SECOND LANGUAGE STUDIES

Volume 42 Issue 1

TABLE OF CONTENTS

SECOND LANGUAGE STUDIES — ONE YEAR MORE

RICKEY LARKIN

Department of Second Language Studies, University of Hawai'i at Mānoa

Before we discuss this year's issue, we would like to pause for a moment of acknowledgement. Last month marked the 1-year anniversary of the wildfires that devastated the town of Lahaina on the island of Maui. Due in no small part to the diversion of water for the sake of tourism and agriculture, this disaster resulted in the loss of at least 102 lives (Lyte, 2024). Many of the graduating students of Lahainaluna High School Class of 2024 will enter the University of Hawai'i system this academic year.

We welcome them.

AWAY AGAIN

Now in its second year under an all-graduate production team, we present you with the 42nd Volume of *Second Language Studies.* We would like to extend our thanks to the submission authors, and for your attention while you read the articles within.

Movement does not cease. We would like to acknowledge the graduation of two of our previous board members, Dr. Ha Nguyen and Dr. Hitoshi Nishizawa, whose dissertation titles you will find within. We congratulate them on their achievements and newly acquired positions. Kristen Urada has also stepped down while she finishes the last phase of her dissertation; we wish her success.

In their place, we add three new members to the board: Milang Shin, Jue Wang, and Joonhee Kim. These three members have dedicated their time and effort into making this issue possible.

THE CURRENT ISSUE

Our first paper, authored by MA graduate Maggie McGehee, was awarded a Harry Whitten Prize for Scholarly Excellence. This research focused on assessing the relationship between English language proficiency (ELP) test scores and academic performance at the University of

Hawai‘i at Mānoa. It explored whether variations in academic success were influenced by different ELP tests used for admission. This study is pivotal for evaluating the validity of the Duolingo English Test (DET) within admissions processes and contributes to existing literature by exploring outcomes at a large, public university in the U.S., employing a broader range of ELP scores and incorporating newer tests. The study analyzed various indicators of student success, including GPA, academic probation rates, and withdrawal rates, and compared these metrics for students with differing ELP scores and those exempt from ELP requirements. The results offer valuable insights into the effective use of DET in conjunction with IELTS and TOEFL scores for admissions at UHM.

Our second paper, authored by MA graduate Sohyeon Lee, investigates the World-Class Instructional Design and Assessment (WIDA) Consortium and the Center for Applied Linguistics' short form assessment—the WIDA Screener—available in both print and online versions. This paper provides a detailed review of the WIDA Screener Online, evaluating its practicality, usability, and validity. Its aim is to affirm its effectiveness as a valuable tool for both students and educators.

Our final paper, by PhD candidate Michaela Nuesser, uses the concepts of linguistic relativity and conceptual transfer for the examination of categorization preferences of drawing- and painting actions between three groups: German monolinguals, English monolinguals, and German–English bilinguals. This is done using similarity judgment tasks containing critical triads of picture stimuli alongside filler triads. A discussion of the paper's relevance to the study of linguistic relativity and transfer is included.

That sums up this issue; please enjoy. If publication in *Second Language Studies* is appealing to you, we are open to a wide variety of submissions, including: empirical articles, theoretical papers, research proposals, language test reviews, pedagogical or research materials, scholarly interviews, and short essays. We thank you again for reading Volume 42 of *Second Language Studies*.

Mahalo nui,

Rickey Larkin, Chief Editor
Micah Mizukami, Review Editor
Joonhee Kim, Copy Editor
Milang Shin, Summary Editor

REFERENCES

Lyte, B. (2024, June 25). Lahaina wildfire death toll now at 102. *Honolulu Civil Beat*. https://www.civilbeat.org/2024/06/lahaina-wildfire-death-toll-now-at-102/

RELATIONSHIP BETWEEN ENGLISH PROFICIENCY TEST SCORES AND ACADEMIC SUCCESS AT THE UNIVERSITY OF HAWAIʻI AT MĀNOA

MAGGIE MCGEHEE

Department of Second Language Studies, University of Hawaiʻi at Mānoa

ABSTRACT

This study compared the relationship between English language proficiency (ELP) test scores and academic success at the University of Hawaiʻi at Mānoa (UHM) and evaluated whether academic outcomes differed for students who entered on the basis of different tests. Locally, this represents one step in evaluating the validity of using the Duolingo English Test (DET) in admissions decisions. More broadly, it fills a gap in the literature by examining outcomes in a new context (a large, public, less selective university in the US), including a newer test (DET), and covering a wider range of ELP scores than is typically represented in such research. In addition to GPA as an indicator of student success, this study considered proportions of students on academic probation or withdrawing in relation to test submitted, and also made comparisons to international students who were exempt from submitting an ELP score for admission. Further, it compared students admitted unconditionally with higher ELP scores, to those with lower ELP scores admitted contingent upon further English language instruction. Findings are relevant to discussing valid use of DET alongside IELTS and TOEFL in admissions at UHM, while incorporating academic outcomes indicators beyond score correlations with GPA.

Keywords: English language proficiency test, university admissions, Duolingo English Test, range restriction, academic success, academic outcomes measures

INTRODUCTION

The present study is one step in evaluating the criterion-related and predictive validity of using the Duolingo English Test (DET) in admissions decisions for international students at the University of Hawai‘i at Mānoa (UHM). DET is a relatively new English Language Proficiency (ELP) test, which became more widely accepted as part of university admissions decisions during the COVID-19 pandemic (Isbell & Kremmel, 2020). This project examined institutional data to explore relationships between ELP test scores and academic success, as represented by first year GPA. Moreover, it evaluated whether there are disparate outcomes in GPA, or in instances of withdrawing or receiving academic probation or warning, for students who entered on the basis of DET compared to those who submitted a different ELP test score or were exempt from submitting an ELP score. The following research questions guided the project:

1. What is the relationship between UHM students' English language proficiency test scores and academic success?
2. Do academic outcomes differ for students who entered on the basis of different English language proficiency tests or without such a score?

Informed by and in partial replication of Isaacs et al. (2023), this study explored how DET compares to more established tests, International English Language Testing System Academic Test (IELTS) or Test of English as a Foreign Language iBT (TOEFL), in terms of association with initial student success, but in a different context. Isaacs et al. (2023) reviewed DET scores and first year grades at a "very selective" (College Board, n.d.) Russell Group institution in the United Kingdom with over 50,000 students (40% international) (Isaacs et al., 2023) and a 22% undergraduate acceptance rate. The present study considered ELP tests operating at a "less selective" (College Board, n.d.) public university in the United States with 19,000 students (7% international) (University of Hawai‘i at Mānoa, n.d.) and a 73% acceptance rate. Findings contribute to the validity discussion of using DET in admissions at UHM specifically and at English-medium universities generally by examining a new setting, and add to the broader literature on ELP tests and academic success by: exploring a wider range of ELP scores than typically represented; including students who were exempt from ELP testing; and

incorporating withdrawal, probation, or warning as academic outcomes indicators in addition to GPA.

BACKGROUND

Before using a test to inform decisions, test developers, takers, and decision-makers ask whether the test is reliable and fair, and whether using the given test takers' scores for certain decisions is valid and just. Validity is not an inherent property of a test, but appertains to specific test takers and uses (Chapelle, 2012). In order to argue that a given use of a test score is valid, it is important to establish that a test takers' score relates to how successfully they are able to use language in the real-world Target Language Use (TLU) domain (Bachman, 2005; Kane, 2013). To explore whether scores reflect ability to use English for completing college-level work, reviewers may evaluate whether scores correlate well with other measures of academic English proficiency (criterion-related validity), and whether they can be extrapolated to predict academic success (predictive validity) (Stoynoff & Chapelle, 2005). This project contributes to the validity discussion on using DET for university admissions, informed by the literature on the predictive validity of high stakes ELP tests for academic success, and on current evidence for DET validity.

Predictive Validity of High Stakes English Proficiency Tests for Academic Success

Requiring college applicants to take ELP tests is based on the premise that scores provide a predictive measure of future success in an English language institution. For criterion-related validity, researchers calculate the correlation between test scores and a second representative criterion. If the correlation is large and positive, this suggests a student's score provides predictive evidence for how they may perform on the other criterion (Stoynoff & Chapelle, 2005). In this framework, what criteria are used to represent academic success?

Researchers have suggested GPA, degree completion, time to graduation, student satisfaction, or professional qualifications as criteria for evaluating academic success. Others propose considering a student's intellectual and inter- and intrapersonal behavior (Zimmermann et al., 2018). There are limitations to using GPA alone, since it does not account for differences in course and grading difficulty between subjects, instructors, levels, or institutions. Still, GPA remains the most commonly used proxy for student success (York et al., 2015).

Beyond framing academic success only as attaining a high GPA, it may be more immediately critical to focus on the opposite end of the scale and scrutinize factors associated with earning a very low GPA. That is, whether a student earns a 3.1 compared to a 3.9 GPA may not matter as much to that individual or the institution in the near or long term, as whether and how many students earn grades below a level that prevents them from persisting to graduation. The present study considered academic success in terms of the threshold GPAs required to maintain good academic standing, which in turn support students in persisting and ultimately completing their degree.

There are limitations to assuming linguistic ability alone predicts GPA and associated academic outcomes. Language proficiency has been shown to correlate more with first-semester or first-year GPA than overall cumulative GPA (Kuncel & Hezlett, 2007). This is logical, since students gain more language exposure over time, making their pre-matriculation proficiency less relevant. The evidence is mixed on predicting GPA from widely used ELP test scores (Daller & Phelan, 2013; Daller & Yixin, 2017), including IELTS (Dang & Dang, 2021; Dooey & Oliver, 2002; Hu & Trenkic, 2021; Schoepp, 2018; Shakibaei & Memari, 2019) and TOEFL (Bridgeman et al., 2016; Chapelle et al., 2008; Cho & Bridgeman, 2012; Ginther & Yan, 2018; Harsch et al., 2017; Zimmermann et al., 2018). Language is necessary to succeed academically, but myriad other factors affect academic performance, including motivation, age, background, intelligence, and major field of study (Dooey & Oliver, 2002). This is also true of native speakers of the language of instruction, who do not all earn perfect GPAs simply by virtue of L1 proficiency. The implication is that researchers do not expect to find especially strong correlations between ELP or other admission test scores and GPA, and this alone is not a criterion for whether using these scores is valid in admissions decisions. Such tests are used to demonstrate whether a student meets a necessary level of proficiency to support their success, rather than to make precisely predictive quantitative estimates of individuals' grades over many years in differing circumstances in and out of the classroom.

Accordingly, meta-analyses consistently report statistically significant but weak to moderate positive correlations between ELP test scores and GPA (Abunawas, 2014; Brown, 2017; Gagen, 2019). A 2017 dissertation evaluated TOEFL iBT and first-year GPA at the University of Hawai‘i, and found a small statistically significant correlation (Brown, 2017). The

most recent peer-reviewed meta-analysis of both TOEFL and IELTS reports only a weak positive correlation between test scores and GPA ($r = .23$, $p < .001$) (Ihlenfeldt & Rios, 2022). This latter study found no difference between the two exams, and no significant moderating variables. Still, reviewers report that TOEFL for example does indeed reflect academic English proficiency (Chapelle et al., 2008), and so it measures what it purports to measure, per Messick (1989). In turn, "even a small correlation might indicate a meaningful relationship between TOEFL iBT…and GPA" (Cho & Bridgeman, 2012).

In all of these studies, researchers recommend including calculations on meaningful sub-samples of participants, such as level and major field of study, country of origin, and subscores on different test sections (Bridgeman et al., 2016; Ginther & Yan, 2018). For example, in 2011, Purdue University initially calculated the predictive validity of TOEFL iBT total scores and first-year GPA as 0, using only overall scores and aggregating all students in a single calculation (Ginther & Yan, 2018). Yet when Ginther and Yan (2018) disaggregated students by subgroups, corrected for range restriction, and investigated subscores for listening, speaking, reading, and writing, the researchers found that the score profile of a single Chinese subgroup was largely responsible for dampening the university's initial correlation calculation. Subsequent changes in Purdue admissions policy excluded this subgroup, and negative correlations then disappeared.

Analyses must also consider the effect of range restriction in test scores and in GPA. In the above study, Purdue's cut score of 80 restricted the range of TOEFL iBT scores to the top 52% of the range of possible scores. Overall, "Data on GPA are only available for the admitted students; therefore, the relationship tends to be underestimated" (Ginther & Yan, 2018). That is, correlation coefficients may be low when comparing admissions test scores to GPA, because the students who are admitted and then earn a GPA by definition exceeded a high cut score on the test. Because validity coefficients are based on sample variance, "having little variance in scores will ultimately lead to a weaker correlation" (Cho & Bridgeman, 2012). As seen in the present study, graduate student GPAs may also exacerbate this effect of restricted range, as graduate programs typically require students to maintain a high GPA.

Evidence for DET Validity

As a newer test, empirical support for DET is still emerging, and there is a general need for more research, especially after the 2019 DET revision. DET was only recently developed, and even more recently accepted by more institutions. In many cases, acceptance was driven by the COVID-19 pandemic when universities sought new options for test administration, including at-home (Isbell & Kremmel, 2020; Wagner, 2020). At that time, there was "very little empirical evidence supporting [DET] use for university admission purposes," including few independent, peer-reviewed studies on the reliability and validity of the current version of the test (Wagner, 2020). Today, more such studies have been published (Isaacs et al., 2023; Isbell et al., 2023), but more are needed to contribute to the validity argument of using DET alongside other accepted ELP tests for university international admissions decisions.

Starting with reliability as a precondition for validity, an early study reported a .79 test-retest reliability coefficient for DET, when participants took the test again after two weeks (Ye, 2014). The author wrote that this value is especially high for a test-retest protocol with different items, compared to identical forms. A year later, DET reported high internal consistency, with a split half reliability coefficient of .96, and test-retest reliability coefficient of .84 for repeat tests taken within 30 days (Settles, 2016). It must be noted that these studies were both commissioned by Duolingo, and Settles is one of DET's developers.

If reliability is established, then reviewers can turn to whether DET correlates well with other established measures of English proficiency (criterion-related validity). For example, this could include comparing students' DET scores with IELTS or TOEFL. As with the reliability evidence above, most existing studies on this come from internal Duolingo Research Reports (DRR) rather than peer reviewed journals. Early DRR studies found a relatively low correlation between DET and TOEFL scores of .41 (Ishikawa et al., 2016). Later studies reported $r = .67$ (Ye, 2014) and .70-.71 (Brenzel & Settles, 2017). Another Duolingo study found a significant correlation ($r = .83$) between DET and IELTS scores, and from this correlation inferred typical US university admission cutoff scores and CEFR levels for DET (Bézy & Settles, 2015).

The revised DET released in 2019 now includes writing and speaking tasks. After this revision, Settles et al. (2020) asked participants to take DET and also submit their IELTS or TOEFL score. In a peer-reviewed study, they found an improved test/retest reliability of .85, and

correlation coefficients of $r = .77$ with TOEFL iBT and .85 with IELTS. Similarly, the TOEFL and IELTS have a .73 correlation to each other, so DET's relationship to IELTS and to TOEFL is comparable to the relationship between IELTS and TOEFL (Brenzel & Settles, 2017; Settles et al., 2020).

Isaacs et al. (2023) is the first published study correlating DET scores and GPA. The researchers analyzed the correlation between first year GPA and DET scores for 1,389 graduate and 492 undergraduate students entering University College London in Fall 2020. Overall, they found a small positive correlation between DET scores and first year GPA for graduate students (adjusted $r = .195$), but a small negative correlation for undergraduate students (adjusted $r = -.112$). Importantly, the same pattern held for students entering on the basis of IELTS and TOEFL, supporting criterion-related validity of DET compared to these two widely accepted tests. These findings lend mixed support for DET but also for IELTS and TOEFL predictive validity, in terms of how well the test scores relate to future academic outcomes as measured by GPA.

Frequent critiques of DET, especially in its earlier version, note that DET items do not appear to reflect realistic, authentic language use in academic contexts (Isbell & Kremmel, 2020; Wagner, 2020; Yao, 2023). This could challenge the content validity of using DET scores as measures of academic English usage. Regarding extrapolation to the TLU, Isbell et al. (2023) asked university stakeholders to rate DET test takers' speech samples in terms of comprehensibility and acceptability for English language use in an academic context. They found a strong correlation between speakers' DET scores (both overall and subscores) and the stakeholders' acceptability and comprehensibility ratings ($r \geq .74$-$.89$). This high accordance between DET score and speech acceptability ratings from university faculty, staff, and students begins to build support for considering DET as reflecting expectations for English language use in an academic setting, at least in the speaking skill.

The present study seeks to add to the validity discussion on DET use in college admissions decisions for students with English as an additional language. How strongly do students' DET scores correlate with success at UHM in terms of GPA? From the other angle, do students who enter with DET experience negative academic outcomes (such as dropping out or receiving a warning or probation) at a higher rate? Further, how do these students' outcomes

compare to their peers who submitted other widely accepted ELP tests, or who were exempt from submitting an ELP score?

METHODS

International Student Admissions and ELP Test Scores at UHM

Data were received from the UHM Undergraduate Admissions Office and Graduate Division for 988 international students entering UHM between Fall 2020 and Spring 2023, the full period when DET was accepted for both graduate (GR) and undergraduate (UG) admissions. 399 of these students submitted ELP test scores to meet admissions criteria for English proficiency (Table 1). Fall 2021 was the first semester that entering students submitted DET scores (two graduate, seven undergraduate). This makes sense, since most students would have applied for Fall 2020 admission or taken an ELP test for Spring 2021 applications by the time UHM began accepting DET during the COVID-19 pandemic. In this three-year period, the number of students submitting DET scores peaked for Fall 2022 entry.

Table 1
ELP tests used by international students admitted and entering UHM, Fall 2020-Spring2023

	DET			IELTS			TOEFL			Total all ELP			No ELP test**		
Entry	GR	UG	total	GR	UG	total	GR	UG	total	GR	UG	total	GR	UG	total
Fall 2020	0	0	0	25	0	25	33	7	40	58	7	65	74	72	146
Spring 2021	0	0	0	4	0	4	16	3	19	20	3	23	13	27	40
Fall 2021	2	7	9	22	6	28	61	12	73	85	25	110	76	96	172
Spring 2022	2	4	6	7	1	8	18	1	19	27	6	33	34	22	56
Fall 2022	18	30*	48	26	9	35	44	12	56	88	51	139	58	74	132
Spring 2023	2	2	4	8	2*	10	11	4	15	21	8	29	20	19	39
Total	24	43	67	92	18	110	183	39	222	299	100	399	275	310	585

* In the UG data, two Fall 2022 DET scores and two Spring 2023 IELTS scores were self-reported only. The Undergraduate Admissions office reports that it is possible for staff to verify a student's score and fail to enter the official score in the database, and that this may have been exacerbated during COVID-19 staff challenges.
** Excludes four GR students who entered in Summer with no ELP score (three in Summer 2021, one in 2022)

This dataset includes 585 students with no recorded ELP test score (excluding four graduate students who entered in Summer terms). Although students born outside of the U.S. must submit a measure of language proficiency, this evidence can include being a native speaker of English, living in an English-speaking country, or completing education in an English language institution (Office of Admissions, n.d.). The latter could include any of a wide range of educational experience, including completing the last three years of high school in the US, transferring 60 college credits from another US institution, or earning an associate degree at one of the seven University of Hawai‘i community colleges before entering UHM. Accordingly, either graduate or undergraduate students may be exempt from providing an ELP score.

