of production coincide with many of the prominent features of MIS' asynchronous writing, specifically the elongation of texts and the use of subordinators (Polat et al., 2020). Indeed, these 10 indices correlated with each other at significant levels varying from .18 to .69 (Table 3). In fact, of the 45 bivariate correlations amongst these indices, only 8 were not significant. Despite being a reliable measurement tool (Lu, 2010; Polio & Yoon, 2018), it is hard to interpret why, in the L2SCA model, length of production is a predictor of academic success, and others, like phrasal complexity (nominal), found to be a moderate predictor of academic writing (Biber et al., 2016; Kyle & Crossley, 2018), are not. One possible explanation for this relates to the genre of samples because noun phrase elaboration is expected in academic writing, but not in less formal genres (Biber et al., 2016; Yoon & Polio, 2017). Thus, despite engaging in academic writing, partially because it was asynchronous (and untimed) and in online forums, students may have perceived the activity as less formal. Yet expectedly, when analyses are performed on the same data sample for numerous variables, confounding effects are inevitable, which calls for caution in the interpretation of such results.

Arguably, the most important finding of this study comes from the hierarchical regression analyses. Both models addressing the respective contributions of standardized test scores (Model 1) and written syntactic complexity (Model 2) to GPA yielded statistical significance, reinforcing the notion that unique assessment information is provided by each of these measures. By examining the R^2 change value, these analyses revealed an additional 5% of the variance that was explained by including complexity indices as predictors of

GPA. Together with language proficiency scores, the final model explained 53% of the variance in students' GPAs, a significant improvement over using only standardized test scores. Considering the contributions of alternative predictors, this is a sizable amount of variance explained. Though limited validity literature has employed this hierarchical method of analysis, our results regarding L2 proficiency are congruent with a handful of studies that have demonstrated the independent explanatory power of TOEFL composite and component scores when modeled with other admissions criteria (e.g., GMAT, GRE) (e.g., Cho & Bridgeman, 2012). For example, by combining all GMAT scores and undergraduate GPAs, Morris and Maxey (2014) accounted for around 15% of students' graduate achievement.

Conclusion and Implications

This study has explored whether standardized overall and writing subtest scores and computational measures of written syntactic complexity can predict graduate-level academic performance. Using linear and hierarchical regression models, our analyses addressed the predictive power of these variables, demonstrating complex relationships among 10 syntactic complexity indices and MIS' GPAs. We believe there are theoretical and practical implications of this work concerning the predictive power of the tests and syntactic complexity indices and use of these norms in university admissions.

Consistent with previous research on the internet-based IELTS exams, our findings confirm that composite English scores (with band 6.5 being a stronger predictor than 7.0) are somewhat predictive of cumulative academic performance. Contrary to many prior studies on exam subtests, we found writing was a stronger predictor of achievement (with band 6.5 being a stronger predictor than 7.0) than overall proficiency scores. It is hard to explain this result, which could be due to the characteristics of the sample population or the discriminant power of test items and the writing ratings around 6.5. level. Regardless, this result should be considered for focused exploration in future studies. Along with the MLS and MLT measures, writing was the other variable that independently contributed to the final regression model as a statistically significant predictor of GPA. This warrants further investigation and could have implications for the use of standardized test scores in university admissions. Specifically, MIS are usually granted admission based on the assumption that their proficiency is represented by the composite scores rather than the individual subtests. This assumption is generally supported since overall scores are computed from the subtest scores. Yet our findings regarding the superior predictive relationship between writing and GPA calls this presumption into question, suggesting greater attention be paid to writing subscores, considering how emergent score

profiles on the subtests may impact students' academic success differently (Ginther & Yan, 2017).