Outside of these known reasons for lacking an ELP score, UHM staff cautioned that admissions staff may fail to enter some scores in the database after verifying them, and that this

may have been exacerbated during the unusually challenging workloads and circumstances during the pandemic. This may have created unknown gaps in score data, and also explains four undergraduate students who have ELP scores in a self-reported score field only that were not also entered in the official score field. These four scores (two DET in Fall 2022, two IELTS in Spring 2023 entry) are included in this study's analyses, as students who misrepresented their test scores would likely be denied admission.

Students required to submit an ELP test must meet or exceed a cut score to enter the university (Table 2), but UHM offers both unconditional admission and conditional admission.

Table 2

Required minimum English proficiency test scores for admission to UHM, Fall 2022, Graduate (GR) and Undergraduate (UG)

	Conditional		Unconditional	
Test	GR	UG	GR	UG
DET	95	90	125	125
IELTS Academic	6.0	5.0	7.0	7.0
TOEFL iBT	61	61	100	100

Applicants admitted conditionally report to the English Language Institute (ELI) for further language instruction (or exemption after placement testing). For unconditional admission, both graduate and undergraduate students must achieve at least 7.0 on IELTS Academic or 100 on TOEFL iBT. In order to be considered for a graduate assistantship, graduate students must further meet a minimum 25 subscore on Listening and on Speaking in TOEFL iBT. Individual graduate programs may set higher cut scores.

Starting in Fall 2020, students scoring 125 or higher on DET were offered unconditional admission, with conditional admission via the ELI for DET scores of 95 (graduate) or 90 (undergraduate) or higher. Effective Spring 2023, the Graduate Division no longer accepts DET, and now only accepts TOEFL (including iBT Home Edition) and IELTS Academic (including Online) (Graduate Division, n.d.-b). Undergraduate Admissions continues to accept DET for Fall

2024 entry, as well as the ACT (American College Testing), Cambridge English Test, English Language Proficiency Test (ELPT), GRE, IELTS, PTE-A, SAT, and TOEFL (iBT, PBT, or Essentials) (Office of Admissions, n.d.). Despite the many ELP test options for undergraduates, the overwhelming majority of students who submitted any of these scores submitted DET, IELTS, or TOEFL (94% of those with a test score on file who entered in Fall 2021, Spring 2022, or Fall 2022). Other ELP tests were excluded from further analysis because only three undergraduate records held an EIKEN score, six reported SAT only, and three reported ACT only, with no scores reported for Cambridge English Test, ELPT, or PTE-A.

This sample is subject to selection bias because it includes only students who were admitted to and entered the university. These participants by definition exceeded an ELP test cut score and also met other high academic requirements. As noted in the literature review, existing studies on ELP and other admission test scores and GPA regularly observe that this restriction of range may dampen correlation coefficients (Ginther & Yan, 2018; Ihlenfeldt & Rios, 2022; Ishikawa et al., 2016). For example, for admissions to University College London (UCL), as for participants in Isaacs et al. (2023), students must exceed one of five cut score levels depending on course of study (University College London, n.d.). The least restrictive of these levels requires a minimum score of 92 on TOEFL (44th percentile for graduate, 52nd percentile for undergraduate test takers (ETS, 2023)), or 115 on DET (68th percentile of test takers) (Cardwell et al., 2023). This means that the analyses in Isaacs et al. (2023) included only at most the highest half of the range of TOEFL test scores, and the highest 32% of the range of DET scores (unfortunately, IELTS does not publish similarly detailed information on score distributions, so it is unknown how IELTS scores relate to percentiles).

By comparison, UHM allows for conditional admission via ELI for students with scores as low as 61 on TOEFL; 6.0 graduate / 5.0 undergraduate on IELTS; and 95 graduate / 90 undergraduate on DET (see Table 2). In percentile terms, with these lower, conditional-admission cut scores, UHM only outright excludes students with the lowest 5% of graduate and 10% of undergraduate TOEFL scores (ETS, 2023), and only the lowest 20% of the distribution of DET test takers (Cardwell et al., 2023). The less-restricted range of ELP test scores in this student population may lead to stronger correlations, since it potentially includes students who score across 90-95% of the distribution of TOEFL test scores and those who earn scores

spanning 80% of the DET test score distribution. While Duolingo provides global mean and standard deviation information which could allow calculations to correct for range restriction, such a range correction is not necessary for this study because the range is not as restricted to begin with.

Study Cohorts and Demographics

In Fall 2022, 988 degree-seeking international students were enrolled at UHM (573 graduate, 415 undergraduate), including continuing students (International Student Services, n.d.). The most-represented home countries were Japan, China, South Korea, Canada, and Taiwan. These were also among the countries most represented in the Fall 2022 ELP score data (Table 3), with the unsurprising exception of Canada, which only had four students with ELP scores (perhaps hailing from Québec, from which ELP test scores are required).

Table 3

Top five home countries among UHM international students

	International students enrolled Fall 2022 (including continuing)			ELP tests taken by newly entering students from these countries, Fall 2020-Spring 2023			
Country	GR	UG	Total	DET	IELTS	TOEFL	Total
Japan	61	115	176	8	15	30	53
China	112	31	143	3	19	31	53
South Korea	69	33	102	2	5	25	32
Canada	40	57	97	2	0	2	4
Taiwan	29	10	39	4	5	7	16

While China, Korea, and Japan are the most represented countries among students submitting any ELP test score, for each ELP test individually, the most common countries vary. For DET test takers, after Japan, the second-most prevalent home country is the Philippines (n =

5). For IELTS, Bangladesh (n = 13) and Thailand (n = 6) are third and fourth, above Korea. For TOEFL, Germany (n = 11), Iran (n = 10), and Nepal (n = 9) precede Taiwan. Available data do not include information suggesting first language or other languages spoken, other than country of birth and citizenship.

This study focused on students who entered in Fall 2021, Spring 2022, and Fall 2022, as each of these cohorts reported DET scores and has had the opportunity to complete at least two semesters of study through Spring 2023. This enabled the researcher to identify students who withdrew or were placed on academic probation or warning, and permitted calculation of first year GPA across two terms. Analysis centered on the n = 282 (200 graduate, 82 undergraduate) who submitted ELP scores, with selected comparisons to the 364 (172 graduate, 192 undergraduate) with no ELP test score recorded. The 282 students with ELP test scores included 200 graduate (119 female, 80 male, 1 not reported, mean age = 29.02 years, SD = 7.09) and 82 undergraduate (54 female, 28 male, mean age = 19.48 years, SD = 2.28).

Additional analyses focused on the cohort entering in Fall 2022, as this was the first semester since March 2020 when mostly in-person instruction resumed at the university. Students in this cohort returned to a more typical semester in terms of instructional methods and experiences, compared to students entering during the peak of COVID-19 when most classes and services were online and campus was extensively disrupted. Of the 139 students with ELP scores entering in Fall 2022 (88 graduate, 51 undergraduate), 131 enrolled in both the Fall 2022 and Spring 2023 semester, enabling comparison of first year GPA. This study's final Fall 2022 cohort of 131 participants included 85 graduate students (51 female, 33 male, 1 not reported, mean age = 28.3 years, SD = 6.80) and 46 undergraduate (31 female, 15 male, mean age = 19.1 years, SD = 1.89).

Available Data Fields

The UHM Undergraduate Admissions Office and Graduate Division provided the following data for international students entering between Fall 2020 and Spring 2023. Not all of these were used in the final reported results, but are documented here:

1. *Demographics*: gender, age, country of birth, citizenship
2. *Academics*: term of entry, level (graduate or undergraduate), major field of study, degree program (i.e., BA, BS, PhD), and previous GPA submitted for admissions purposes
3. *Test scores*: overall scores and date taken for the most recent test date. Section scores were requested, but not available in most cases. Admissions staff reported that section scores are never recorded for undergraduate applicants. ELP tests included DET, IELTS, and TOEFL, as well as Cambridge English Test, EIKEN, ELPT, or PTE-A (although as described above, only DET, IELTS, and TOEFL appeared frequently enough in the data to be compared). The dataset also included scores from general admissions tests such as GRE, SAT, and ACT, but these were not directly compared to the ELP tests as they do not purport to measure the same construct.
4. *Grades*: for each semester, total credits attempted and earned GPA (out of 4.0). The dataset also included grades for required freshmen English composition courses (ENG 100 or ESL 100), and the number of courses undertaken that are designated as Writing Intensive (a UHM designation applicable for undergraduate courses only).
5. *Other*: graduate student data included Graduate Assistant Exemption status (Y or N) and English Language Institute (ELI) exemption code. This code indicates whether admissions officers determined the student should report to the ELI for further English language study, or were exempted due to living in an English-speaking country, holding a degree from an English medium institution, or exceeding UHM's higher ELP test cut scores for unconditional admission (see Table 2).

Variables Included in Statistical Analyses

This study investigated the relationship between DET, TOEFL, and IELTS scores and first year academic outcomes, and evaluated whether meaningful differences exist between those who submitted scores from different tests or were exempt. Analyses further considered possible differences by academic level or field of study. Accordingly, key continuous variables were GPA by semester and first full year, and ELP test scores for DET, IELTS, and TOEFL. First year GPA was calculated from credits attempted and grade points earned for each enrolled semester.

Key nominal variables for comparing subgroups included ELP test taken, academic level (graduate or undergraduate), and field of study. Following Isaacs et al. (2023), field of study was coded into three groups based on European Research Council classifications: life sciences, physical sciences and engineering, and social sciences and humanities (European Research Council, 2021). Students were also classified on their cohort based on entry term (Fall 2021, Spring 2022, or Fall 2022), and whether they were admitted unconditionally with an ELP score exempting them from further English language instruction, or conditionally through the ELI.

As part of the second research question regarding differential academic outcomes for students who entered with different ELP tests or none, in addition to comparing mean GPA, students were flagged if they withdrew or received an academic action (warning or probation) after their first or second semester. At UHM, graduate students are placed on probation when their GPA falls below 3.0 after earning eight credits (Graduate Division, n.d.-a). Undergraduate students receive an academic warning if they earn less than a 2.0 GPA in their first semester, and probation if their cumulative GPA is under 2.0 thereafter (Mānoa Advising Center, n.d.). This dichotomous flag provided another nominal variable, to compare proportions of students on academic action (warning, probation, or withdrawal) by ELP test taken. This is a coarser outcomes measure, but has more serious consequences than fractions of GPA points. It also in part addresses the ceiling effect in graduate student grades: because graduate students are required to maintain a 3.0 or higher GPA, graduate GPAs tend to cluster around 4.0, reducing the meaningfulness of correlation calculations on their face with this restricted range of GPAs.

Analyses

This study used an alpha of $p < .05$ for statistical significance, for comparison with the level utilized by other ELP test score ~ GPA correlation studies including Isaacs et al. (2023). All analyses were performed using R Statistical Software v4.3.1 (R Core Team, 2023).

First, correlations (Pearson's *r*) with 95% confidence intervals were calculated between ELP test scores and first year GPA, to examine the relationship between UHM students' English language proficiency test scores and academic success. Correlation assumes that the variables are continuous (which is true of both test scores and GPA), and that the two have a linear relationship. Scatter plots were created to check the linear relationship assumption. Correlations were then calculated for subgroups, to identify any differences in the associations. Subgroups included academic level (graduate or undergraduate), field of study (life sciences, physical sciences and engineering, or social sciences and humanities), and whether the student was admitted unconditionally or conditioned on further English language coursework via the ELI. Analyses combining all three cohorts (entering Fall 2021, Spring 2022, and Fall 2022) were compared with analyses for the Fall 2022 cohort only, to cross-check for effects that may have accrued during the COVID-impacted semesters.

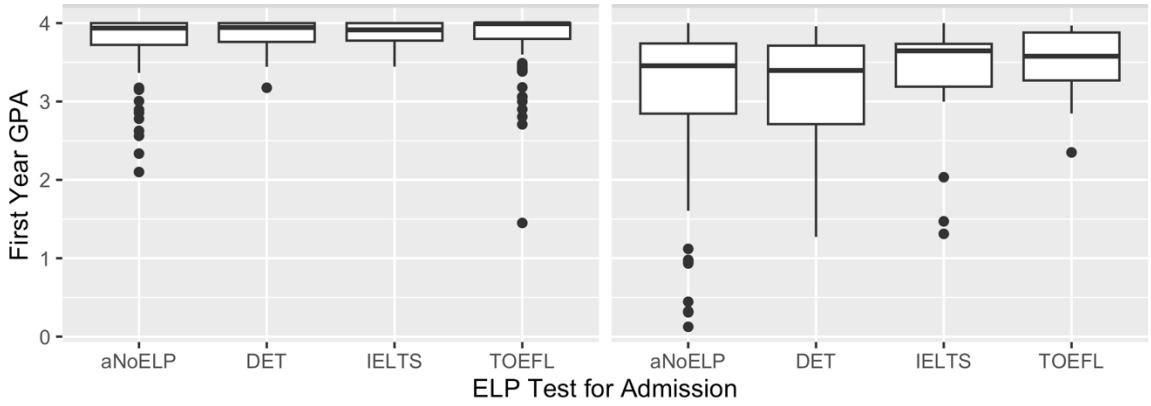
The second research question asked whether academic outcomes differ for international students who entered on the basis of different English language proficiency tests, or were exempt from submitting an ELP score, based on significant experience living in an English-speaking country or studying in an English medium institution. One-way ANOVA with post-hoc contrasts tests were used to check whether mean first year GPA differs significantly based on which of the three ELP tests (or no ELP score) students submitted (a four-level categorical variable). Next, chi square and difference of proportion tests were created to compare proportions of students who withdrew, or were placed on academic warning or probation, as a more serious indicator of whether there are meaningful differences in academic outcomes for those who had no ELP score or submitted DET, IELTS, TOEFL for admission during the same timeframe. Although ELP test scores alone were not expected to be especially powerful for predicting academic success, it would be instructive if any test bore a stronger relationship than others, or if those who entered on the basis of one ELP test had notably differently outcomes than other subgroups.

RESULTS

Initial data exploration revealed an accumulation of graduate GPAs near 4.0, reflecting the requirement for graduate students to maintain a minimum cumulative GPA of 3.0 (Figure 1).

Figure 1

Box and whiskers plot of mean first year GPA by ELP test submitted (or no ELP score), cohorts entering Fall 2021, Spring 2022, and Fall 2022, graduate and undergraduate disaggregated



In these cohorts, only graduate students entering with no ELP or TOEFL earned a first year GPA under 3.0. This plot also showed that for undergraduate students, first year GPA varied more for those entering with no ELP score or with a DET score on file, seen in the long whiskers and wider box spreads compared to IELTS and TOEFL. The no ELP score group had outliers with the lowest GPAs, under 1.0. These descriptive statistics informed subsequent decisions to conduct analyses separated by level (graduate and undergraduate), and provided context for evaluating correlations.

Correlations Between ELP Test Scores and First Year GPA: by Level

Overall and when grouped by academic level, ELP scores correlated moderately with first year GPA for the cohort entering in Fall 2022 (Table 4). All of the correlations showed a positive association (r = .30-.45), with the exception of graduate students who submitted DET scores (r = -.40). The correlations were statistically significant for IELTS scores for graduate students (p = .019), DET scores for undergraduate students (p = .021), and when both graduate and undergraduate students were combined for DET (p = .024) and for TOEFL (p = .001).

Table 4

Correlation between ELP test scores and first year GPA by level: Fall 2022 cohort

Level	ELP test	n	M	SD	min	max	r	95% CI LL	95% CI UL	p	R^2
All	DET	45	118	14.6	90	150	.34	.05	.57	.024*	.11
		GPA	3.49	0.54	1.91	4.00					
	IELTS	34	6.96	0.73	5.5	8.5	.32	-.01	.59	.057	.10
		GPA	3.69	0.56	1.31	4.00					
	TOEFL	52	92.2	19.4	30	117	.42	.17	.62	.001**	.18
		GPA	3.70	0.48	1.45	4.00					
GR	DET	17	121	12.8	95	140	-.40	-.74	.10	.112	.16
		GPA	3.84	0.23	3.18	4.00					
	IELTS	25	6.96	0.57	6.0	8.0	.45	.08	.71	.019*	.20
		GPA	3.88	0.13	3.65	4.00					
	TOEFL	43	96.6	15.2	64	117	.30	.00	.54	.052	.09
		GPA	3.73	0.49	1.45	4.00					
UG	DET	28	116	15.5	90	150	.44	.07	.70	.021*	.19
		GPA	3.29	0.57	1.91	3.96					
	IELTS	9	6.94	1.13	5.5	8.5	.43	-.32	.85	.243	.19
		GPA	3.16	0.90	1.31	3.97					
	TOEFL	9	74.5	24.6	30	107	.45	-.21	.83	.168	.20
		GPA	3.58	0.44	2.85	3.97					

Results were similar when calculating correlations for all three cohorts of students combined entering in Fall 2021, Spring 2022, and Fall 2022 (Table 5), with no remarkable differences compared to looking at the Fall 2022 cohort alone. With the combined cohorts, the negative correlation (non-statistically significant) for DET and graduate student GPA was weaker (from -.40 in all cohorts compared to -.05 for Fall 2022), and the non-statistically significant correlation for IELTS for all student levels combined was dampened from .32 to .10. The correlation for TOEFL and graduate student GPA became statistically significant (r = .27, p = .004). For undergraduate students, the correlation between DET and GPA was no longer statistically significant, while TOEFL became significantly correlated (r = .42, p = .040).

Table 5

Correlation between ELP test scores and first year GPA by level: Cohorts entering Fall 2021, Spring 2022, and Fall 2022

Level	ELP test	n	M	SD	min	max	r	95% CI LL	95% CI UL	p	R^2
All	DET	58	118	14.9	90	150	.29	.03	.51	.027*	.08
		GPA	3.42	0.63	1.27	4.00					
	IELTS	67	6.87	0.79	5.0	8.5	.10	-.14	.33	.413	.01
		GPA	3.73	0.50	1.31	4.00					
	TOEFL	134	93.4	15.9	30	117	.40	.25	.53	<.001***	.16
		GPA	3.77	0.39	1.45	4.00					
GR	DET	21	122	14.2	95	140	-.05	-.48	.39	.814	.00
		GPA	3.84	0.23	3.18	4.00					
	IELTS	52	6.89	0.66	5.0	8.5	.36	.10	.57	.008**	.13
		GPA	3.87	0.14	3.44	4.00					
	TOEFL	112	95.8	13.7	61	117	.27	0.09	.43	.004**	.07
		GPA	3.82	0.36	1.45	4.00					
UG	DET	37	115	15.0	90	150	.29	-.04	.56	.082	.08
		GPA	3.19	0.67	1.27	3.96					
	IELTS	15	6.8	1.16	5	8.5	.02	-.50	.52	.956	.00
		GPA	3.23	0.89	1.31	4.00					
	TOEFL	22	82.1	20.7	30	109	.42	.02	.71	.040*	.18
		GPA	3.51	0.42	2.35	3.97					

Figures 2 and 3 provide scatterplots for GPA and ELP test scores for the Fall 2022 cohort and for the three cohorts combined. Scatterplots with separate fit lines by level, showing the differences that emerge when disaggregating graduate and undergraduate, are included in Figures 4 and 5.

Figure 2

ELP test scores and first year GPA, all levels (graduate and undergraduate combined): Fall 2022 cohort

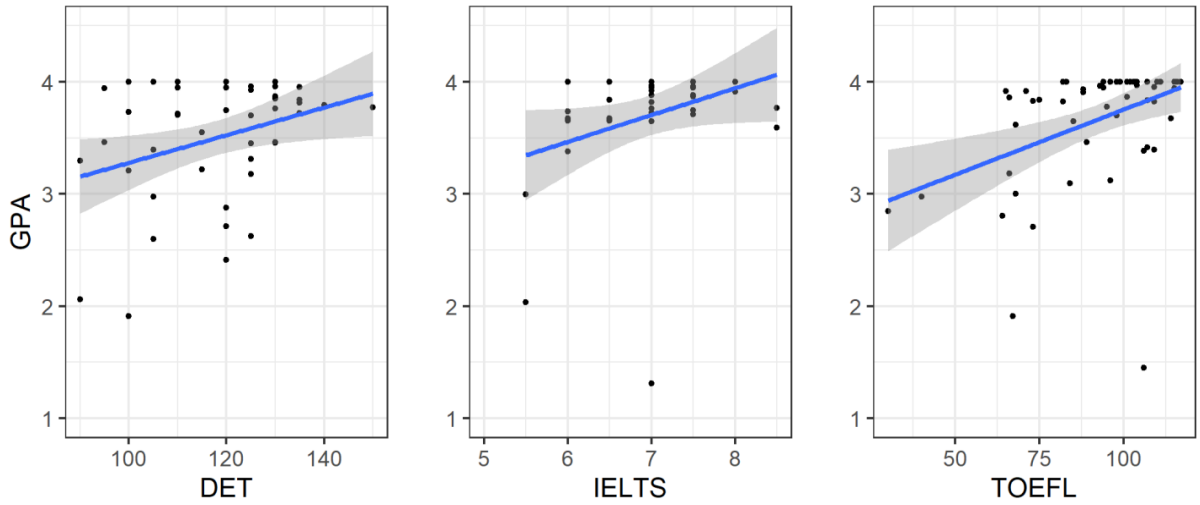


Figure 3

ELP test scores and first year GPA, all levels (graduate and undergraduate combined): Combined cohorts entering Fall 2021, Spring 2022, and Fall 2022

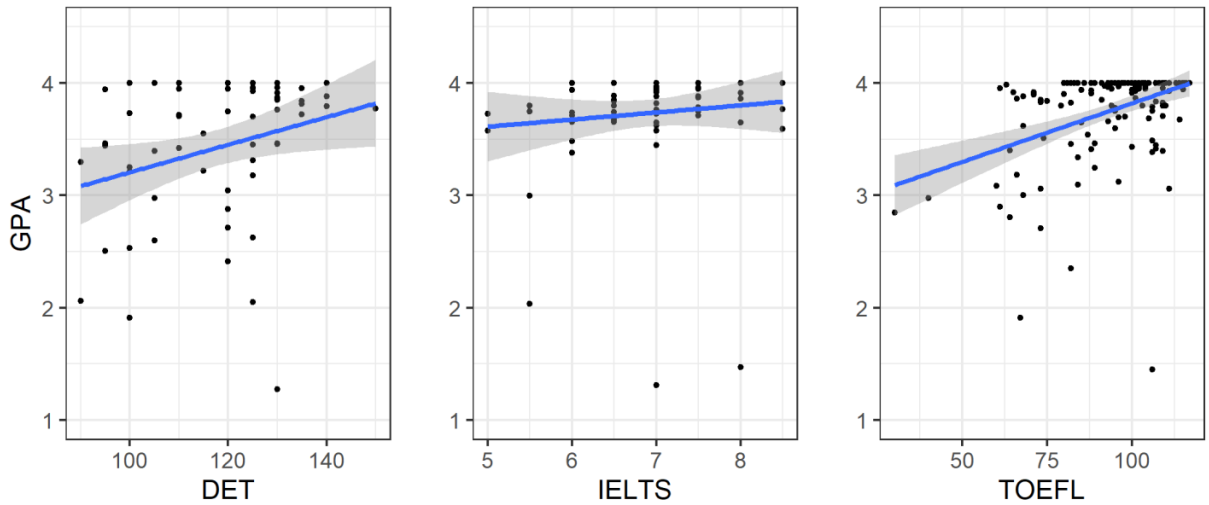


Figure 4

ELP test scores and first year GPA, by level (graduate and undergraduate separated): Fall 2022 cohort

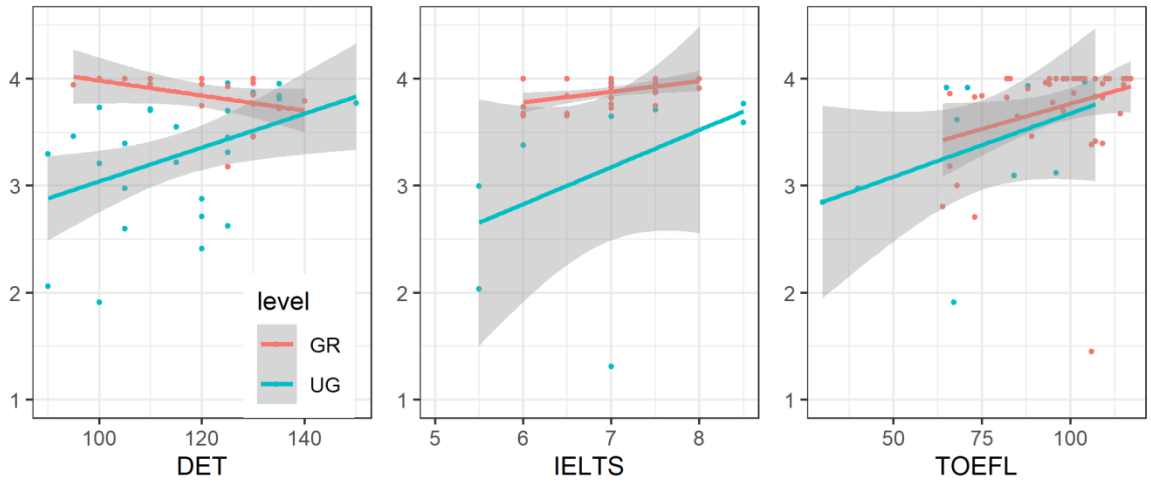
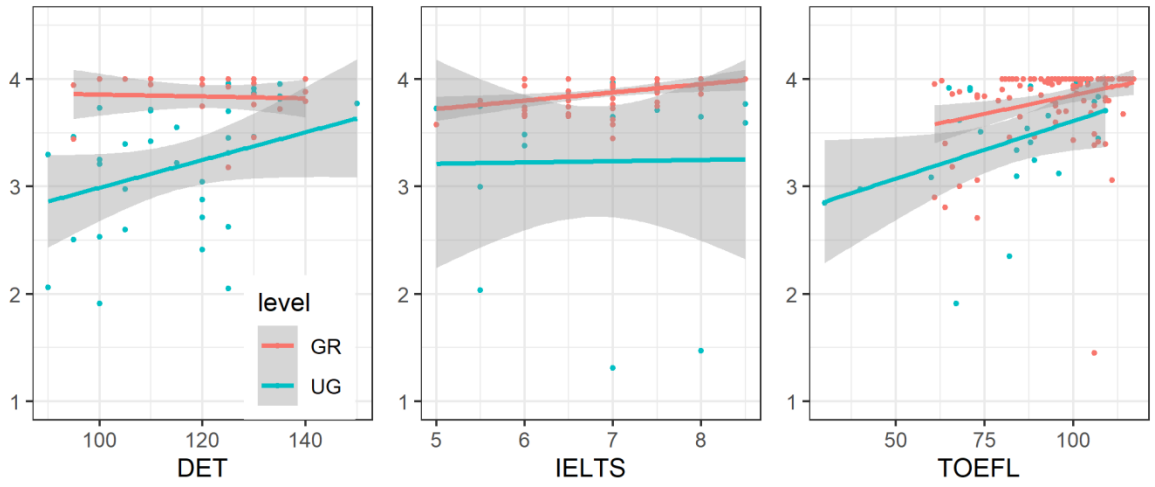


Figure 5

ELP test scores and first year GPA, by level (graduate and undergraduate separated): Combined cohorts entering Fall 2021, Spring 2022, and Fall 2022



Correlations between ELP Test Scores and First Year GPA: by Field of Study

Next, correlations were calculated for first year GPA and ELP test score for subgroups by field of study (life sciences, physical sciences and engineering, or social sciences and humanities). However, sample sizes became too small in many of the subgroups to draw meaningful conclusions. For the combined cohorts entering Fall 2021, Spring 2022, and Fall 2022, correlations between GPA and ELP test score only met the alpha for statistical significance

for social sciences and humanities students who submitted TOEFL ($r = .49$, $p < .001$, $n = 70$). Fall 2022 cohort correlations met the alpha for social sciences and humanities students submitting TOEFL ($r = .45$, $p = .017$, $n = 26$) or IELTS ($r = .83$, $p < .001$, $n = 12$). Correlation tables are reported for reference in Appendix A, but caution in interpretation is necessary due to the small subgroups.

Correlations between ELP Test Scores and GPA: Unconditional vs. Conditional Admission

As described above, at UHM, students may be admitted conditionally with DET scores of 95 for graduate or 90 for undergraduates, IELTS scores of 6.0 for graduate or 5.0 undergraduate, or 61 on TOEFL, while unconditional admission requires reaching 125 DET, 7.0 IELTS, or 100 TOEFL (see Table 2). For students admitted unconditionally (with higher ELP test scores), there were no statistically significant correlations between ELP test score and GPA (Table 6; Figures 6-8). For those admitted conditionally, with ELP scores reflecting a lower range of English proficiency that might make the linguistic demands of study in English more difficult, stronger and statistically significant correlations emerged (Table 7; Figures 9-11). This is notable, that for students admitted conditionally, their relative position among this lower entering ELP score range (for example, TOEFL 61-99) correlated more with their first year GPA outcomes, compared to unconditionally admitted students who exceeded higher ELP cut scores (for example, TOEFL 100-120) and for whom relative position above the already high cut score related less to their first year GPA.

The following tables and figures are for all three cohorts combined (students entering in Fall 2021, Spring 2022, and Fall 2022). See Appendix B for the Fall 2022 cohort alone.

Table 6

Students admitted unconditionally: Correlation between ELP test score and first year GPA, cohorts entering Fall 2021, Spring 2022, and Fall 2022

	ELP							95% CI			
Level	test	n	M	SD	min	max	r	LL	UL	p	R^2
All	DET	27	131	6.31	125	150	.25	-.14	.58	.203	.06
		GPA	3.57	0.64	1.27	4.00					
	IELTS	42	7.33	0.49	7.0	8.5	-.09	-.39	.22	.554	.01
		GPA	3.76	0.56	1.31	4.00					
	TOEFL	61	106	4.42	100	117	-.06	-.30	.19	.645	.00
		GPA	3.84	0.37	1.45	4.00					
GR	DET	12	132	5.77	125	140	.25	-.38	.72	.436	.06
		GPA	3.79	0.25	3.18	4.00					
	IELTS	33	7.26	0.42	7.0	8.5	.21	-.14	.52	.236	.04
		GPA	3.9	0.14	3.44	4.00					
	TOEFL	55	106	4.58	100	117	-.04	-.30	.22	.742	.00
		GPA	3.84	0.39	1.45	4.00					
UG	DET	15	130	6.81	125	150	.24	-.31	.67	.391	.06
		GPA	3.4	0.80	1.27	3.96					
	IELTS	9	7.61	0.65	7.0	8.5	.04	-.64	.69	.919	.00
		GPA	3.23	1.06	1.31	4.00					
	TOEFL	6	106	2.80	101	109	-.66	-.96	.33	.154	.44
		GPA	3.78	0.20	3.44	3.97					

Figure 6

Students admitted unconditionally, graduate and undergraduate combined: ELP test score and first year GPA, cohorts entering Fall 2021, Spring 2022, and Fall 2022

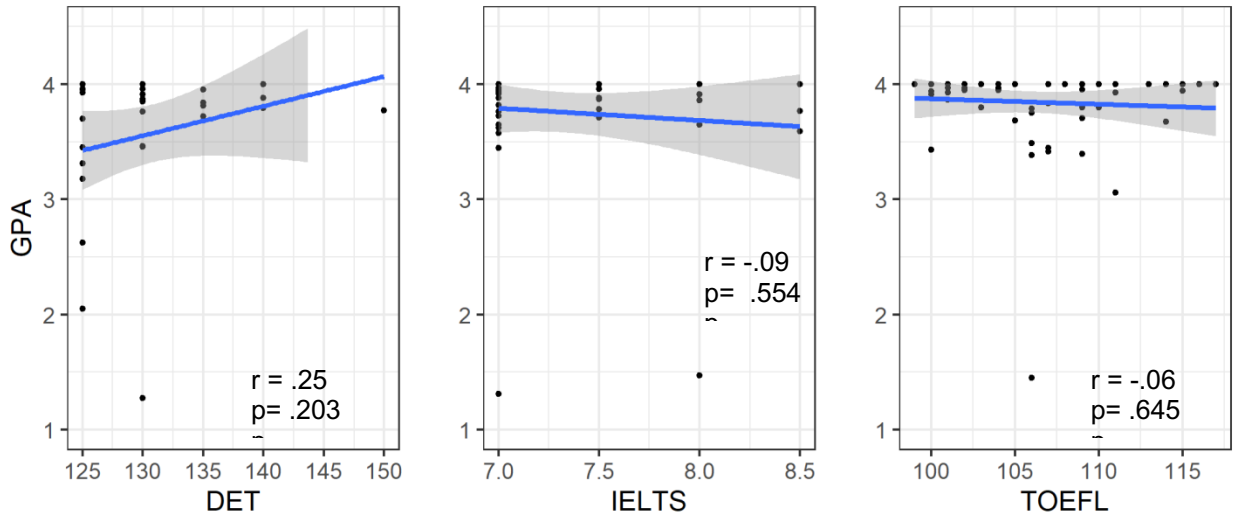


Figure 7

Students admitted unconditionally, graduate only: ELP test score and first year GPA, cohorts entering Fall 2021, Spring 2022, and Fall 2022

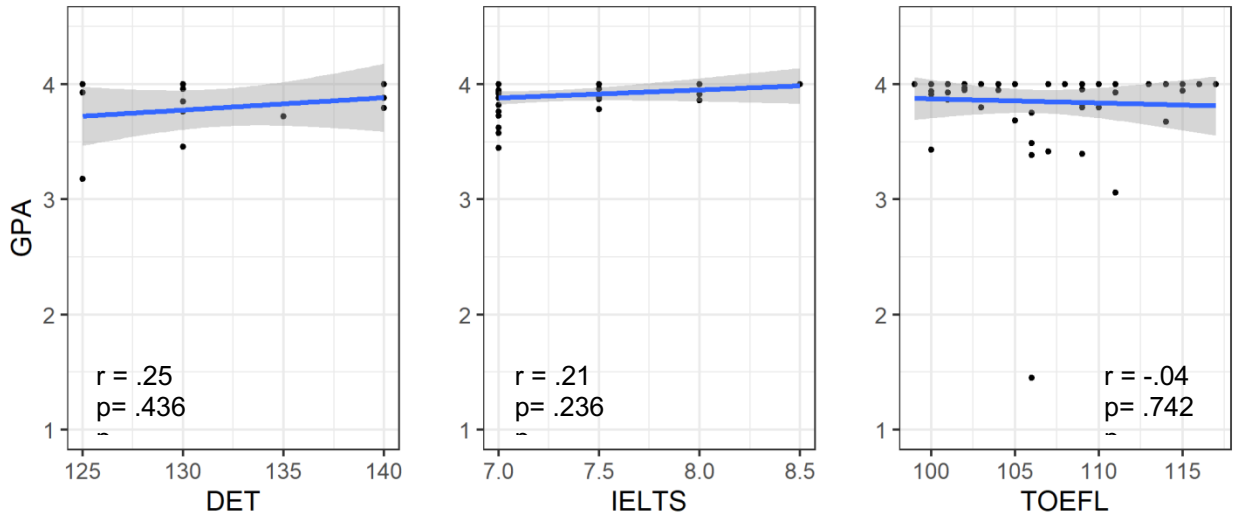


Figure 8

Students admitted unconditionally, undergraduate only: ELP test score and first year GPA, cohorts entering Fall 2021, Spring 2022, and Fall 2022

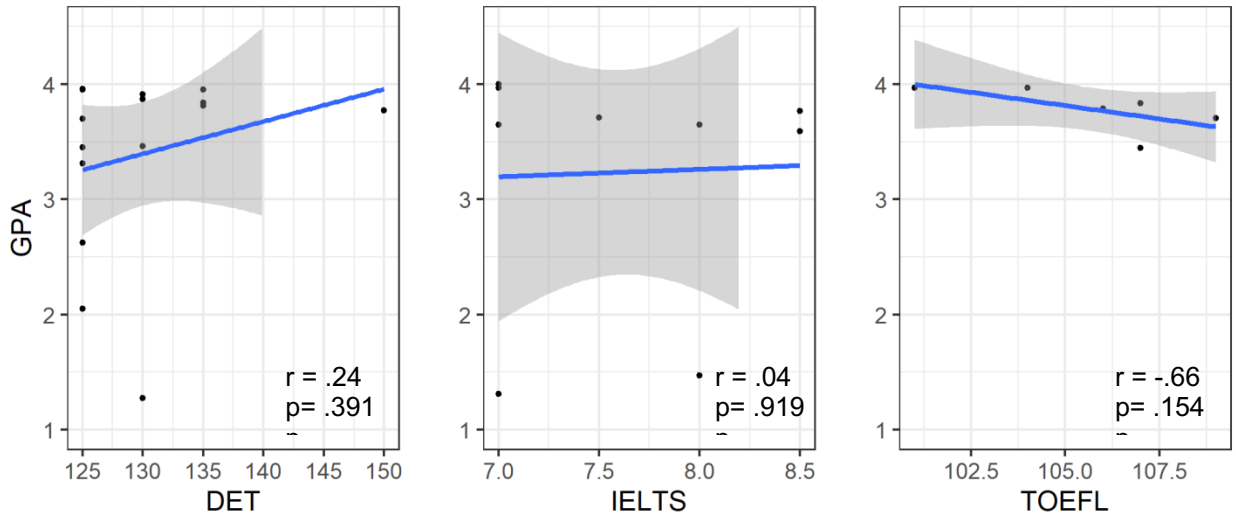


Table 7

Students admitted conditionally via ELI: Correlation between ELP test score and first year GPA, cohorts entering Fall 2021, Spring 2022, and Fall 2022

Level	ELP test	n	M	SD	min	max	r	95% CI LL	95% CI UL	p	R^2
All	DET	31	106	9.91	90	120	.16	-.20	.49	.377	.03
		GPA	3.29	0.6	1.91	4					
	IELTS	25	6.04	0.48	5	6.5	.39	.01	.67	.044*	.15
		GPA	3.68	0.41	2.03	4					
	TOEFL	73	83.4	13.2	30	99	.50	.31	.66	<.001***	.25
		GPA	3.71	0.39	2.35	4					
GR	DET	9	108	10.3	95	120	.29	-.46	.80	.443	.09
		GPA	3.89	0.19	3.44	4					
	IELTS	19	6.18	0.42	5	6.5	.30	-.16	.64	.193	.09
		GPA	3.82	0.14	3.58	4					
	TOEFL	57	85.3	11.6	61	99	.55	.34	.71	<.001***	.30
		GPA	3.79	0.33	2.71	4					
UG	DET	22	105	9.87	90	120	.08	-.35	.49	.710	.01
		GPA	3.05	0.53	1.91	3.73					
	IELTS	6	5.58	0.38	5	6	-.04	-.82	.80	.945	.00
		GPA	3.23	0.65	2.03	3.74					
	TOEFL	16	76.9	16.4	30	96	.23	-.26	.63	.348	.06
		GPA	3.4	0.44	2.35	3.93					

Figure 9

Students admitted conditionally via ELI, graduate and undergraduate combined: ELP test score and first year GPA, cohorts entering Fall 2021, Spring 2022, and Fall 2022