Another conclusion is that written syntactic complexity scores are modestly correlated with academic outcomes. With our findings regarding higher correlation values between MLS, MLT, and GPA, future research may replicate our analyses using offline writing samples or other genres to determine if these indices remain predictive across writing modalities. Further, the hierarchical regression results evidence the incremental validity of these complexity indicators and their usefulness as additional predictors of GPA. Coupled with the composite and writing scores, the full range of complexity indicators accounted for an additional 5% of the variability in MIS' GPA for an explained total variance of 53%. The unique contribution of syntactic complexity scores greatly exceeds the incremental value reported for many popular admissions criteria, including the GRE and GMAT. However, many admission decision-makers have little knowledge of the nuances of language-related criteria beyond the minimum score required by their university (Ginther & Yan, 2017). Although still premature, written syntactic complexity scores may be incorporated (with further research) into the admissions process for MIS in several ways.

First, since standardized tests already require MIS to produce a writing sample that is rated by human raters, the same sample could also be analyzed through computational (free) tools, such as the L2SCA, and used along with writing subtest scores. At the very least, these results highlight the need for additional research examining the complexity indices as predictors. Second, especially in graduate programs that are writing intensive or language focused, application reviewers could use some predictive complexity scores in conjunction with writing scores from samples that the students would be asked to produce upon arrival at the university. Beyond compensating for the lack of predictive validity of test scores, this can also ensure that students' most current English proficiency is well captured, without any doubts about possible use of AI (e.g., ChatGPT) in personal statements and other application documents that are also considered in admissions. This could prove helpful, particularly with admission decisions for students who submit scores lower than required minimums (whether being 6.5 or 7) but fall within one standard deviation. Programs could also evaluate these samples to advise students and offer additional writing courses to help them succeed in their programs (Johnson & Tweedie, 2021).

There are some limitations that may restrict the generalizability of our results to other educational contexts. Importantly, our sample represents the case of one academic program within a U.S.-based institution. Considering our findings on the comparative magnitude of the correlations between overall proficiency and academic performance, further investigation of other academic disciplines and applied linguistics programs beyond the U.S. context are needed to

ascertain these observations. Moreover, our dataset represented naturalistic language production and lacked the size to support more nuanced multivariate analyses. We were unable to control for the effect of MIS' backgrounds and other potential intervening factors on the complexity measures (Lu & Ai, 2015). Future studies should also consider larger (stratified) samples that allow for the inclusion of certain demographic factors (e.g., gender, country) in the analyses. Lastly, we used 1st-year GPA as a proxy for academic success and acknowledge that a partial single metric cannot fully capture achievement in graduate school. Although GPA is a commonly used metric and is necessary for measurement purposes, future studies could use cumulative GPA in combination with other factors (e.g., program completion rates) to explore other aspects of MIS' academic performance. Future studies could also explore, through a between-group design, how similar or different the relationships between these variables are for students from whom English is their first language versus speakers who learned English as an additional language.

Finally, although domestic students are not required to submit language proficiency scores, there is documented evidence (Mancilla et al., 2017) that there is significant amount of variation (in syntactic complexity, use of lexical resources, coherence and cohesion, etc.) within that group as well. Thus, requiring these language proficiency tests for MIS is inequitable. This research could help critique and problematize the distinction between MIS' use of English and so-called "standard academic English" and the overall language expectations of graduate school from a critical language perspective (Matsumoto, 2022). Overall, despite its shortcomings, we believe that this study not only adds to the literature on the predictive capacity of standardized tests, but it is also the first study that explores the utility of syntactic complexity indices in conjunction with test scores as possible predictors of academic achievement of MIS.

Appendix A

Examples of Online Posts

Example 1

In the LeDoux article, I can see how the phonics approach of teaching reading is advantageous as it is very structured and helps decode words, but focusing on it too much can affect the whole reading experience of the learners. Whole language on the other hand, encourages inquiry-based learning, but can be "disorganized" as the teacher "goes with the flow" of the class' learning. Like on page 11 in LeDeoux, the most effective way of teaching reading would be for the "two methods to work together as part of one program." The issue becomes how phonics should be integrated into Whole Language classrooms. I could see this integration as using the phonics method to teach how to do

basic reading in the beginning, which can introduce new concepts, but this learning can only go as far as being stored in the short-term memory.