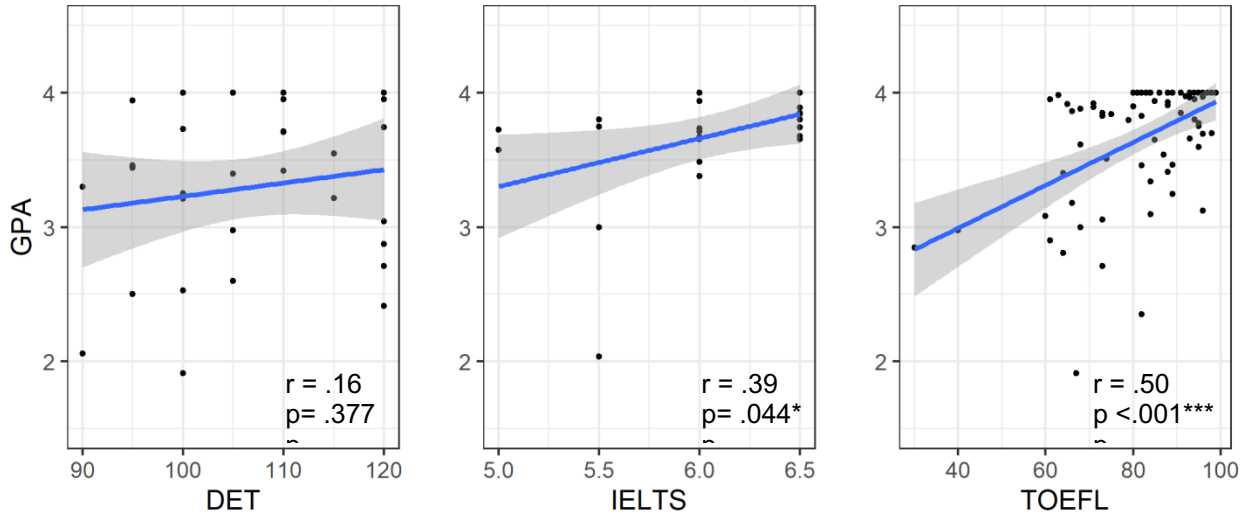


Figure 10

Students admitted conditionally via ELI, graduate only: ELP test score and first year GPA, cohorts entering Fall 2021, Spring 2022, and Fall 2022

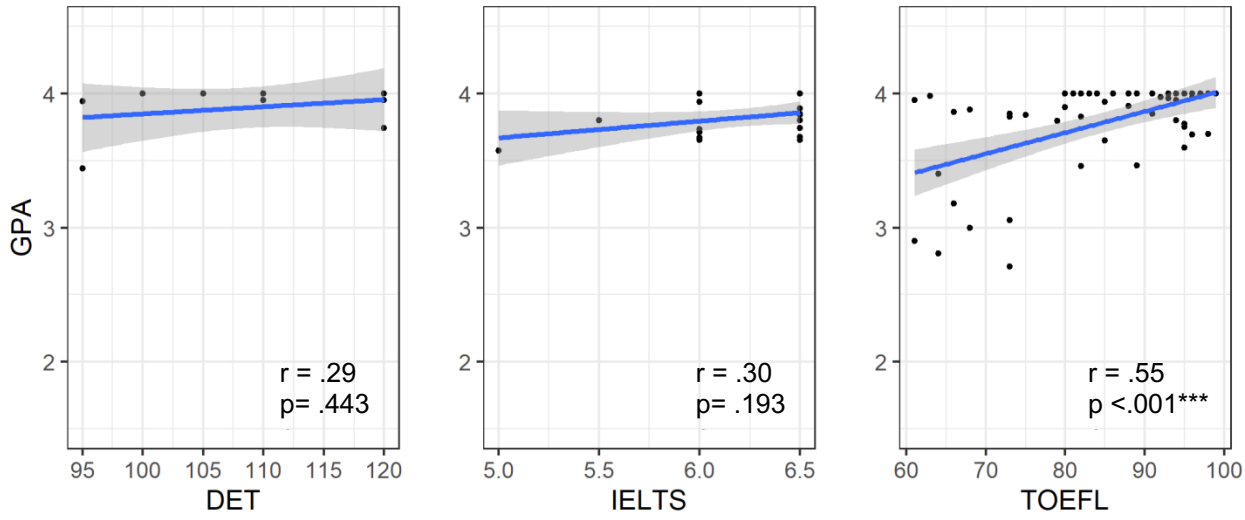
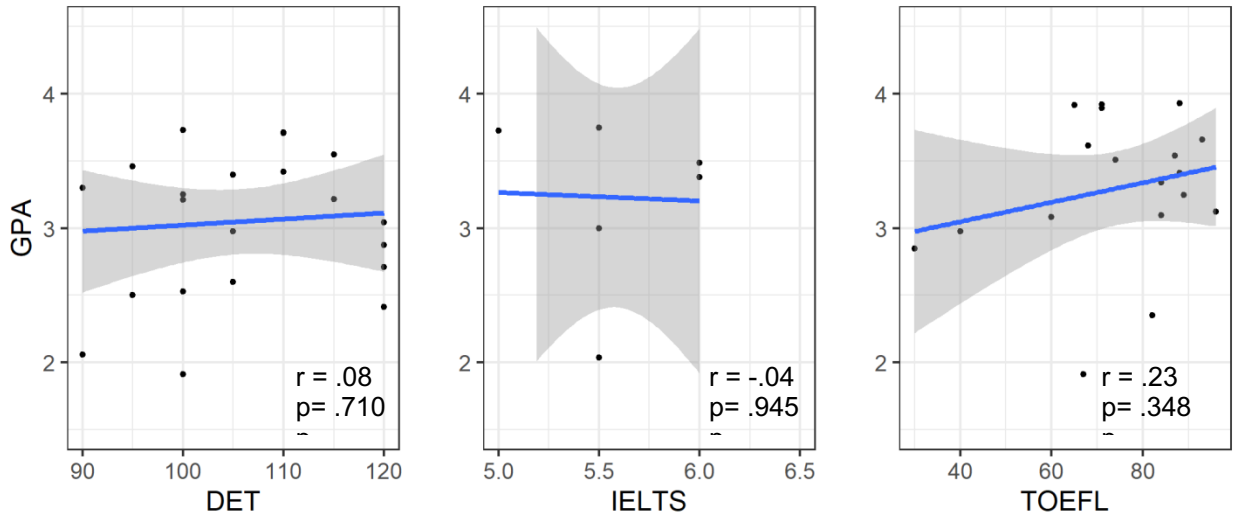


Figure 11

Students admitted conditionally via ELI, undergraduate only: ELP test score and first year GPA, cohorts entering Fall 2021, Spring 2022, and Fall 2022



Academic Outcomes: Mean GPA For Those Entering with Different ELP Tests

After exploring each test's correlations with first year GPA for different subgroups, this study turned to whether students entering on the basis of different ELP tests displayed significantly different academic outcomes. In this phase of the research, international students who did not submit an ELP test score were included as a fourth comparison group (no ELP). This latter group consists of students who fulfilled the ELP requirement in another way, such as by living or studying for an extended previous period in an English-speaking country.

A one-way ANOVA was conducted to compare the mean first year GPAs of students grouped by ELP test taken (a four-level categorical variable: no ELP, DET, IELTS, or TOEFL). For the Fall 2022 cohort alone, no statistically significant difference was found in mean GPAs between the groups, when considering all students combined or when separated by level (graduate or undergraduate) (Appendix C).

With all three cohorts combined (students entering in Fall 2021, Spring 2022, and Fall 2022 considered together), and with both levels combined (graduate and undergraduate), one-way ANOVA revealed a significant difference in mean GPA between at least two of the groups ($R^2 = .05$, $F(3, 572) = 9.65$, $p < .001$) (Table 8). However, importantly, when one-way ANOVA was applied to students separated by level (graduate and undergraduate disaggregated), any statistically significant differences in mean GPA by ELP disappeared.

Table 8

One-way ANOVA of mean first year GPA by ELP test taken, combined cohorts entering Fall 2021, Spring 2022, and Fall 2022, by level (graduate (GR) and undergraduate (UG))

Level of study	ELP test	First year GPA n	M	SD	SE	95% CI LL	UL	F	p	DF	R^2
Both *levels (GR & UG)*	no ELP	317	3.48	0.69	0.04	3.41	3.56	9.65	<.001***	572	0.05
	DET	58	3.42	0.63	0.08	3.26	3.58				
	IELTS	67	3.73	0.50	0.06	3.60	3.85				
	TOEFL	134	3.77	0.39	0.03	3.70	3.83				
GR only	no ELP	148	3.79	0.34	0.03	3.74	3.85	0.76	.516	329	0.01
	DET	21	3.84	0.23	0.05	3.74	3.93				
	IELTS	52	3.87	0.14	0.02	3.83	3.91				
	TOEFL	112	3.82	0.36	0.03	3.75	3.88				
UG only	no ELP	169	3.21	0.79	0.06	3.09	3.33	1.07	.362	239	0.01
	DET	37	3.19	0.67	0.11	2.97	3.40				
	IELTS	15	3.23	0.89	0.23	2.72	3.68				
	TOEFL	22	3.51	0.42	0.09	3.33	3.68				

Post hoc contrasts tests (with Tukey HSD adjustment for comparing a family of four estimates) were then conducted on the combined cohorts (entering Fall 2021, Spring 2022, and Fall 2022) and with graduate and undergraduate levels combined (Table 9). This post hoc test revealed that – again, only when both levels of students were aggregated – students entering either with no ELP score or with DET earned a statistically significantly lower mean GPA compared to those entering with either IELTS or TOEFL.

Table 9

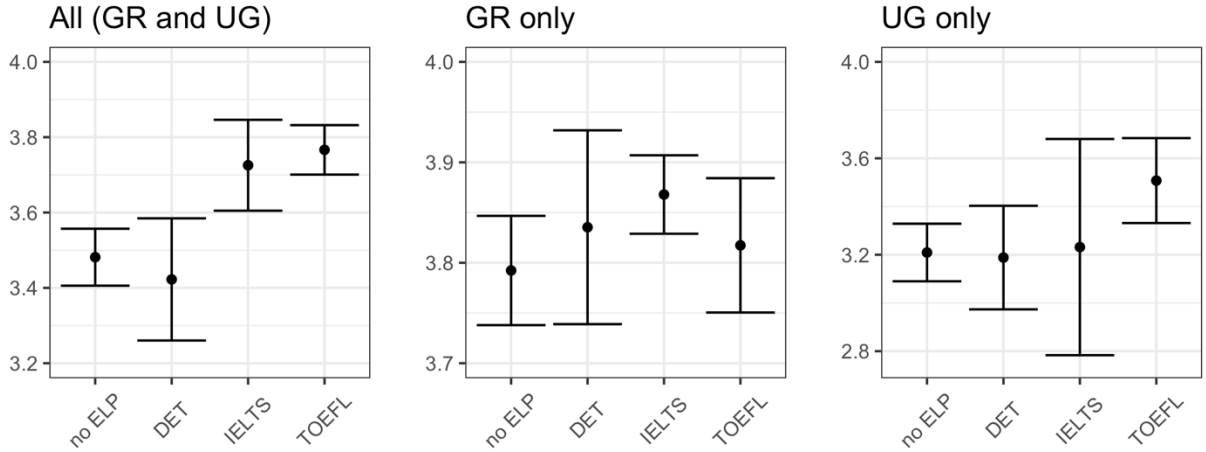
Post hoc contrasts test (Tukey method for comparing a family of four estimates), first year GPA by ELP test taken, combined cohorts entering Fall 2021, Spring 2022, and Fall 2022, all levels combined (graduate and undergraduate)

Contrast			estimate	SE	df	p	t ratio
no ELP	-	DET	0.06	0.09	572	.903	0.69
no ELP	-	IELTS	-0.24	0.08	572	.015*	-3.01
no ELP	-	TOEFL	-0.29	0.06	572	<.001** *	-4.59
DET	-	IELTS	-0.30	0.11	572	.027*	-3.51
DET	-	TOEFL	-0.34	0.09	572	.002**	-4.55
IELTS	-	TOEFL	-0.04	0.09	572	.969	-0.57

These results are graphically displayed in Figure 12. The first plot aggregates both student levels (graduate and undergraduate combined), showing that the 95% CI for the mean GPA of students with no ELP and with DET test scores did not overlap the mean GPA of IELTS or TOEFL test takers. Again, it is essential to note that these statistically significant differences only appeared with graduate and undergraduate students aggregated. When separated by level, any statistical significance disappeared, with the 95% CI of mean GPA for each group overlapping the other groups' mean GPA (final two plots in Figure 12).

Figure 12

Mean first year GPA with 95% confidence intervals by ELP test submitted, combined cohorts entering Fall 2021, Spring 2022, and Fall 2022, by level (graduate (GR) and undergraduate (UG))



Academic Outcomes: Probation or Withdrawal for Those Entering with Different ELP Tests

Moving to an academic outcomes measure with more serious consequences for students, a chi square test of equal proportions was used to determine whether the proportion of international students with an academic action – academic warning, probation, or withdrawn – was comparable by ELP test submitted (no ELP, DET, IELTS, or TOEFL). Pearson's chi square difference of proportions test found that the proportion of students with an academic action did not differ by groups who submitted no ELP score or each of the different ELP tests. This holds for the Fall 2022 cohort alone, as well as for the three cohorts combined, and whether separately evaluating graduate and undergraduate or combining both levels (Table 10). Two-sample tests for equality of proportions of students on academic action by test taken also returned no statistically significant differences (Table 11).

Table 10

Pearson's chi square difference of proportions tests for proportions of students on academic action (probation, warning, or withdrawal) by ELP test taken or no ELP test submitted

		ELP test score submitted						
Level		No ELP	DET	IELTS	TOEFL	X^2	df	p
Fall 2022 cohort alone								
All	Total	130	48	35	55	6.262	3	.100
	# academic action	22	4	1	6			
	Proportion acad. action	.17	.08	.03	.11			
GR	Total	56	18	26	44	6.884	3	.076
	# academic action	10	1	0	4			
	Proportion acad. action	.18	.06	.00	.09			
UG	Total	74	30	9	11	.869	3	.833
	# academic action	12	3	1	2			
	Proportion acad. action	.16	.10	.11	.18			
Combined cohorts entering Fall 2021, Spring 2022, and Fall 2022								
All	Total	352	63	70	145	4.659	3	.199
	# academic action	49	8	4	14			
	Proportion acad. action	.14	.13	.06	.10			
GR	Total	160	22	55	121	2.660	3	.447
	# academic action	16	1	2	11			
	Proportion acad. action	.10	.05	.04	.09			
UG	Total	192	41	15	24	.462	3	.927
	# academic action	33	7	2	3			
	Proportion acad. action	.17	.17	.13	.13			

Table 11

Matrix of p-values for two-sample tests for equality of proportions of students on academic warning, probation, or withdrawing, by ELP test submitted, for the Fall 2022 cohort alone, or the combined cohorts entering Fall 2021, Spring 2022, and Fall 2022

Fall 2022 cohort alone				Combined cohorts entering Fall 2021, Spring 2022, and Fall 2022			
	No ELP	DET	IELTS		No ELP	DET	IELTS
DET	.230			DET	.952		
IELTS	.063	.570		IELTS	.090	.271	
TOEFL	.413	.915	.324	TOEFL	.250	.682	.475

The primary finding here is that there was no statistically significant difference in the proportions of students on academic action based on which ELP test they submitted for admission. Still, it is worth noting patterns of interest in the raw proportions. First, overall, graduate students experienced academic probation or dropped out in lower raw proportions than undergraduates. Among students who submitted any ELP test score, the academic action or withdrawal rate ranged from .00-.09 of graduate students by ELP test taken, and .10-.18 of undergraduates with such test scores. Perhaps it is not surprising that graduate students with more advanced education, more stringent admissions requirements, more maturity, and potentially more well-defined academic plans persisted more successfully.

Potentially problematic, the raw proportion of students who experienced academic warning or probation or withdrew was higher in general among those who were exempt from submitting an ELP test score compared to those who were required to do so. For example, in the Fall 2022 cohort (the first semester with mostly in-person on-campus operations since COVID-19 began), 17% of international students entering without an ELP test score had an academic action or withdrew, compared to 3-11% of those required to submit such a score. Again, while these differences did not reach the level of statistical significance, it is worth flagging this rate of students experiencing critical negative academic outcomes after having been exempted from providing an ELP score to evidence their preparation for academic study in English.

DISCUSSION

Existing literature, theory, and logic tell us that many other factors influence first year academic success beyond English language proficiency alone. Thus, small to moderate correlations with small R^2 are not a surprise. Such mixed results in terms of statistical significance and strength of relationships are common in correlation studies of GPA and ELP test scores (Bridgeman et al., 2016; Dang & Dang, 2021; Ginther & Yan, 2018; Hu & Trenkic, 2021; Ihlenfeldt & Rios, 2022). Compounding this, graduate programs often require that students maintain a higher minimum GPA (at UHM, 3.0), leading to a GPA ceiling effect.

However, the correlations in the present study were stronger compared to other studies of ELP test scores and GPA, including those reported in Ihlenfeldt and Rios's (2022) meta-analysis and the even weaker correlations reported by Isaacs et al. (2023). This is in part thanks to the wider range of ELP test scores in this sample, because UHM offers a conditional admission option for students with lower ELP test scores. UHM admissions decisions include students who scored across 90-95% of the TOEFL score distribution, and 80% of the DET score distribution. By comparison, Isaacs et al. (2023) included students with scores across only about 50% of the range of TOEFL scores, and 32% of the range of DET scores. That the present study found stronger correlations between ELP score and GPA reflects some mitigation of the restriction of range seen in most other correlation studies of entrance exams and first year GPA.

Further, when students in the present dataset were separated into those admitted conditionally with lower ELP test scores (DET 90-124, IELTS 5.0-6.5, TOEFL 61-99), and those admitted unconditionally with higher ELP cut scores (DET 125, IELTS 7.0, TOEFL 100), statistically significant correlations were found for conditionally admitted students with lower ELP test scores, but no statistically significant correlations emerged for unconditionally admitted students with higher scores. For conditionally admitted students with a lower range of entering English ability, we might expect them to struggle more with the linguistic demands of study overall. Indeed for those students, it appears that their relative starting position along the range of lower English proficiency scores may have mattered more in terms of their eventual grades, than specific score above the higher cut score mattered for unconditionally admitted students. This could be evidence that the unconditional admit cut scores are well set, as students who exceeded

them achieved less varied eventual GPAs and were presumably comparably well prepared regardless of relative score placement over the cut.

Of greater interest than specific separate correlation results is whether different ELP tests perform differently when used to decide that a student is linguistically prepared for college study in English. In the correlations, the unexpectedly strong negative correlation for graduate students' DET scores and first year GPA in the Fall 2022 cohort (r = -.40) bears further attention, although this is moderated in the three cohorts combined (r = -.05 for students entering in Fall 2021, Spring 2022, and Fall 2022) and does not reach statistical significance in either group (p = .112 and p = .814, respectively). An important caveat when considering graduate student grade correlations is the ceiling effect stemming from the Graduate Division requirement to maintain a 3.0 GPA (see Figure 1). Virtually all graduate first year GPAs in this dataset were over 3.0, and cluster near 4.0. In this subsample, the restricted range of graduate GPA moderated correlation strength, significance, and perhaps direction.

A one-way ANOVA comparing mean GPA of international students who were exempt from submitting ELP scores (no ELP) or each of the different ELP tests at first suggested there was a statistically significant difference between at least two of these groups, with graduate and undergraduate levels combined, and with the three entry semesters combined. However, it is necessary to emphasize that these mean GPA differences only appeared when graduate and undergraduate students were aggregated together in a combined analysis. When graduate and undergraduate students were disaggregated, the 95% confidence intervals for each ELP test group largely overlapped. Importantly, graduate and undergraduate students did not submit the different tests in the same proportions (see Table 1): across the Fall 2021, Spring 2022, and Fall 2022 cohorts, DET was submitted by only 22 graduate students but 41 undergraduates; IELTS was submitted by 55 graduates but only 16 undergraduates; and TOEFL was overwhelmingly preferred by graduate students, with 123 scores compared to 25 undergraduate TOEFL scores. This different representation from student levels within each test group, and the fact that graduate GPAs cluster near 4.0 while undergraduates spread to 2.0, demand caution in making conclusions based on calculations that aggregate graduate and undergraduate students. In summary, statistically, we cannot reject the possibility that mean GPA among the ELP test

groups (or no ELP test) may be the same, and that there is no difference in a students' probable GPA outcome solely on the basis of which ELP test they submitted.

More critical for students' academic outcomes than fractions of GPA points is the risk of dropping out or earning a low enough GPA to warrant academic action (probation or warning). Difference of proportions tests found no statistically significant difference in the proportion of students withdrawing or on probation or warning between those who submitted the three different ELP tests (or no ELP test). In this analysis, as in the ANOVA previously reported, there was no apparent significant difference in risk of these serious negative academic outcomes based on belonging to the group of students who submitted any of these different ELP tests or no ELP test, for either graduate or undergraduate students. However, an unexpected finding here was that students exempt from submitting an ELP test score had a higher proportion of academic probation, warning, or withdrawal than expected, with raw rates higher than for students who submitted an ELP test score as evidence of linguistic preparation.

Returning again to Figure 1, the spread of individual students' GPAs extended lowest among those with no ELP score. While no DET or IETLS graduate student test takers earned a GPA under 3.0, outliers below the whisker extended below 3.0 for those with no ELP score (though also for TOEFL). Among undergraduates, the outliers extended the lowest for students with no ELP score, including the only GPAs in the sample below 1.0. These findings suggest an important opportunity to review ELP test exemption, and to break this group of students into smaller meaningful subgroups for further analyses of potential patterns.

LIMITATIONS

As is common, a larger sample size would benefit this study. In terms of subgroups in this dataset, statistically significant correlations with GPA were found most often in relationship to TOEFL test scores. This group had the largest sample size, as TOEFL was the most submitted ELP test. By contrast, no meaningful conclusions could be drawn about differences by field of study in this population, as the subgroups by field of study yielded sample sizes too small for interpretable statistical evaluation, and were ultimately moved to Appendix A of this report. In addition, no single country of origin represented a large enough proportion of students at UHM

in this study to warrant considerations by country or suggest effects by L1. Incorporating data from additional semesters, especially if the number of applicants submitting DET increased, would support more confident recommendations based on findings and may support further subgroup comparisons.

Fall 2022 was the first "post-COVID" semester, when UHM began returning to majority in-person teaching and on-campus operations and student services. This is another reason it will be instructive to evaluate data from additional future semesters, to consider whether and how patterns changed after pandemic restrictions eased, both in terms of ELP testing and in terms of student experience at UHM. It would also enable making important comparisons with awareness of recent revisions in DET structure and items. Future data on UHM graduate students will be limited because the Graduate Division discontinued accepting DET as an ELP test in Spring 2023. Unless Graduate Division policy changes, continued analyses should still be made for undergraduate students alone.

Initially, this study intended to model previously earned GPA in combination with ELP test score as potential predictors of future GPA. However, there is no detail about previous GPA recorded in this institutional database regarding the country or region of the prior institution, how the previous GPA was calculated, or the level of study represented (i.e., whether high school, associate degree, or prior undergraduate or master's degree grades). These factors make it challenging to assign meaning to any individual student's previous GPA or in comparison to others. Although this limitation cannot presently be addressed, exploring potential predictive power of previous GPA in association with ELP test score is an area for future study.

CONCLUSION

Although overall these analyses returned no glaring statistically significant differences in these ELP tests' functioning at UHM, there are subgroups where DET looks at least on the surface different than IELTS and TOEFL, in spread of undergraduate GPA and in a (statistically non-significant) negative correlation with graduate GPA. Still, crucially, statistically significant differences in mean GPA of DET test takers or those with no ELP compared to IELTS and

TOEFL only appeared if graduate and undergraduate students were combined; no significant differences were found when graduate and undergraduate level students were disaggregated.

Notably, the present analyses indicated that international students exempt from submitting an ELP test score for UHM admission also earned lower mean GPAs than IELTS and TOEFL takers (with graduate and undergraduate students combined), and experienced academic probation or warning or withdrew in higher raw rates than any of the ELP groups. Similarly, previous research in Australia found ELP tests to be more reliable in establishing a student's readiness to complete university study in English compared to other exemption criteria (Oliver et al., 2012). This suggests an opportunity to reevaluate UHM's ELP exemption criteria by further breaking the ELP test exempt students into subgroups based on why they were exempted, and examining their outcomes.

While differences among these tests in terms of predictive validity cannot be definitively ruled out, the present results suggest there are no immediate red flags against accepting DET alongside IELTS and TOEFL. Perhaps most critically for student outcomes, no significant difference was found in proportions of international students withdrawing or on academic action or probation after their first two semesters when grouped by ELP test submitted (or compared to those who submitted no ELP test). Among Graduate students who submitted DET for admission, none earned a first year GPA under the 3.0 threshold for probation. Especially in light of ongoing revisions to improve the DET, and the inarguable savings in price, time, and easier access to DET for potential test takers, these findings support recommending that the Graduate Division reconsider accepting DET.

Locally, this study contributes to ongoing discussions of using different ELP scores in admissions decisions for international students at UHM. More broadly, it fills a gap in the existing literature by examining outcomes in a new context (a large, public, less selective state university in the US) and including a newer test, the DET, building directly on Isaacs et al.'s (2023) study in a highly-selective UK institution. If valid for admissions decisions, continued use of DET could support justice, in providing more access to ELP testing options more broadly to more potential students. DET is cheaper than most other accepted tests, takes far less time to complete, and can be taken anywhere and at any time (Settles et al., 2020). For students in socioeconomic or geopolitical situations that make it difficult to access the more expensive

IELTS and TOEFL, DET provides an option. UHM, along with other English-medium institutions, may wish to re-evaluate DET cut scores in light of more student outcomes data, newer linking studies, and in combination with other evidence of prospective students' academic readiness. Universities do not seek to do harm by admitting those not prepared for success, but neither do admissions committees want to unjustly hinder access for those who are prepared. Good ELP tests are useful tools for identifying students who are ready to study, but these potential students must first be able to access the test.

ACKNOWLEDGEMENTS

Thank you to Dr. Daniel Isbell for guidance throughout this project, from its conception. Thank you to Dr. Dustin Crowther for valuable feedback as second reader. With appreciation to UHM staff Duncan Hui (Graduate Division), and Roy Suda, Ryan Yamaguchi, and Pheng Xiong (Undergraduate Admissions and Enrollment Management) for their work to understand and fulfill the data request.

REFERENCES

Abunawas, M. E. (2014). *A meta-analytic investigation of the predictive validity of the test of English as a foreign language (TOEFL) scores on GPA* [Dissertation]. Texas A&M University.

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, *2*(1), 1–34. https://doi.org/10.1207/s15434311laq0201_1

Bézy, M., & Settles, B. (2015). *The Duolingo English Test and East Africa: Preliminary linking results with IELTS & CEFR* (Duolingo Research Report DRR-15-01). https://s3.amazonaws.com/duolingo-papers/reports/DRR-15-01.pdf

Brenzel, J., & Settles, B. (2017). *The Duolingo English Test—Design, validity, and value* (Duolingo Research Report DET Whitepaper). https://s3.amazonaws.com/duolingo-papers/other/DET_ShortPaper.pdf

Bridgeman, B., Cho, Y., & DiPietro, S. (2016). Predicting grades from an English language assessment: The importance of peeling the onion. *Language Testing*, *33*(3), 307–318. https://doi.org/10.1177/0265532215583066

Brown, L. G. M. (2017). *A quantitative study of the Test of English as a Foreign Language (TOEFL) exam and its effects on international student academic achievement* [Dissertation]. University of Southern California.

Cardwell, R., Naismith, B., LaFlair, G., & Nydick, S. (2023). *Duolingo English Test: Technical manual*. Duolingo. https://doi.org/10.46999/CQNG4625

Chapelle, C. A. (2012). Conceptions of validity. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 21–33). Routledge.

Chapelle, C. A., Enright, M., & Jamieson, J. (2008). *Building a validity argument for the Test of English as a Foreign Language*. Routledge.

Cho, Y., & Bridgeman, B. (2012). Relationship of TOEFL iBT® scores to academic performance: Some evidence from American universities. *Language Testing*, *29*(3), 421–442. https://doi.org/10.1177/0265532211430368

College Board. (n.d.). *BigFuture: College Search*. College Search - BigFuture. Retrieved November 19, 2023, from https://bigfuture.collegeboard.org/college-search/filters

Daller, M., & Phelan, D. (2013). Predicting international student study success. *Applied Linguistics Review*, *4*(1), 173–193. https://doi.org/10.1515/applirev-2013-0008

Daller, M., & Yixin, W. (2017). Predicting study success of international students. *Applied Linguistics Review*, *8*(4), 355–374. https://doi.org/10.1515/applirev-2016-2013

Dang, C. N., & Dang, T. N. Y. (2021). The predictive validity of the IELTS Test and contribution of IELTS preparation courses to international students' subsequent academic study: Insights from Vietnamese international students in the UK. *RELC Journal*, 003368822098553. https://doi.org/10.1177/0033688220985533

Dooey, P., & Oliver, R. (2002). An investigation into the predictive validity of the IELTS Test as an indicator of future academic success. *Prospect: An Australian Journal of TESOL*, *17*(1), 36–54.

ETS. (2023). *TOEFL iBT test and score data summary 2022*. https://www.ets.org/pdfs/toefl/toefl-ibt-test-score-data-summary-2022.pdf

European Research Council. (2021). *Panel structure for ERC calls 2021 and 2022 (revised)*. https://erc.europa.eu/sites/default/files/document/file/ERC_Panel_structure_2021_2022.pdf

Gagen, T. (2019). *The Predictive validity of IELTS scores: A meta-analysis*. The University of Western Ontario.

Ginther, A., & Yan, X. (2018). Interpreting the relationships between TOEFL iBT scores and GPA: Language proficiency, policy, and profiles. *Language Testing*, *35*(2), 271–295. https://doi.org/10.1177/0265532217704010

Graduate Division. (n.d.-a). *Disciplinary Actions*. Retrieved November 19, 2023, from https://manoa.hawaii.edu/graduate/disciplinary-actions/

Graduate Division. (n.d.-b). *English Proficiency*. Retrieved May 4, 2023, from https://manoa.hawaii.edu/graduate/english-proficiency/

Harsch, C., Ushioda, E., & Ladroue, C. (2017). Investigating the predictive validity of TOEFL iBT® test scores and their use in informing policy in a United Kingdom university setting. *ETS Research Report Series*, *2017*(1), 1–80. https://doi.org/10.1002/ets2.12167

Hu, R., & Trenkic, D. (2021). The effects of coaching and repeated test-taking on Chinese candidates' IELTS scores, their English proficiency, and subsequent academic achievement. *International Journal of Bilingual Education and Bilingualism*, *24*(10), 1486–1501. https://doi.org/10.1080/13670050.2019.1691498

Ihlenfeldt, S. D., & Rios, J. A. (2022). A meta-analysis on the predictive validity of English language proficiency assessments for college admissions. *Language Testing*, 026553222211123. https://doi.org/10.1177/02655322221112364

International Student Services. (n.d.). *International Student Statistics*. Retrieved November 4, 2023, from https://www.hawaii.edu/issmanoa/?page_id=3633

Isaacs, T., Hu, R., Trenkic, D., & Varga, J. (2023). Examining the predictive validity of the Duolingo English Test: Evidence from a major UK university. *Language Testing*, *40*(3), 748–770. https://doi.org/10.1177/02655322231158550

Isbell, D. R., Crowther, D., & Nishizawa, H. (2023). Speaking performances, stakeholder perceptions, and test scores: Extrapolating from the Duolingo English test to the university. *Language Testing*, Online First. https://doi.org/10.1177/02655322231165984

Isbell, D. R., & Kremmel, B. (2020). Test review: Current options in at-home language proficiency tests for making high-stakes decisions. *Language Testing*, *37*(4), 600–619. https://doi.org/10.1177/0265532220943483

Ishikawa, L., Hall, K., & Settles, B. (2016). *The Duolingo English Test and academic English* (Duolingo Research Report DRR-16-01).

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73. https://doi.org/10.1111/jedm.12000

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, *315*(5815), 1080–1081. https://doi.org/10.1126/science.1136618

Mānoa Advising Center. (n.d.). *Academic Actions*. Retrieved November 19, 2023, from https://manoa.hawaii.edu/undergrad/mac/academic-actions/

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13–103). Macmillan Publishing Co.

Office of Admissions. (n.d.). *International Admission Requirements*. Retrieved May 4, 2023, from https://manoa.hawaii.edu/admissions/international/

Oliver, R., Vanderford, S., & Grote, E. (2012). Evidence of English language proficiency and academic achievement of non-English-speaking background students. *Higher Education*

Research & Development, *31*(4), 541–555. https://doi.org/10.1080/07294360.2011.653958

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Schoepp, K. (2018). Predictive validity of the IELTS in an English as a medium of instruction environment. *Higher Education Quarterly*, *72*(4), 271–285. https://doi.org/10.1111/hequ.12163

Settles, B. (2016). *The Reliability of Duolingo English Test scores* (Duolingo Research Report DRR-16-02). https://s3.amazonaws.com/duolingo-papers/reports/DRR-16-02.pdf

Settles, B., T. LaFlair, G., & Hagiwara, M. (2020). Machine learning–driven language assessment. *Transactions of the Association for Computational Linguistics*, *8*, 247–263. https://doi.org/10.1162/tacl_a_00310

Shakibaei, G., & Memari, M. (2019). On the predictive validity (generalizability) of IELTS academic tests. *Teaching English Language*, *13*(2), 299–317.

Stoynoff, S., & Chapelle, C. A. (2005). Understanding language tests and testing practices. In *ESOL tests and testing*. TESOL.

University College London. (n.d.). *English language requirements*. Prospective Students Undergraduate. Retrieved January 8, 2024, from https://www.ucl.ac.uk/prospective-students/undergraduate/how-apply/english-language-requirements

University of Hawai‘i at Mānoa. (n.d.). *About UH Mānoa*. Retrieved November 4, 2023, from https://manoa.hawaii.edu/about/

Wagner, E. (2020). Duolingo English Test, revised version July 2019. *Language Assessment Quarterly*, *17*(3), 300–315. https://doi.org/10.1080/15434303.2020.1771343

Yao, D. (2023). Examining the subjective fairness of at-home and online tests: Taking Duolingo English Test as an example. *PLOS ONE*, *18*(9), e0291629. https://doi.org/10.1371/journal.pone.0291629

Ye, F. (2014). *Validity, reliability, and concordance of the Duolingo English Test* [Technical Report]. University of Pittsburgh. https://s3.amazonaws.com/duolingo-papers/other/ye.testcenter14.pdf

York, T. T., Gibson, C., & Rankin, S. (2015). Defining and Measuring Academic Success. *Practical Assessment, Research, and Evaluation*, *20*(Article 5). https://doi.org/10.7275/hz5x-tx03

Zimmermann, J., von Davier, A. A., Buhmann, J. M., & Heinimann, H. R. (2018). Validity of GRE General Test scores and TOEFL scores for graduate admission to a technical university in Western Europe. *European Journal of Engineering Education*, *43*(1), 144–165. https://doi.org/10.1080/03043797.2017.1343277

APPENDIX A

Table A1

Correlation between ELP test scores and first year GPA by field of study, graduate and UG combined, combined cohorts entering Fall 2021, Spring 2022, and Fall 2022, and Fall 2022 cohort only, by field of study

	ELP							95% CI			
Field	test	n	M	SD	min	max	r	LL	UL	p	R^2
Combined cohorts entering Fall 2021, Spring 2022, and Fall 2022											
Life	DET	11	125	15.4	95	150	.58	-.03	.88	.060	.34
Sciences		GPA	3.38	0.61	2.41	4					
(LS)	IELTS	14	6.79	0.85	5.5	8.5	.37	-.20	.75	.193	.14
		GPA	3.71	0.50	2.03	4					
	TOEFL	23	91.9	9.98	63	107	.05	-.37	.45	.824	.00
		GPA	3.82	0.28	3.1	4					
Physical	DET	15	123	13.9	95	140	.08	-.45	.57	.790	.01
Sciences		GPA	3.52	0.68	1.27	4					
and	IELTS	24	6.84	0.75	5	8.5	.00	-.40	.39	.984	.00
Engineering		GPA	3.76	0.54	1.31	4					
(PE)	TOEFL	41	97.4	13.4	61	117	.07	-.24	.36	.668	.00
		GPA	3.8	0.24	3	4					
Social	DET	32	113	13.7	90	140	.30	-.06	.59	.098	.09
Sciences		GPA	3.39	0.63	1.91	4					
and	IELTS	29	6.93	0.82	5	8.5	.06	-.31	.41	.747	.00
Humanities		GPA	3.71	0.49	1.47	4					
(SH)	TOEFL	70	91.6	18.4	30	116	.49	.29	.65	<.001***	.24
		GPA	3.73	0.48	1.45	4					
Fall 2022 cohort only											
LS	DET	8	127	13.1	110	150	.30	-.51	.83	.465	.09
		GPA	3.46	0.59	2.41	4.00					
	IELTS	7	6.79	0.99	5.5	8.5	.62	-.25	.94	.140	.38
		GPA	3.54	0.68	2.03	3.97					
	TOEFL	11	91.1	7.06	82	102	.32	-.35	.77	.342	.10
		GPA	3.83	0.29	3.10	4.00					
PE	DET	13	125	12.3	95	140	.48	-.09	.82	.095	.23
		GPA	3.7	0.27	3.18	4.00					
	IELTS	15	7	0.63	6	8.5	-.05	-.53	.46	.853	.00
		GPA	3.73	0.68	1.31	4.00					
	TOEFL	15	99.3	16.3	65	117	.26	-.28	.67	.340	.07
		GPA	3.73	0.30	3.00	4.00					
SH	DET	24	111	13	90	130	.26	-.16	.60	.224	.07
		GPA	3.39	0.61	1.91	4.00					
	IELTS	12	7	0.74	5.5	8	.83	.52	.95	<.001***	.70
		GPA	3.72	0.29	3.00	4.00					
	TOEFL	26	88.6	23.2	30	116	.45	.09	.70	.017*	.20
		GPA	3.63	0.61	1.45	4.00					

Table A2

Correlation between ELP test scores and first year GPA by field of study: Cohorts entering Fall 2021, Spring 2022, and Fall 2022