Example 2

I just want to share some information about “Phonics.” In teaching reading, there are two main schools of thought: phonics and whole language. Phonics involves explicitly teaching a student how to pronounce words by analyzing its individual components, whereas whole language relies heavily on reading comprehension and context to memorize the sounds of words. Whole language works well for some students who can pick up on these patterns intuitively; for less fortunate students, such as those suffering from dyslexia, a phonics approach works best. On this website, I will advocate a combination of both phonics instruction and whole language, as each has its drawbacks when used exclusively, but combined, these two theories will complement each other.

Appendix B

Discussion Board Guidelines

In the Syllabus:

Reflections are not summaries of the assigned readings. Reflections will help students go beyond pure knowledge and comprehension to analyze, evaluate, critique, synthesize, and personalize implications and applications of theories and research covered in the assigned readings.

Posted Within Blackboard:

Below is the information about our asynchronous discussion.

The discussion will begin on *(date and time)* and end on *(date and time)*. We have TWO forums. Make sure you use academic language and make references to the content of each assigned reading.

Here are some reminders:

1. Everyone must post THREE messages.
2. Everyone MUST refer to the assigned readings in TWO of the posts: one for each topic/forum.
3. Your THIRD (or more if you wish) messages can be responses to your classmates’ comments. **Avoid simply agreeing or disagreeing. Rather, you should explain how your own thinking differs from the thinking of the person whose post you are responding to.**
4. The length of each message should contain at minimum of 5 sentences (around 200 words).
5. Everyone MUST read all messages in the forums.
6. Everyone MUST post one message a DAY. You CANNOT post two or all three at once.

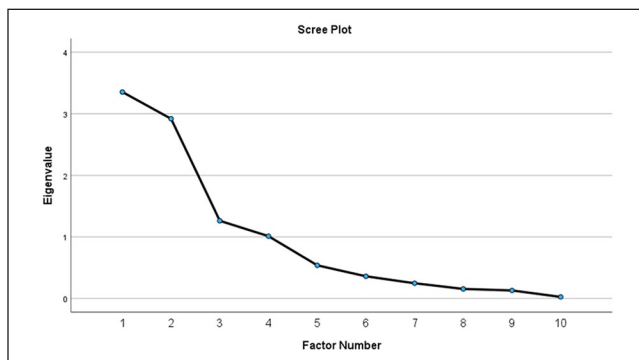
Appendix C

Factor Analysis Results

1. Factor Loadings for Exploratory Factor Analysis With Promax Rotation of the 10 Syntactic Complexity Indices Around Four Measures

Indices/ Factors	1: Amount of subordination	2: Length of production unit	3: Amount of coordination	4: Degree of phrasal sophistication
MeanDCT	.921			
MeanDCC	.813			
MeanMLS		.784		
MeanMLT		.668		
MeanMLC		.643		
MeanCPT			.769	
MeanTS			.729	
MeanCPC			.717	
MeanCNC				.906
MeanCNT				.714

2. Scree Plot for the Syntactic Complexity Indices



Acknowledgments

We would like to thank the participants, reviewers, and AERA Open’s editorial team for their contributions to this work.


Declaration of Conflicting Interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Nihat Polat  <https://orcid.org/0000-0002-6657-4985>

Laura Mahalingappa  <https://orcid.org/0000-0001-6644-5799>

Note

1. The Graduate Record Examination (GRE®) is a standardized test that is used in graduate school admissions worldwide. The Graduate Management Admission Test (GMAT) is a standardized test that is used in admissions into graduate management programs worldwide.