	ELP							95% CI			
Field	test	n	M	SD	min	max	r	LL	UL	p	R^2
Graduate only											
LS	DET	2	125	21.2	110	140	not enough observations				
		GPA	4	0	4	4					
	IELTS	10	6.60	0.57	6	7.5	.08	-.58	.68	.816	.01
		GPA	3.85	0.13	3.66	4					
	TOEFL	17	91.6	10.5	63	105	-.02	-.50	.46	.935	.00
		GPA	3.92	0.16	3.46	4					
PE	DET	9	126	12.9	95	140	.45	-.30	.86	.222	.20
		GPA	3.73	0.27	3.18	4					
	IELTS	22	6.76	0.69	5	8	.40	-.01	.70	.056	.16
		GPA	3.88	0.14	3.58	4					
	TOEFL	38	97.7	12.6	61	117	.11	-.21	.41	.494	.01
		GPA	3.8	0.25	3	4					
SH	DET	10	118	14.6	95	140	-.49	-.86	.20	.152	.24
		GPA	3.89	0.17	3.45	4					
	IELTS	20	7.17	0.60	6	8.5	.44	.01	.73	.047*	.19
		GPA	3.87	0.16	3.44	4					
	TOEFL	57	95.7	15.1	61	116	.37	.12	.57	.004**	.14
		GPA	3.8	0.46	1.45	4					
Undergraduate only											
LS	DET	9	124	15.5	95	150	.74	.14	.94	.024*	.54
		GPA	3.25	0.59	2.41	3.95					
	IELTS	4	7.25	1.32	5.5	8.5	.82	-.67	1.00	.181	.67
		GPA	3.35	0.89	2.03	3.97					
	TOEFL	6	92.7	9.14	84	107	.33	-.66	.90	.528	.11
		GPA	3.54	0.35	3.1	3.97					
PE	DET	6	119	15.6	95	135	-.19	-.87	.73	.716	.04
		GPA	3.21	0.98	1.27	3.96					
	IELTS	2	7.75	1.06	7	8.5	not enough observations				
		GPA	2.45	1.61	1.31	3.59					
	TOEFL	3	93.7	24.8	65	109	-.82	--	--	.391	.67
		GPA	3.82	0.11	3.7	3.91					
SH	DET	22	111	13.1	90	130	.33	-.11	.66	.139	.11
		GPA	3.16	0.63	1.91	3.95					
	IELTS	9	6.39	1.02	5	8	-.43	-.85	.33	.248	.18
		GPA	3.35	0.76	1.47	4					
	TOEFL	13	75.6	21.5	30*	106	.40	-.14	.76	.141	.16

	GPA	3.42	0.47	2.35	3.97	

*One student with TOEFL score of 30 earned a 2.85 GPA in the first two semesters

Table A3

Correlation between ELP test scores and first year GPA by field of study: Fall 2022 cohort only

	ELP							95% CI			
Field	test	n	M	SD	min	max	r	LL	UL	p	R^2
Graduate only											
LS	DET	1	110	--	110	110	not enough observations				
		GPA	4	--	--	--					
	IELTS	4	6.62	0.63	6	7.5	.81	-.68	1.00	.190	.66
		GPA	3.76	0.14	3.66	3.96					
	TOEFL	9	92.2	7.31	82	102	.08	-.62	.71	.835	.01
		GPA	3.91	0.17	3.46	4.00					
PE	DET	8	129	6.23	120	140	.21	-.58	.80	.615	.04
		GPA	3.77	0.26	3.18	4.00					
	IELTS	13	6.89	0.53	6	8	.21	-.36	.66	.481	.04
		GPA	3.92	0.11	3.65	4.00					
	TOEFL	13	101	14.5	68	117	.41	-.16	.77	.147	.17
		GPA	3.71	0.32	3.00	4.00					
SH	DET	8	114	13.3	95	130	-.67	-.93	.07	.069	.45
		GPA	3.88	0.19	3.45	4.00					
	IELTS	8	7.22	0.57	6	8	.68	.02	.92	.045*	.46
		GPA	3.86	0.12	3.67	4.00					
	TOEFL	21	95.5	17.7	64	116	.33	-.12	.66	.149	.11
		GPA	3.67	0.64	1.45	4.00					
Undergraduate only											
LS	DET	7	129	12.1	115	150	.63	-.24	.94	.132	.39
		GPA	3.38	0.59	2.41	3.95					
	IELTS	3	7	1.50	5.5	8.5	.81			.394	.66
		GPA	3.26	1.06	2.03	3.97					
	TOEFL	2	86	2.83	84	88	not enough observations				
		GPA	3.51	0.59	3.10	3.93					
PE	DET	5	117	16.4	95	135	.55	-.64	.96	.333	.31
		GPA	3.59	0.29	3.31	3.96					
	IELTS	2	7.75	1.06	7	8.5	not enough observations				
		GPA	2.45	1.61	1.31	3.59					
	TOEFL	2	86	29.7	65	107	not enough observations				
		GPA	3.87	0.06	3.83	3.91					
SH	DET	16	109	13.1	90	130	.35	-.18	.72	.190	.12
		GPA	3.15	0.60	1.91	3.87					
	IELTS	4	6.5	0.91	5.5	7.5	.95	-.09	1.00	.047*	.91
		GPA	3.43	0.32	3.00	3.71					
	TOEFL	5	68	26.8	30	104	.42	-.49	.89	.350	.18
		GPA	3.49	0.50	2.85	3.97					

APPENDIX B

Correlations between ELP test scores and first year GPA, Fall 2022 cohort

	ELP							95% CI			
Level	test	n	M	SD	min	max	r	LL	UL	p	R^2
Students admitted Unconditionally											
All	DET	20	131	6.34	125	150	.35	-.11	.69	.132	.12
		GPA	3.67	0.34	2.62	4.00					
	IELTS	23	7.33	0.49	7.0	8.5	.07	-.36	.47	.763	.00
		GPA	3.77	0.55	1.31	4.00					
	TOEFL	25	108	4.67	101	117	.06	-.34	.43	.789	.00
		GPA	3.79	0.53	1.45	4.00					
GR	DET	9	131	4.64	125	140	.23	-.51	.78	.543	.06
		GPA	3.74	0.27	3.18	4.00					
	IELTS	17	7.24	0.36	7.0	8.0	.15	-.36	.59	.562	.02
		GPA	3.92	0.09	3.72	4.00					
	TOEFL	23	108	4.79	101	117	.07	-.35	.46	.754	.00
		GPA	3.78	0.55	1.45	4.00					
UG	DET	11	131	7.69	125	150	.40	-.26	.81	.221	.16
		GPA	3.61	0.39	2.62	3.96					
	IELTS	6	7.58	0.74	7.0	8.5	.33	-.65	.90	.517	.11
		GPA	3.33	1.00	1.31	3.97					
	TOEFL	2	106	2.12	104	107	not enough observations				
		GPA	3.9	0.10	3.83	3.97					
Students admitted Conditionally through the ELI											
All	DET	25	107	10	90	120	.12	-.28	.50	.548	0.01
		GPA	3.35	0.625	1.91	4					
	IELTS	11	6.09	0.38	5.5	6.5	.58	.05	.86	.036*	0.34
		GPA	3.51	0.57	2.03	4					
	TOEFL	27	80.1	15.3	30	98	.52	.19	.75	.004**	0.27
		GPA	3.63	0.427	2.71	4					
GR	DET	8	110	9.64	95	120	-.38	-.86	.44	.353	0.14
		GPA	3.95	0.0865	3.74	4					
	IELTS	8	6.25	0.27	6	6.5	.22	-.47	.75	.540	0.05
		GPA	3.78	0.15	3.65	4					
	TOEFL	20	83.1	11.7	64	98	.63	.26	.84	.003**	0.40
		GPA	3.67	0.416	2.71	4					
UG	DET	17	106	10.1	90	120	.00	-.48	.48	.994	0.00
		GPA	3.08	0.569	1.91	3.73					
	IELTS	3	5.67	0.29	5.5	6	.72	--	--	.489	0.52
		GPA	2.8	0.69	2.03	3.38					
	TOEFL	7	71.7	21.6	30	96	0.28	-.47	.80	.468	0.08
		GPA	3.49	0.462	2.85	3.93					

APPENDIX C

One-way ANOVA of mean first year GPA by ELP test taken, Fall 2022 cohort only, aggregated and separated by level (graduate (GR) and undergraduate (UG))

Level of study	ELP test	n	First year GPA M	SD	SE	95% CI LL	UL	F	p	DF	R^2
Both levels	no ELP	116	3.46	0.71	0.07	3.33	3.59	2.54	.057	243	0.03
	DET	45	3.49	0.54	0.08	3.34	3.65				
(GR & UG	IELTS	34	3.69	0.56	0.10	3.50	3.87				
combined)	TOEFL	52	3.7	0.48	0.07	3.57	3.83				
GR only	no ELP	50	3.77	0.39	0.06	3.66	3.88	0.92	.435	131	0.02
	DET	17	3.84	0.23	0.05	3.73	3.94				
	IELTS	25	3.88	0.13	0.03	3.83	3.93				
	TOEFL	43	3.73	0.49	0.07	3.58	3.88				
UG only	no ELP	66	3.23	0.80	0.10	3.03	3.42	0.70	.555	108	0.02
	DET	28	3.29	0.57	0.11	3.08	3.50				
	IELTS	9	3.16	0.90	0.30	2.57	3.75				
	TOEFL	9	3.58	0.44	0.15	3.29	3.87				

TEST REVIEW: WIDA SCREENER

SOHYEON LEE

Department of Second Language Studies, University of Hawai'i at Mānoa

INTRODUCTION

Over the past two decades, there has been a significant increase in the number of multilingual students in U.S. schools from kindergarten to Grade 12. This trend is expected to continue, with projections suggesting that 40% of all students could come from families where English is not the first language (Fu, 2009; Kim et al., 2018). To support these multilingual learners and their educators, the World-Class Instructional Design and Assessment (WIDA) Consortium and the Center for Applied Linguistics have developed the Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs). This is a comprehensive, standards-based English language proficiency test. Additionally, the WIDA Screener, which is available in both paper and online formats, serves as a shorter version of ACCESS for K-12 English Language Learners (ELLs). My review will focus exclusively on the WIDA Screener Online, considering its practicality, usability, and validity. By evaluating the WIDA Screener, particularly its online version, I aim to substantiate its effectiveness. This, in turn, will illustrate how it can be a valuable resource for both students and educators.

TEST REVIEW

Test Purpose

The WIDA Screener is an English language proficiency assessment specifically designed for newly enrolled students in grades K-12. Unlike ACCESS, which is used for both placement and achievement evaluation, the WIDA Screener focuses solely on the identification and placement of ELLs. While it offers a more streamlined approach compared to ACCESS, the Screener plays a crucial role in helping educators determine whether a new student requires additional language support.

WIDA has developed a set of English Language Development (ELD) standards aimed at promoting equity for multilingual learners in curriculum, instruction, and assessment. These standards address several key areas. The Social Instructional Language standard evaluates the ability to use English in social contexts and for instructional purposes, such as following directions and engaging in classroom discussions. The Language of Language Arts standard focuses on skills related to language arts, including reading comprehension, writing, and literary analysis. The Language of Math emphasizes understanding and communicating mathematical concepts, as well as reasoning and problem-solving. The Language of Science assesses language skills necessary for interacting with scientific content, including experiment descriptions and scientific discussions. Lastly, the Language of Social Studies standard gauges language proficiency in social studies topics, covering historical texts, geographical concepts, and civics and economics discussions. These diverse standards reflect WIDA's holistic approach to language development, addressing both academic and social language skills that are vital for the progress of multilingual learners (Wida Consortium, 2020).

While the WIDA Screener aligns with the WIDA ELD Standards, it primarily assesses general language proficiency, rather than directly evaluating language skills in each specific academic content area (Macgregor & Sahakyan, 2020). The ELD Standards provide a framework for English language instruction and guide curriculum development and instructional practices across various content areas. Although integral to WIDA's philosophy, these standards play a more significant role in ongoing instruction and formative assessments within classrooms, rather than in initial placement assessments like the WIDA Screener (Wida Consortium, 2020).

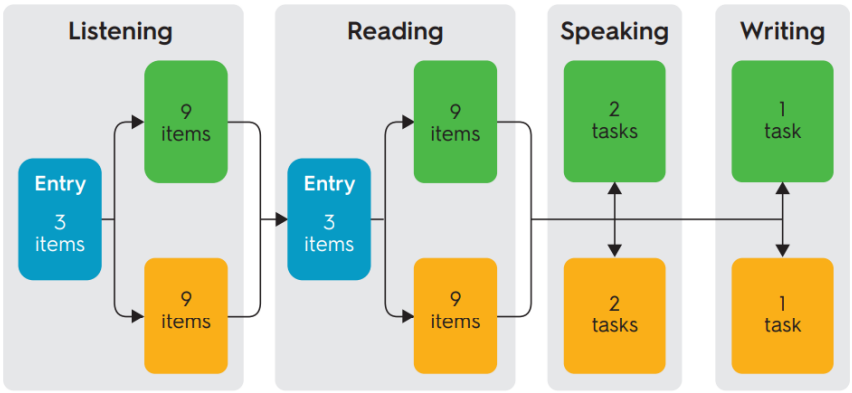
Nonetheless, by incorporating these varied standards, the WIDA Screener ensures a comprehensive assessment of a student's English language proficiency, addressing both academic and social language skills crucial for determining the appropriate level of language support needed.

Test Method

The WIDA Screener is divided into five grade-level clusters: 1, 2-3, 4-5, 6-8, and 9-12. Each cluster includes four test sections: Listening, Reading, Speaking, and Writing. The Listening and Reading sections feature an adaptive testing design with three groups of test items. Students initially encounter a set of items, and their performance on these determines whether they are directed to an easier or more challenging group of items, as depicted in Figure 1. This adaptive mechanism ensures a tailored assessment experience, reflecting each student's English proficiency level.

Figure 1

Overview of the Test Design of Screener Online. Adapted from "Examining the relationship between the WIDA Screener and ACCESS for ELLs assessments," by D. Macgregor and N. Sahakyan, 2020, in WIDA Technical Report (p.5).



The tasks' level in the Speaking and Writing sections is determined by the student's scale scores on the Listening and Reading tests (Macgregor & Sahakyan, 2020). This approach maintains a consistent difficulty level across all test sections, based on the student's demonstrated language skills. The entire Screener Online test is flexible in scheduling, with a total test duration of about 80 minutes. The Reading and Listening sections are allocated 15 minutes each, while the Speaking and Writing sections are given a longer duration of 25 minutes each, considering the different complexities of these skills.

Furthermore, WIDA offers teachers and decision-makers access to a test demo and a practice test through a portal link (available via WIDA customer service). These resources

provide step-by-step simulations with model answers for each test item. For example, in a speaking task simulation, a teacher on a computer screen interacts with a student, demonstrating what is expected in a student's response (see Appendix A). Additionally, the screener evaluates language proficiency in subject matter contexts. Some items are specifically designed to assess classroom language skills for content-based learning, like a listening item that combines understanding reading graphs with listening comprehension (see Appendix B).

The WIDA Screener Interpretive Guide for Score Reports outlines the scoring and calculation of test items. Listening and Reading scores are computed automatically online, producing raw scores based on the number of correct items. For the Speaking and Writing tests, assessments are conducted by local school or district staff, referred to as 'local raters,' using the WIDA Screener Scoring Scales. Once the speaking and writing scores are entered into the WIDA Assessment Management System, the system generates speaking and writing scores as well as composite scores.

Scores

According to the WIDA Screener Interpretive Guide for Score Reports, the scoring process involves converting raw scores into scale scores using statistical methods to account for varying test difficulties. To facilitate easier interpretation, these scale scores are then translated into proficiency levels. The reports provide proficiency level scores for each language domain (Listening, Speaking, Reading, Writing) and three composite scores: Oral Language, Literacy, and Overall. The Oral Language score is calculated as the average of the Listening and Speaking scores, while the Literacy score is the average of the Reading and Writing scores. For the Overall score, a weighted average of all four domain scores is computed, with Listening and Speaking each contributing 15%, and Reading and Writing contributing 35% each. This weighting reflects the relative importance of each skill in overall language proficiency.

The same proficiency level cut scores are applied in both the WIDA Screener and ACCESS for ELLs. These cut scores delineate the boundaries between proficiency levels and are defined for each grade based on expert judgments during a standard-setting process. To assist parents in understanding their children’s proficiency levels, scores are aligned with the six WIDA English language proficiency levels: Entering, Emerging, Developing, Expanding,

Bridging, and Reaching, where 'Entering' is the initial level, and 'Reaching' is the highest achievable level.

While ACCESS is primarily a criterion-referenced test, focusing on assessing ELL students' achievement against set standards, the Screener incorporates aspects of both norm-referenced and criterion-referenced testing. Its norm-referenced nature is evident in its use for initial placement decisions (Brown, 2005), assessing general language proficiency among ELL students. However, it also functions as a criterion-referenced test, with proficiency levels based on the WIDA standards serving as criteria. Thus, each proficiency level acts as a criterion to determine students' language proficiency status.

Cost and Publisher

The WIDA Screener is available free of charge, offering a significant advantage to members of the WIDA consortium. This accessibility enables schools and educators to use the assessment tool flexibly throughout the year, without financial constraints. WIDA is situated within the Wisconsin Center for Education Research at the University of Wisconsin-Madison. For any inquiries or additional support, the WIDA Client Services Center can be reached via email or phone, providing users with ready assistance.

Moreover, the WIDA website is an encompassing resource for information regarding the Screener and other related tools. The website, which can be accessed at http://wida.wisc.edu, is designed to be user-friendly and informative, making it easier for educators and decision-makers to navigate and efficiently find necessary information.

Validity

Substantial evidence supporting the reliability of ACCESS has been documented in numerous reports. This includes detailed information on the assessment's conceptualization and the procedures for standard-setting. One notable aspect is the conversion of grade-level cluster cut scores to grade-specific scores, which has been demonstrated to have high reliability rates for composite scores (Fox & Fairbairn, 2011). A more recent Technical Report by Macgregor & Sahakyan (2020) further strengthens this evidence by examining the relationship between the WIDA Screener and the ACCESS for ELLs assessment. The study found significant correlations

between the overall composite scale scores of the Screener and ACCESS across all grades, with correlations ranging from 0.67 for Grade 1 to a high of 0.86 for Grades 7-9. This indicates a strong alignment between the two assessments, underscoring the Screener's validity.

Furthermore, the Screener's proficiency score has been identified as a critical factor in making identification and placement decisions for students (Kim et al., 2018). This highlights the Screener's practical value, especially in scenarios where new students from diverse backgrounds need to be quickly and accurately identified as ELLs and placed in appropriate language support programs. The fact that the Screener is freely available to states and territories enrolled in the WIDA Consortium and can be accessed virtually anytime enhances its practicality. Its relatively shorter duration, requiring about 80 minutes as opposed to the 265 minutes needed for ACCESS, facilitates a more efficient assessment process. This efficiency is particularly valuable for schools in managing and identifying the language proficiency status of new students. Moreover, WIDA provides extensive support and guidelines for teachers and administrators on their website, including resources like 'suggested sample item scripts', 'interpretive guides for score reports', and various workshops, further cementing the Screener's role as a vital tool in the educational landscape.

Areas for Improvement

The WIDA Screener Online provides proficiency scores across different domains, but as presented in Appendix C, the score report lacks detailed explanations regarding an ELL student's practical English usage. While an interpretive guide for score reports is available, it does not sufficiently clarify a student's proficiency in practical terms. For instance, terms like 'Bridging' and 'Expanding' in WIDA proficiency levels are not immediately clear, leaving parents and educators uncertain about the specific language skills their students have or need. Questions like whether a child needs more vocabulary development or whether they can understand narrative stories but struggle with academic writing remain unanswered. This lack of detail makes it challenging to fully understand a child's language proficiency.

Furthermore, in the context of ELL placements, there is a lack of specific guidelines on how to use and integrate Screener scores with other data sources. For instance, the process of incorporating ACCESS scores, Screener results, home language surveys, or parent interviews in

the decision-making process at the school or district level is not clearly outlined, as noted by Kim et al. (2018). This ambiguity raises concerns about the effectiveness of the Screener as a standalone measurement tool. In practice, combining Screener scores with these additional data sources could offer a more comprehensive understanding of a student's English language proficiency and needs. However, without clear guidance on integrating these different sources, educators may struggle to make well-informed placement decisions. Conducting thorough studies on this matter would be beneficial for future research, potentially leading to more structured guidelines for integrating various assessment tools and data in ELL placements.

Additionally, the accessibility of the Screener poses a limitation. While it is freely available to WIDA Consortium members, educators and students in non-member states are excluded from using this resource. This restriction is particularly problematic for states with fewer ELLs, where funding for such assessments might not be prioritized. Denying access to non-members seems neither equitable nor educationally sound. Offering the Screener, even for a fee, to non-member states would be a more inclusive approach. Also, providing additional or differentiated assessment materials for member states with smaller ELL populations would ensure appropriate support (Kim et al., 2018).

The need for a more transparent and comprehensive approach to interpreting Screener results, coupled with the accessibility issue, highlights areas for improvement. Addressing these concerns could significantly enhance the utility and fairness of the WIDA Screener in the assessment and placement of ELL students.

CONCLUSION

In conclusion, the process of ELL identification and placement involves considering a myriad of factors, including student English language proficiency scores, home language surveys, academic achievements, and parent interviews. The WIDA Screener has been established as a reliable tool in this context, demonstrating a high correlation with ACCESS and serving as a robust predictive source for ELL identification and placement (Kim et al., 2018; Macgregor & Sahakyan, 2020). However, the WIDA Consortium cautions against relying solely on the Screener for making these critical decisions. This recommendation underscores the

complexity of language learning and the necessity of a holistic approach to the educational decision-making process.

Furthermore, it is emphasized that those responsible for ELL identification and placement should possess a deep understanding of second language acquisition and teaching practices. This is particularly crucial given that these decisions are often made by a single individual or a small group of educators, which can limit the diversity of perspectives in the decision-making process (Kim et al., 2018).

Despite these considerations and the need for a multifaceted approach, the WIDA Screener Online stands as a validated and valuable instrument. Its expanding use is a testament to its effectiveness in aiding educators and decision-makers. Moving forward, it is essential to continue refining and supplementing this tool with additional resources and training for educators. This will ensure that it not only remains a robust instrument for assessment but also becomes an integral part of a comprehensive strategy for effectively supporting the diverse needs of ELL students.

REFERENCES

Brown, J. D. (2005). *Testing in language program: A comprehensive guide to English language assessment* (1st ed.). McGraw-Hill.

Fox, J., & Fairbairn, S. (2011). Test review: ACCESS for ELLs®. *Language Testing*, *28*(3), 425–431. https://doi.org/10.1177/0265532211404195

Fu, D. (2009). New Chinese immigrant students' literacy development: From heritage language to bilingualism. In J. C. Scott, D. Y. Straker & L. Katz (Eds.), *Affirming students' right to their own language: Bridging language policies and pedagogical practices* (pp. 247-261). Routledge.

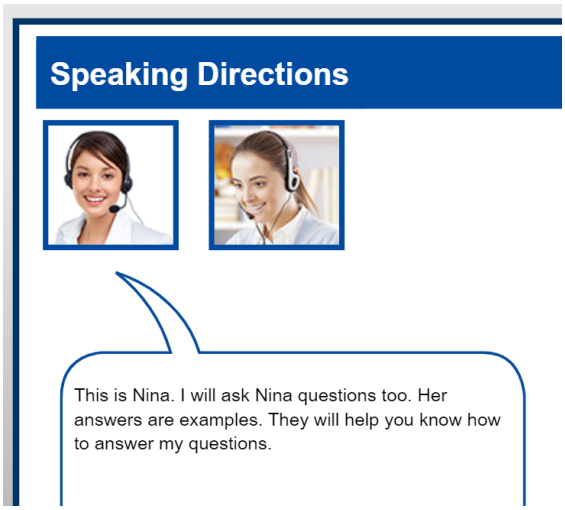
Kim, A. A., Molle, D., Kemp, J., & Cook, H. G. (2018). Examination of Identification and Placement Decisions Made for K-12 English Learners. WCER Working Paper No. 2018-12. *ERIC*. https://eric.ed.gov/?id=ED593631

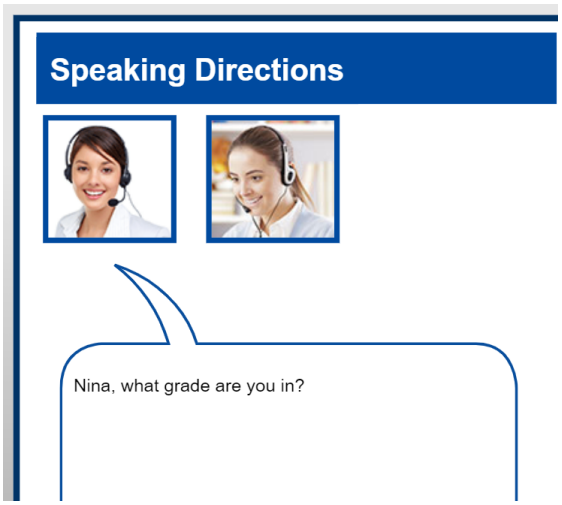
Macgregor, D., & Sahakyan, N. (2020). *Examining the relationship between the WIDA Screener and ACCESS for ELLs assessments*. https://wida.wisc.edu/sites/default/files/resource/technical-report-examining-relationship-between-WIDA-screener-ACCESS.pdf

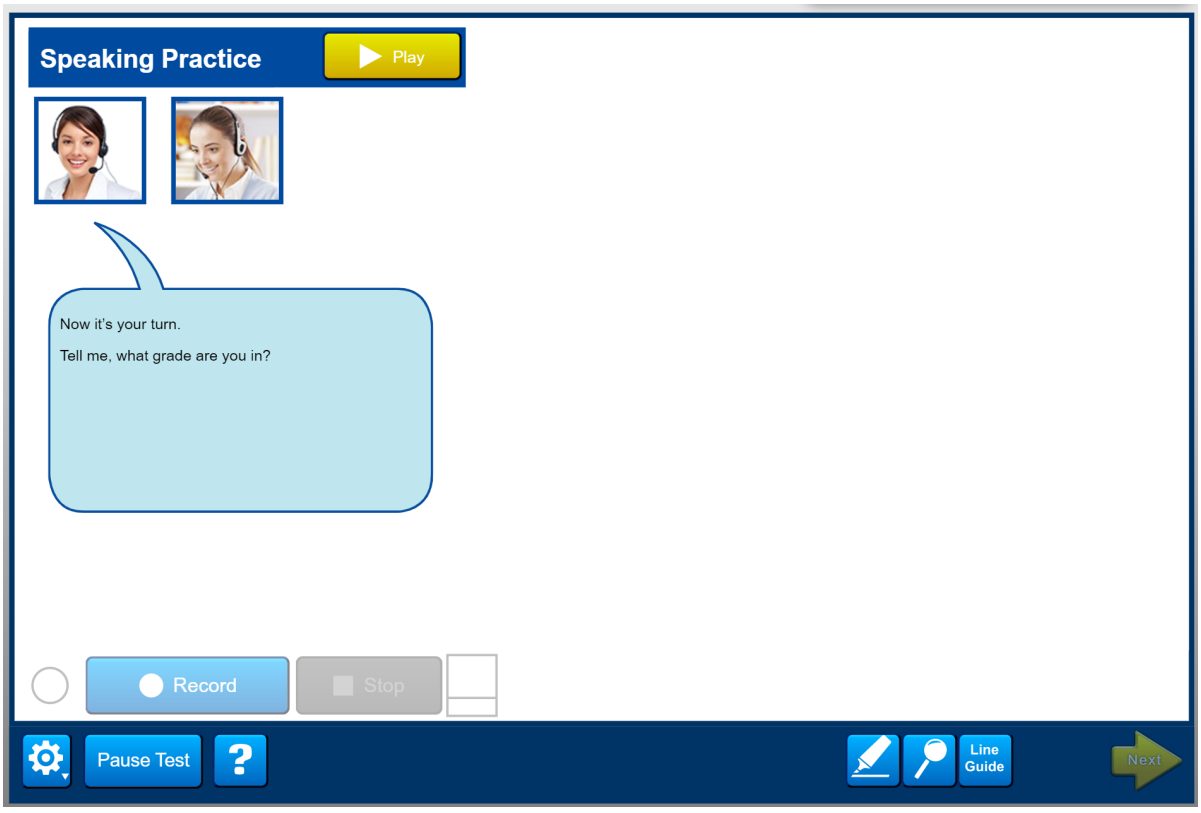
WIDA Consortium. (2020). *WIDA English Language Development Standards Framework, 2020 Edition: Kindergarten–Grade 12*. Board of Regents of the University of Wisconsin System.

WIDA. (n.d.). *WIDA Screener Interpretive Guide for Score Reports Grades 1-12.* Retrieved from https://wida.wisc.edu/sites/default/files/resource/WIDA-Screener-Interpretive-Guide.pdf

APPENDIX A

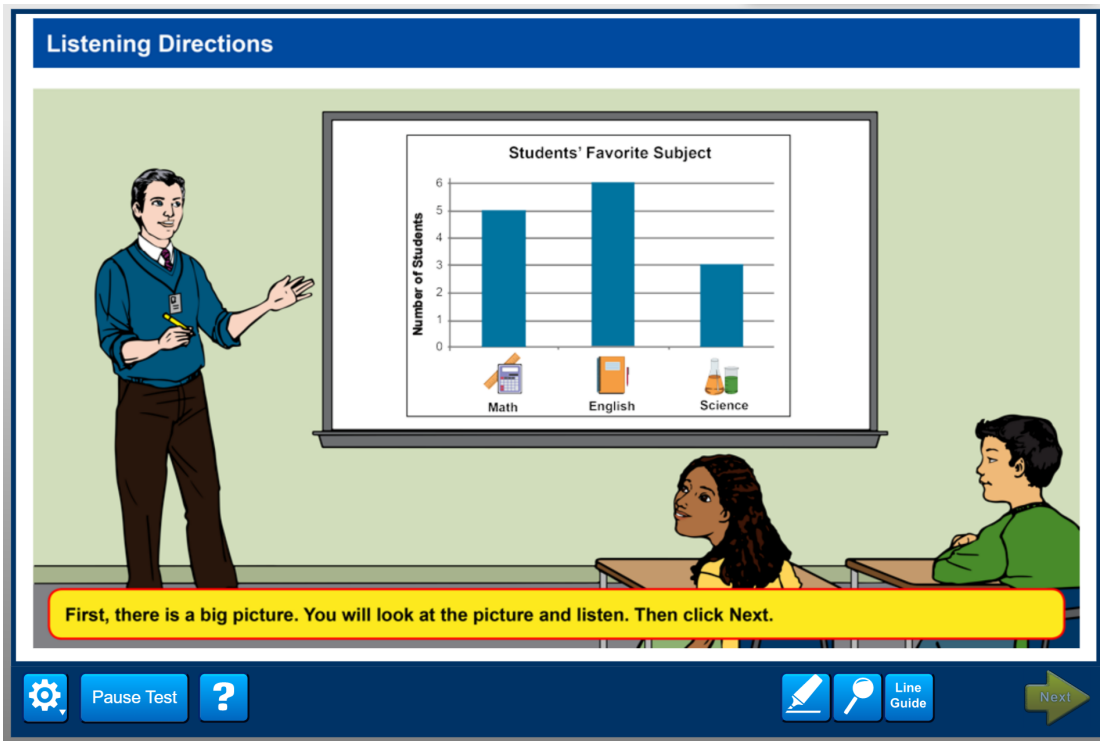

Speaking Directions
This is Nina. I will ask Nina questions too. Her answers are examples. They will help you know how to answer my questions.

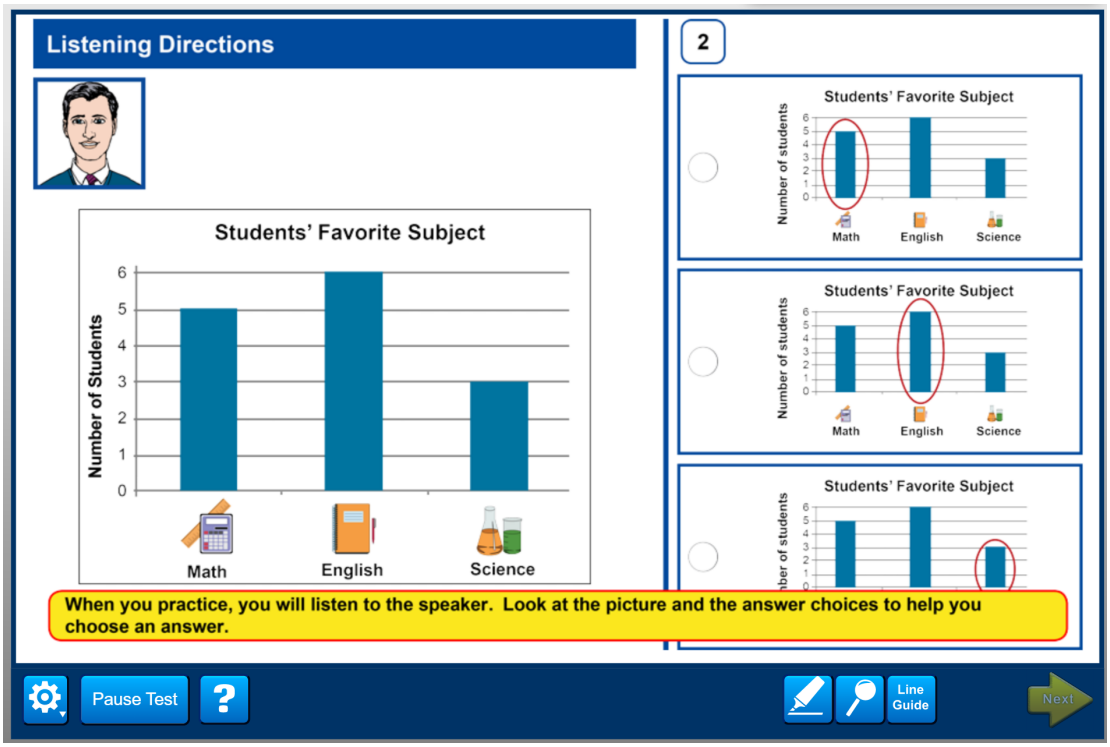

Speaking Directions
Nina, what grade are you in?


Speaking Practice
Play
Now it's your turn.
Tell me, what grade are you in?
Record
Stop
Pause Test
Line Guide
Next

Appendix B. WIDA Screener Practice Test Listening Section

(Retrieved from WIDA secure portal)





Appendix C. WIDA Screener Online Score Report

(Retrieved from WIDA Screener Interpretive Guide for Score Reports)

WIDA Screener Online Score Report

WIDA™

WIDA Screener Online

Score Report

Test Date: 09/29/2021

Test Administrator/Scorer:

Student Information

First Name: Student
Last Name: Name 4
Birthdate: 05/05/2011
Current Grade: 05
State ID: 3456546880
School: WIDA Demo School 1
District: WIDA Demo District 1
State: ST
Home Language(s):

This report provides information about the student's scores on the WIDA Screener Online. The assessment is designed to provide a snapshot of an English language learner's general English language proficiency. This test is based on the WIDA English Language Development Standards and is used as one criterion to determine a student's eligibility for English language services. Please refer to state policy for making decisions about a student's eligibility for English language services.

Scores are reported as language proficiency levels. Proficiency levels (PLs) are reported as whole numbers for the language domain scores and the composite scores are reported to the whole PL or half PL.

Language Domain Scores	Proficiency Level
Listening	4.0
Speaking	2.0
Reading	3.0
Writing	3.0
Composite Scores	**Proficiency Level**
Oral Language	3.0
Literacy	3.5
Overall*	3.0

*Overall composite score is calculated only when all four domains have been assessed.

145200-000004-33521

Screener-ISR

PHENOMENA ASSOCIATED WITH LINGUISTIC RELATIVITY AND COGNITIVE TRANSFER: DRAWING AND PAINTING ACTIONS IN GERMAN-ENGLISH BILINGUALS

MICHAELA NUESSER

Department of Second Language Studies, University of Hawai'i at Mānoa

ABSTRACT

This study investigates whether linguistic relativity and conceptual transfer extends to categorization preferences of drawing- and painting actions in German and English monolingual speakers as well as German–English bilinguals. In alignment with the hypotheses, a triad similarity judgment revealed significant differences in separations among all three groups between stimuli showing two types of painting actions which are commonly distinguished in German but not in English. Bilinguals showed categorization preferences of both monolingual groups. However, separations between stimuli showing drawing- and painting actions, two verbs commonly distinguished in English but not in German, were not significantly different between groups. This is likely due to a limitation of the task itself. The study has implications for the study of both linguistic relativity and conceptual transfer which had not been shown for the partial (in-) equivalence among drawing- and painting actions before. In contrast to previous studies, participants had the option to arrange stimuli on a scale that allowed for more than just an odd-one-out arrangement, not restricting the participants to one of two outcomes predefined by the researcher. The results can be interpreted as evidence that speakers may be guided by their native categorization habits but not restricted to them and that categorization preferences can be influenced by habits from two speech communities at the same time.

Keywords: Linguistic Relativity; Whorf; Conceptual Transfer; cross-linguistic differences; conceptual overlap

INTRODUCTION

In this study, I investigated whether phenomena typically associated with linguistic relativity and conceptual transfer (Whorf, 1956; Bylund & Athanasopoulos, 2014; Park and Ziegler, 2014) extend to categorization preferences of drawing and painting actions in German and English speakers.

Linguistic relativity, most famously coined by Benjamin Lee Whorf, claims that speakers of different languages are pointed toward different aspects of the world through being part of their speech communities (Whorf, 1956, p. 221). Prime examples of this study are object categorization studies (Malt et al., 2003), studies of motion event perception (Bylund & Athanasopoulos, 2014), and studies of placement awareness (Koster & Cadierno, 2018).

The study of conceptual transfer concerns language learners who, because of linguistic relativity phenomena, may be required to think about the world in new ways. As learning a foreign language may co-occur with continuing native ways to think about the world, Jarvis (2016, p. 608) defines conceptual transfer as "cross-linguistic influence in the expression and interpretation of conceptual meaning". The field aims to investigate whether cross-linguistic influence in categorization preferences is observable in people who know more than one language.

Signs of phenomena commonly associated with conceptual transfer have been found in intermediate learners especially, while advanced learners often—but not always as Gullberg (2009) found—seem to show behavior that is closer to that of native speakers on experimental tasks (Park & Ziegler, 2014; Bylund & Athanasopoulos, 2014).

BACKGROUND

Verb Categorization and Linguistic Relativity

The focus domain of this study is drawing and painting actions. Though most research in the field has been carried out in the domains of object categorization and motion events, select studies have investigated whether linguistic relativity and conceptual transfer phenomena extend to other action categorization like that of placement events (van Bergen & Flecken, 2017),

throwing actions (Nicoladis & Gao, 2021), and "putting in" vs "putting on" (Park & Ziegler, 2014).

Park and Ziegler (2014) asked Korean and English monolinguals as well as Korean-English bilinguals to choose the odd-one-out of stimuli picture triads showing "put in" and "put on" actions. The two languages differ in how they indicate and group this direction. Results revealed phenomena associated with linguistic relativity as Korean and English monolinguals demonstrated statistically significant differences in choices. The bilinguals with the lowest proficiency behaved most similarly to the Korean monolingual group, which demonstrates phenomena associated with conceptual transfer among the participants. Frequency of Korean use negatively correlated with English monolingual-like behavior on the task.

Nicoladis and Gao (2021) investigated Mandarin-English bilinguals and how they refer to Mandarin prototypical throwing actions in English and Mandarin. Participants labeled video clips that showed prototypical examples of throwing actions habitually referred to with distinct words by Chinese speakers, but difficult to label in English. Nicoladis and Gao found that bilinguals chose a significantly larger variety of labels than English monolinguals. After analysis of patterns, the authors suggest that bilinguals showed several strategies to express what is not translatable to English. Nicoladis and Gao interpreted the findings as evidence of phenomena associated with conceptual transfer in throwing actions.