References

- Abunawas, M. (2014). *A meta-analytic investigation of the predictive validity of the Test of English as a Foreign Language (TOEFL) scores on GPA* [Unpublished doctoral dissertation]. Texas A&M.
- Ai, H., & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students writing. In N. Ballier, A. Díaz-Negrillo, & P. Thompson (Eds.), *Automatic treatment and analysis of learner corpus data* (pp. 249–264). John Benjamins.
- Arcuino, C. L. T. (2013). *The relationship between the Test of English as a Foreign Language (TOEFL), the International English Language Testing System (IELTS) scores and academic success of international master's students* [Unpublished doctoral dissertation]. Colorado State University. Retrieved March 15, 2024, from <https://www.proquest.com/dissertations-theses/relationship-between-test-english-as-foreign/docview/1413309058/se-2>
- Arrigoni, E., & Clark, V. (2015). *Investigating the appropriateness of IELTS cut-off scores for admissions and placement decisions at an English-medium university in Egypt* (IELTS Research Reports, 29). <https://bandscore.ielts.org/pdf/Arrigoni%20%20Clark%20FINAL.pdf>
- Aryadoust, V. (2023). Topic and accent coverage in a commercialized L2 listening test: Implications for test-takers' identity. *Applied Linguistics*, 1–20. <https://doi.org/10.1093/applin/amad062>
- Barkhuizen, G., & Strauss, P. (2020). *Communicating identities*. Routledge.
- Biber, D., & Gray, B. (2013). Discourse characteristics of writing and speaking task types on the TOEFL iBT® Test: A lexico-grammatical analysis. *ETS Research Report Series, 2013*(1), i–128. <https://dx.doi.org/10.1002/j.2333-8504.2013.tb02311.x>
- Biber, D., Gray, B., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, 37, 639–668. <https://doi.org/10.1093/applin/amu059>
- Biber, D., Reppen, R., & Staples, S. (2017). Exploring the relationship between TOEFL iBT scores and disciplinary writing performance. *TESOL Quarterly*, 51, 948–960. <https://doi.org/10.1002/tesq.359>
- Bista, K., & Dagley, A. (2015). Higher education preparation and decision-making trends among international students. *College and University*, 90, 2–11.
- Bridgeman, B., Cho, Y., & DiPietro, S. (2016). Predicting grades from an English language assessment: The importance of peeling the onion. *Language Testing*, 33, 307–318. <https://doi.org/10.1177/0265532215583066>
- Bright, L. (2020, January 15). *IELTS vs. TOEFL: What's the difference?* Canada.com. Retrieved January 15, 2023, from <https://www.studyincanada.com/Discover/Article/1/5116/IELTS-vs-TOEFL:-What's-the-Difference?>
- British Council. (2019). IELTS grows to 3.5 million a year. Retrieved January 15, 2023, from <https://takeielts.britishcouncil.org/about/press/ielts-grows-three-half-million-year#:~:text=IELTS%20is%20the%20International%20English,taken%20in%20the%20last%20year>
- Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices*. Pearson.
- Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, 26, 42–65. <https://doi.org/10.1016/j.jslw.2014.09.005>
- Cadena, M. A., Amaya, C., Duan, D., Rico, C. A., García-Bayona, L., Blanco, A. T., Santana Agreda, Y., Rodríguez, G. J. V., Ceja, A., Martínez, V. G., Goldman, O. V., & Fernández, R. W. (2023). Insights and strategies for improving equity in graduate school admissions. *Cell*, 186(17), 3529–3547. <https://doi.org/10.1016/j.cell.2023.07.029>
- Chapelle, C. A. (2012). Validity argument for language assessment: the framework is simple. *Language Testing*, 29, 19–27. <https://doi.org/10.1177/0265532211417211>
- Cho, Y., & Bridgeman, B. (2012). Relationship of TOEFL iBT® scores to academic performance: Some evidence from American universities. *Language Testing*, 29, 421–442. <https://doi.org/10.1177/0265532211430368>
- Cotton, F., & Conrow, F. (1998). An investigation of the predictive validity of IELTS amongst a group of international students studying at the University of Tasmania. In S. Wood (Ed.), *IELTS Research Reports 1*, (pp. 1–72). IELTS Australia.
- Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*, 26, 66–79. <https://doi.org/10.1016/j.jslw.2014.09.006>
- Deygers, B., & Malone, M. E. (2019). Language assessment literacy in university admission policies, or the dialogue that isn't. *Language Testing*, 36(3), 347–368. <https://doi.org/10.1177/0265532219826390>
- Dooley, P., & Oliver, R. (2002). An investigation into the predictive validity of the IELTS test as an indicator of future academic success. *Prospect*, 17(1), 36–54.
- Flores, N., & Rosa, J. (2015). Undoing appropriateness: Raciolinguistic ideologies and language diversity in education. *Harvard Educational Review*, 85(2), 149–171. <https://doi.org/10.17763/0017-8055.85.2.149>
- Gagen, T. (2019). *The predictive validity of IELTS scores: A meta-analysis* [Doctoral dissertation]. University of Western Ontario.
- Ginther, A., & Elder, C. (2014). A comparative investigation into understandings and uses of the TOEFL iBT® Test, the International English Language Testing Service (Academic) test, and the Pearson Test of English for graduate admissions in the United States and Australia: A case study of two university contexts. *ETS Research Report Series, 2014*(2), 1–39. <https://doi.org/10.17763/0017-8055.85.2.149>
- Ginther, A., & Yan, X. (2017). Interpreting the relationships between TOEFL iBT scores and GPA: Language proficiency, policy, and profiles. *Language Testing*, 35, 271–295. <https://doi.org/10.1177/0265532217704010>
- Graham, J. G. (1987). English language proficiency and the prediction of academic success. *TESOL Quarterly*, 21, 505–521. <https://doi.org/10.2307/3586500>

- Hassler, U., & Thadewald, T. (2003). Nonsensical and biased correlation due to pooling heterogeneous samples. *The Statistician*, 52, 367–379. <https://doi.org/10.1111/1467-9884.00365>
- Hill, K., Storch, N., & Lynch, B. (1999). A comparison of IELTS and TOEFL as predictors of academic success. In R. Tulloh (Ed.), *International English Language Testing System (IELTS) Research Report 2*, (pp. 63–73). IELTS Australia.
- Ihlenfeldt, S. D., & Rios, J. A. (2023). A meta-analysis on the predictive validity of English language proficiency assessments for college admissions. *Language Testing*, 40(2), 276–299. <https://doi.org/10.1177/02655322221112364>
- Institute of International Education. (2022). *International Students*. Retrieved January 15, 2023, from <https://opendoorsdata.org/annual-release/international-students/>
- Israel, B., & Batalova, J. (2021). *International students in the United States*. Retrieved January 15, 2023, from <https://www.migrationpolicy.org/article/international-students-united-states-2020>
- Johnson, R. C., & Tweedie, M. G. (2021). “IELTS-out/TOEFL-out”: Is the end of general English for academic purposes near? Tertiary student achievement across standardized tests and general EAP. *Interchange*, 52(1), 101–113. <https://doi.org/10.1007/s10780-021-09416-6>
- Keith, T. Z. (2014). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling*. Routledge.
- Kerstjens, M., & Nery, C. (2000). Predictive validity in the IELTS test: A study of the relationship between IELTS scores and students’ subsequent academic performance. In R. Tulloh (Ed.), *International English Language Testing System (IELTS) Research Report 3*, (pp. 3–85). IELTS Australia.
- Kyle, K., & Crossley, S. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *Modern Language Journal*, 102, 333–349. <https://doi.org/10.1111/modl.12468>
- Laad, S., & Sharma, A. (2021). Global student mobility trends–2021 and beyond. *LEK Consulting Footer Utility Privacy Notice and Legal*. <https://www.lek.com/insights/ar/global-student-mobility-trends-2021-and-beyond> by LEK Consulting
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15, 474–496. <https://doi.org/10.1075/ijcl.15.4.02lu>
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers’ language development. *TESOL Quarterly*, 45, 36–62. <https://doi.org/10.5054/tq.2011.240859>
- Lu, X., & Ai, H. (2012). *Web-based interface to L2 Syntactic Complexity Analyzer*. Retrieved January, 2023, from <https://sites.psu.edu/xx113/l2sca/>
- Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, 29, 16–27. <https://doi.org/10.1016/j.jslw.2015.06.003>
- Mahalingappa, L., Kayi-Aydar, H., & Polat, N. (2021). Institutional and faculty readiness for teaching multilingual international students in university programs in the field of education. *TESOL Quarterly*, 55, 47–77. <https://doi.org/10.1002/tesq.3083>
- Mancilla, R., Polat, N., & Akcay, A. (2017). An investigation of native and non-native English speakers’ levels of written syntactic complexity in asynchronous online discussions. *Applied Linguistics*, 38, 112–134. <https://doi.org/10.1093/applin/amv012>
- Matsumoto, Y. (2022). Multilingual international students’ communicative practices in US university classrooms: Rethinking appropriate Englishes through English as a lingua franca perspectives. *Harvard Educational Review*, 92(4), 486–507. <https://doi.org/10.17763/1943-5045-92.4.486>
- Mazgutova, D., & Kormos, J. (2015). Syntactic and lexical development in an intensive English for Academic Purposes programme. *Journal of Second Language Writing*, 29, 3–15. <https://doi.org/10.1016/j.jslw.2015.06.004>
- Mertler, C. A., & Vannatta, R. A. (2010). *Advanced and multivariate statistical methods: Practical application and interpretation*. Pyrczak.
- Morris, M., & Maxey, S. (2014). The importance of English language competency in the academic success of international accounting students. *Journal of Education for Business*, 89, 178–185.
- Neal, M. E. (1998). The predictive validity of the GRE and TOEFL exams with GGPA as the criterion of graduate success for international graduate students in science and engineering. (ERIC Document Reproduction Service No. ED424294).
- Norris, J. M., & Ortega, L. (2009). Toward an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30, 555–578. <https://doi.org/10.1093/applin/amp044>
- O’Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences*, 43, 971–990. <https://doi.org/10.1016/j.paid.2007.03.017>
- Oliver, R., Vanderford, S., & Grote, E. (2012). Evidence of English language proficiency and academic achievement of non-English-speaking background students. *Higher Education Research & Development*, 31, 541–555. <https://doi.org/10.1080/07294360.2011.653958>
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492–518. <https://doi.org/10.1093/applin/24.4.492>
- Ortega, L. (2015). Syntactic complexity in L2 writing: Progress and expansion. *Journal of Second Language Writing*, 29, 82–94. <https://doi.org/10.1016/j.jslw.2015.06.008>
- Piller, I., & Bodis, A. (2024). Marking and unmarking the (non) native speaker through English language proficiency requirements for university admission. *Language in Society*, 53(1), 1–23. <https://doi.org/10.1017/S0047404522000689>
- Polat, N. (2016). *L2 learning, teaching, and assessment: a comprehensible input perspective*. Multilingual Matters.
- Polat, N., Mahalingappa, L., & Mancilla, R. (2020). Longitudinal growth trajectories of written syntactic complexity: The case of Turkish learners in an intensive English program. *Applied Linguistics*, 41, 688–711. <https://doi.org/10.1093/applin/amz034>
- Polio, C., & Yoon, H. (2018). The reliability and validity of automated tools for examining variation in syntactic complexity across genres. *International Journal of Applied Linguistics*, 28, 1–24.
- Posselt, J. R. (2016). *Inside graduate admissions: Merit, diversity, and faculty gatekeeping*. Harvard University Press.

- Posselt, J., Southern, D., Hernandez, T., Desir, S., Alleyne, F., & Miller, C. W. (2023). Redefining merit through new routines: Holistic admissions policy implementation in graduate education. *Educational Evaluation and Policy Analysis*, 01623737231201612. <https://doi.org/10.3102/01623737231201612>
- Pritasari, A., Reinaldo, H., & Watson, C. (2019) English-medium instruction in Asian business schools: A case study. *Journal of Multilingual and Multicultural Development*, 40, 1–13. <https://doi.org/10.1080/01434632.2018.1458855>
- Richards, J. C. (2010). Competence and performance in language teaching. *RELC Journal*, 41, 101–122. <https://doi.org/10.1177/0033688210372953>
- Sá, C. M., & Sabzalieva, E. (2018). The politics of the great brain race: public policy and international student recruitment in Australia, Canada, England and the USA. *High Education*, 75, 231–253. <https://doi.org/10.1007/s10734-017-0133-1>
- Schoepp, K. (2018). Predictive validity of the IELTS in an English as a medium of instruction environment. *Higher Education Quarterly*, 72, 271–285. <https://doi.org/10.1111/hequ.12163>
- Shulman, L. S. (1970). Reconstruction of educational research. *Review of Educational Research*, 40, 371–396. <https://doi.org/10.3102/00346543040003371>
- Stockwell, G., & Harrington, M. (2003). The incidental development of L2 proficiency in NS-NNS email interactions. *CALICO Journal*, 20, 337–360. <https://doi.org/10.1558/cj.v20i2.337-359>
- Sullivan, L. M., Velez, A. A., Longe, N., Larese, A. M., & Galea, S. (2022). Removing the Graduate Record Examination as an admissions requirement does not impact student success. *Public Health Review*, 43, 1605023. <https://doi.org/10.3389/phrs.2022.1605023>
- Thorpe, A., Snell, M., Davey-Evans, S., & Talman, R. (2017). Improving the academic performance of non-native English-speaking students: The contribution of pre-sessional English language programs. *Higher Education Quarterly*, 71, 5–32. <https://doi.org/10.1111/hequ.12109>
- Wang-Taylor, Y., & Daller, M. (2014). Predicting Chinese students' academic achievement in the UK. In J. Angouri, T. Harrison, S. Schnurr, & S. Wharton (Eds.), *Proceedings of the 47th Annual Meeting of the British Association for Applied Linguistics, Learning, Working and Communicating in a Global Context* (pp. 217–227). Coventry.
- Wolfram, W. (2023). Addressing linguistic inequality in higher education: A proactive model. *Dædalus*, 152(3), 36–51.
- Wongtrirat, R. (2010). *English language proficiency and academic achievement of international students: A meta-analysis* [Unpublished doctoral dissertation]. Old Dominion University. doi:10.25777/y7yt-m587
- Wood, S. (2022). The complete guide to the TOEFL test. *U.S. News and World Report*. <https://www.usnews.com/education/best-colleges/articles/the-complete-guide-to-the-toefl-test>
- Yang, W., Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing*, 28, 53–67. <https://doi.org/10.1016/j.jslw.2015.02.002>
- Yoon, H. J., & Polio, C. (2017). ESL students' linguistic development in two written genres. *TESOL Quarterly*, 51, 275–301. <https://doi.org/10.1002/tesq.296>

Authors

NIHAT POLAT is chair and professor of applied linguistics in the Department of Teaching and Learning, Policy and Leadership at the University of Maryland, College Park, MD, 20912; npolat@umd.edu. His research foci include teacher education, applied linguistics, and the education of under-served populations.

LAURA MAHALINGAPPA is an associate professor of applied linguistics and language education in the Department of Teaching and Learning, Policy and Leadership at the University of Maryland, College Park, MD, 20912; lmaha@umd.edu. Her research focuses on educator preparation for linguistically marginalized students and writing development.

RAE MANCILLA is assistant director of the Office of Online Learning for the School of Health and Rehabilitation Sciences at the University of Pittsburgh, Pittsburgh, PA, 15260; raemancilla@pitt.edu. Her research focuses on online learning, digital accessibility, and instructional design.