Van Bergen & Flecken (2017) explored German, Dutch, English, and French native speakers' anticipatory eye movements when listening to Dutch sentences in correlation to whether their native language specifies position in placement verbs (e.g. "put into standing position" vs. "put into lying position", which is common in German and Dutch but not in English and French). All non-Dutch native speakers were learners of Dutch. The authors' eye-tracking results showed consistent related anticipatory eye movements in German and Dutch native speakers when listening to Dutch sentences, whereas English and French native speakers did not predict placement position once they heard the verb in the recorded Dutch sentences. All participants demonstrated proper understanding of Dutch placement verbs. The findings can thus be interpreted as signs of both linguistic relativity and conceptual transfer in placement events.

The Current Study

There is, to the best of my knowledge, no study to date that investigates cross-linguistic differences in cognitive preferences within the lexical realm of actions like painting and drawing. English and German differ with regard to how they categorize painting and drawing actions: drawing on my experience in both languages, I argue that the two English verbs "draw" and "paint" both commonly appear in similar contexts as the German word "malen", except when "paint" is used to refer to coating something in paint as in painting a wall or furniture, in which case "streichen" appears more commonly in German. Following Pavlenko (2009), terms with this relationship are called partial non-equivalents.

Using photograph stimuli depicting realistic drawing and painting actions, I investigated how functional monolinguals of English, functional monolinguals of German, and advanced German–English bilinguals categorize drawing and painting actions. I recorded their similarity judgments when asked to match three different stimuli pictures at a time. The task itself did not require any overt language use. The study included functional monolingual groups as common in linguistic relativity research and a German–English bilingual group, situated in the field of conceptual transfer.

Following review of the literature, I identified the following exploratory research questions: 1) Do previously attested linguistic relativity phenomena extend to the categorization of painting and drawing actions in English and German monolinguals? 2) Do German–English sequential bilinguals show evidence of conceptual transfer when categorizing painting and drawing actions?

As in Park & Ziegler (2014), differences in similarity judging behavior between two monolingual groups is interpreted as a sign of linguistic relativity (see RQ1). Following Jarvis (2016), differences in similarity judging behavior on the nonverbal tasks comparing bilinguals and their native language's monolingual group is interpreted as a sign of conceptual transfer (see RQ2).

METHODS

Participants

In line with sample sizes of related studies (Cook et al., 2006: n = 36; Park & Ziegler, 2014: n = 17–31; Stepanova & Coley, 2002: n = 22), 32 native English and functional monolinguals from the community of a university in the USA and 32 native German and functional monolinguals studying at or residing around a university in Germany took part in this experiment. Even though all participants had experience learning a second language in formal education, they self-reported a proficiency of 4 out of 10 or lower in English or German as a second language. After exclusion due to all filler triads being answered incorrectly or the report of advanced proficiency in second languages, 24 English monolinguals and 25 German monolinguals were included in the analysis. The groups were recruited to address the first research question of the study.

Thirty-four German–English sequential bilinguals were recruited from the two communities mentioned above. Bilinguals all had at least 10 years of experience learning English, reported a total average proficiency of 7.3, and indicated an average of 32.6 hours of exposure to English media per week. Their average score on the online available 5-min LexTale English proficiency test was 69.5%.[1] For comparison, the creators of the LexTale indicate that an advanced group of Dutch and Korean English learners averaged at 70.7% (Lemhöfer & Broersma, 2012). 5 bilingual participants were tested in Hawai'i, all other bilingual participants were tested in Germany. After exclusion due to German not being a participant's first language, 31 German–English bilinguals were included in the analysis. The bilingual group was recruited to address the second research question of the study. The samples are convenience samples. Participants were offered course credit or $10/10€ in reward for their time and consented to participate.

Materials

The experimental materials for the similarity judgment task were 12 critical triads of picture stimuli. The experiment also included 12 filler triads to interrupt similar looking triads.

[1] The LexTale has proven to be a valid and standardized predictor of English proficiency that exceeds self-rating scores in terms of correlation with other proficiency scores (Lemhöfer & Broersma, 2012).

Critical triads showed painting- or drawing situations, filler triads showed other crafting or work done by hand. Critical item sets were designed to yield two different interpretations so that participants' tendencies to German habits (categorizing artistic painting and drawing as different from covering an object or wall in paint) would prompt an answer that is different from that of those participants that tend to English habits of discriminating (categorizing painting objects, walls, or paintings as different from drawing). Examples of expected results are shown in Figures 1 and 2. The 12 filler triads only had one obvious correct answer in which two stimuli were more similar to each other than the third.

Figure 1

Expected Arrangement by German Monolinguals: Drawing/malen and painting/malen are shown to be more similar to each other than to painting/streichen

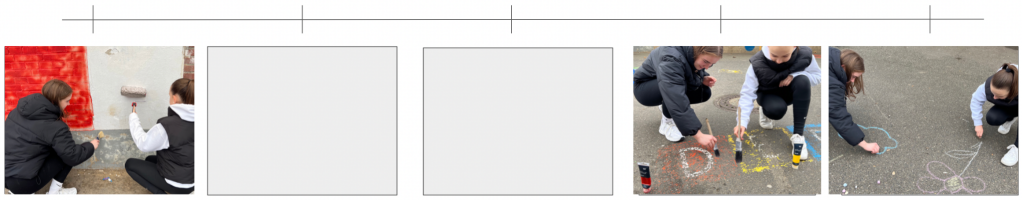
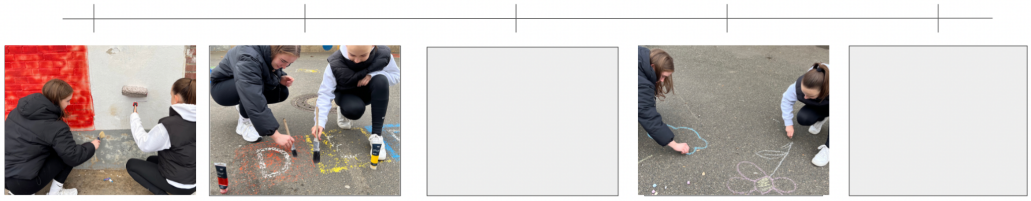


Figure 2

Expected Arrangement by English Monolinguals: Painting/malen and painting/streichen are shown to be more similar to each other than to drawing/malen



The stimuli were presented on an editable Google slideshow displaying the picture triad at the top and a scale with five marks, each complemented by a rectangle underneath, for possible stimuli placement. To avoid results based on the influence of the order of trials, participants were assigned to see one of 12 lists of triads. Each list had a pseudo-random order allowing for critical and filler trials to be alternating. To avoid results based on the influence of the order the three stimuli within each triad are presented in, their order was randomized to be different in each list.

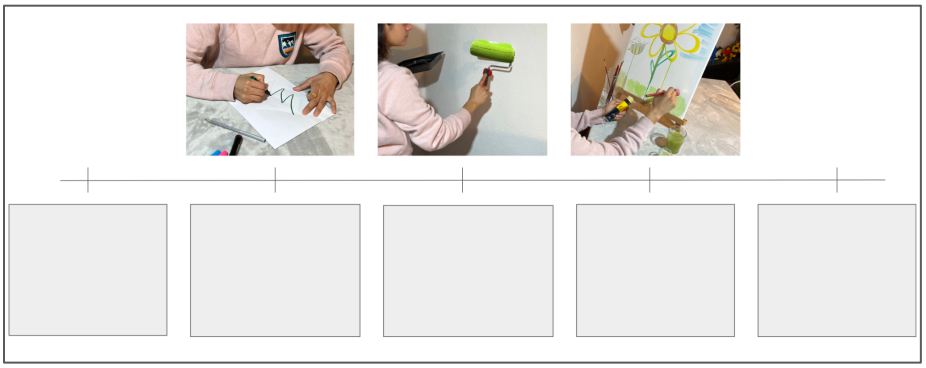
A list of all 24 critical and filler triads, the raw data spreadsheet, and the R code used to analyze the data are available in the OSF repository of this project: https://osf.io/bfdcp/?view_only=cb01ecab387e4e1e81760a1391e29b36.

Procedure

Participants were tested individually in 20-min sessions in a quiet room in Germany or in a lab at the university of the researcher. Before the task, I asked participants to complete the English LexTale test (Lemhöfer & Broersma, 2012). This was not only to solidify group placement and instructional language choice but also to introduce bilingual participants into an English-speaking environment. Then, participants were presented with the slideshow, first showing the scale of 5 marks and boxes, and then containing 24 slides with the experimental triads. According to the instructions, participants dragged each picture into a box on the scale representing the similarity of the actions portrayed. Participants were not encouraged to speak aloud, and the researcher never mentioned the lexical items in question. With all participants conducting the same activity, this experiment addresses both research questions. Figure 3 shows a sample slide from the experiment. Backtracking was not allowed during the task to discourage changing previous answers.

Figure 3

Experiment Interface



Analysis

As in Stepanova and Coley (2002), the investigated dependent measures are separations of stimuli. Since there are two distinctions of interest, I investigated both separations of drawing- and painting stimuli and streichen and malen stimuli. The dependent variable, in either method of analysis, is not to be understood as a measure of correctness. The chosen by placement on the scale intends to measure perceived similarity of drawing- and painting related actions. Width of separations, i.e., whether stimuli were placed with one or two empty boxes in between, was not regarded to avoid overcomplication of the analysis. The independent variable is group membership. Using an alpha level of 0.05, I conducted two logistic mixed effects models with both participant and trial as random effects using the R function glmer. One model was fitted to predict separations of drawing- and painting stimuli, the other one to predict separations of streichen and malen stimuli. By default, the model uses the English monolingual group as the comparison group. A multinomial logistic regression model was not adequate as the task allowed participants to separate all stimuli or to keep both pairs together. *Emmeans* was used for post-hoc comparisons. It applies the tukey method for p-value adjustment for comparing a family of 3 estimates.

RESULTS

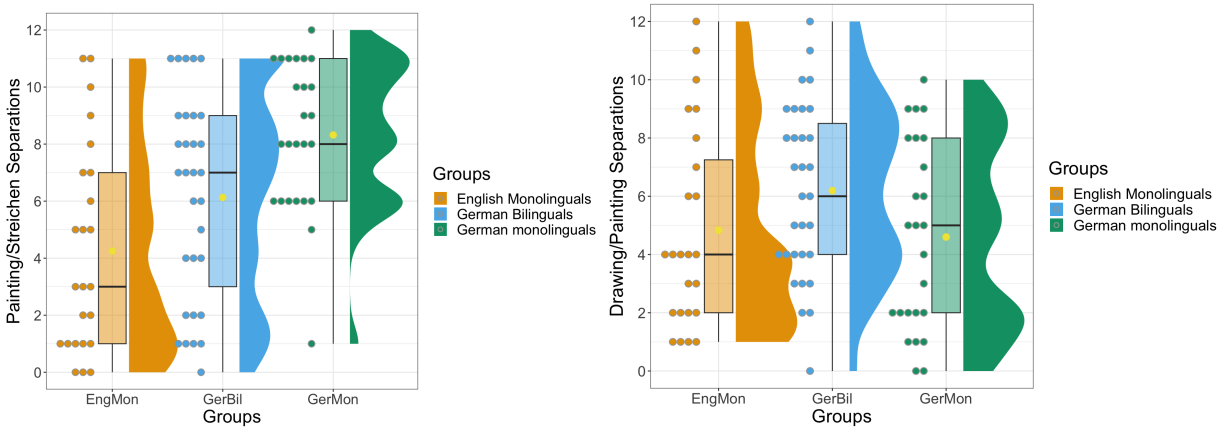
Results are expressed in numbers of separations per participant. Participants solved filler items, which had a correct and incorrect option, at near ceiling in terms of accuracy. Figure 4 shows the distribution of separations in raincloud plots per group. Means are indicated by yellow dots. Drawing- and painting stimuli were separated similarly by all groups: German functional monolinguals separated drawing- and painting stimuli on average 4.6 out of 12 times, English functional monolinguals separated them 4.8 out of 12 times, and German bilinguals separated them 6.2 out of 12 times.

Looking at painting and streichen separations, German functional monolinguals separated the stimuli at an average of 8.3 out of 12 times, while English functional monolinguals separated them less often, 4.3 out of 12 times. German–English bilinguals separated them 6.1 out of 12 times, which falls in between the two monolingual groups.

Visual inspection of the graphs reveals that most participants in the English monolingual group (yellow rain clouds) are clustered in the lower end of both drawing and painting and painting and streichen separations. This is because, in addition to the hypothesized arrangements, participants could also choose to not separate either or to separate all stimuli in a triad. The observed clustering of the English monolinguals suggests that they frequently chose to place all stimuli next to each other. Similarly, the green rainclouds show that separation of both drawing and painting stimuli and painting and streichen stimuli was common—though more likely for the latter. This suggests that, at least at times, German monolinguals separated all three stimuli. The raindrops reflect each individual participant's separations out of all 12 trials.

Figure 4

Distribution of Stimuli Separations Across Groups



Tables 1 and 2 show group sizes, means, standard deviations, standard errors, ranges, and confidence intervals.

Table 1

Descriptive Statistics for Drawing/Painting Separations by Group

Group	n	M	SD	SE	Minimum Separations	Maximum Separations	CI
English Monolinguals	24	4.83	3.42	0.70	1	12	[3.46, 6.20]
German Bilinguals	31	6.19	2.97	2.97	0	12	[5.15, 7.24]
German Monolinguals	25	4.60	3.23	3.23	0	10	[3.33, 5.87]

Table 2

Descriptive Statistics for Painting/Streichen Separations by Group

Group	n	M	SD	SE	Minimum Separations	Maximum Separations	CI
English Monolinguals	24	4.25	3.63	0.74	0	11	[2.80, 6.70]
German Bilinguals	31	6.13	3.55	0.64	0	11	[4.88, 7.38]
German Monolinguals	25	8.32	2.61	0.52	1	12	[7.30, 9.34]

A logistic mixed effects model revealed that the differences in drawing and painting separation between English monolinguals and German monolinguals was not statistically significant, $p = .682$; neither was the difference between English monolinguals and German bilinguals, $p = .124$. English and German monolinguals as well as German–English bilinguals had similar odds of separating drawing- and painting stimuli. R-squared was calculated using the r.squaredGLMM function of the lmerTest package. The fixed effects explain about 2% of the variance. The random effects add 43% to the variance explained. See Table 3 for the regression coefficients, standard error, p-values, and r-squared of the mixed effects model.

A post-hoc pairwise comparison (see Table 4) revealed that the difference in odds of separating drawing- and painting stimuli between German bilinguals and German monolinguals is also not statistically significant (p = .116).

Table 3

Mixed-Effects Model Results for Drawing/Painting Separations

Fixed Effects					
Predictors	Coefficient	SE	OR	95% CI in OR	p
Intercept	0.80	0.34	0.58	[0.26, 1.28]	0.16
German Bilinguals	-0.90	0.42	1.88	[0.83, 4.31]	0.12
German Monolinguals	-1.86	0.44	0.84	[0.35, 1.99]	0.68
Random Effects					
Groups	Variance	SD			
Participant	1.77	1.33			
Item	0.73	0.86			

Note. * p < .05, ** p < .01, *** p < .001
marginal R squared: 0.02, conditional R squared: 0.45

Table 4

Post-hoc Pairwise Comparison for Mixed-Effects Model Results of Drawing/Painting Separations

Contrast	Coefficient in OR	SE	p
English Monolinguals – German Bilinguals	0.53	0.41	0.2739
English Monolinguals – German Monolinguals	1.95	0.43	0.9119
German Bilinguals – German Monolinguals	2.25	0.41	0.1157

Note. * p < .05, ** p < .01, *** p < .001

The logistic mixed effects model predicting the odds of painting and streichen separation revealed that the difference between the English monolinguals and German monolingual groups was statistically significant, $p < .001$. The difference between English monolinguals and German bilinguals was also statistically significant with $p = .033$. Table 5 shows the regression coefficients, standard error, p-values, and r-squared of the mixed effects model. The fixed effects explain about 10% of the variance. The random effects add another 34% of the variance.

A post-hoc pairwise comparison (see Table 6) revealed that here, German bilinguals and German monolinguals also differ significantly ($p = .036$).

Table 5

Mixed-Effects Model Results for Painting/Streichen Separations

Fixed Effects						
	Predictors	Coefficient	SE	OR	95% CI in OR	p
	Intercept	-0.86	0.34	0.42	[-0.21, 0.83]	0.0112 *
	German Bilinguals	0.89	0.42	2.45	[1.06, 5.76]	0.0331 *
	German Monolinguals	1.91	0.44	6.79	[2.84, 17.04]	< 0.001 ***
Random Effects						
	Groups	Variance	SD			
	Participant	1.84	1.36			
	Item	0.17	0.41			

Note. * $p < .05$, ** $p < .01$, *** $p < .001$
marginal R squared: 0.1, conditional R squared: 0.44

Table 6

Post-hoc Pairwise Comparison for Mixed-Effects Model Results of Painting/Streichen Separations

Contrast	Coefficient in OR	SE	p
English Monolinguals – German Bilinguals	0.41	0.42	0.0837
English Monolinguals – German Monolinguals	0.15	0.45	<.0001 ***
German Bilinguals – German Monolinguals	0.36	0.41	0.0359 *

Note. * p < .05, ** p < .01, *** p < .001

DISCUSSION

This study shows differences among German and English monolinguals as well as German–English bilinguals in judging the similarity of painting related actions as correlated with common lexical categorization in English and German, which has not been shown before. Following Park & Ziegler (2014), differences in similarity judging among monolinguals is interpreted as a sign of linguistic relativity (RQ1). Following Jarvis (2016), differences in similarity judging comparing German–English bilinguals and the German monolingual group is interpreted as a sign of conceptual transfer (RQ2).

For both research questions of this study, the null hypothesis is rejected partly. English and German monolinguals differed significantly in the number of separations between painting and streichen stimuli, which suggests that English and German monolinguals regarded their similarities differently. German monolinguals separated painting and streichen stimuli significantly more frequently than English monolinguals, which correlates with the more frequent separation of the two in German due to German speakers habitually referring to them with distinct words (“malen” and “streichen”).

Though, the range of responses varied greatly among both groups. Park and Ziegler (2014) interpret their similar findings as support for Bassetti and Cook’s (2011) and Boroditsky’s

(2001) interpretations that speakers may be guided by their native categorization habits but not restricted to them.

However, experimental behavior was not found to correlate with the distinction between "drawing" and "painting" in English (or the non-distinction between them in German). English and German monolinguals did not differ significantly in separating drawing and painting stimuli. The null hypothesis must be accepted in part for the first research question. This does not necessarily suggest that linguistic relativity does not extend to the drawing and painting distinction. In fact, it may show how diverse action-portraying stimuli may be perceived: pictures of drawing and painting can both be seen as doing art, while painting a wall can be perceived as doing work.

According to English monolingual participants' informal reports after the task, this is what prompted them to arrange "drawing" and "painting" next to each other and separating "streichen" from the two—opposite of what was hypothesized. With similar frequency, English monolinguals also decided to arrange all three stimuli without any separations. According to their unrecorded feedback after the task, they wanted to show that drawing is related to painting (as both are art) and that painting (a wall) is also related to painting (as they both use paint).

Similarly, German participants sometimes decided to split apart all three stimuli. Even though they still separated painting and streichen stimuli significantly more often than drawing and painting stimuli (as hypothesized), they reported informally that they often saw differences among all stimuli and that the drawing stimulus could be perceived as someone doing technical drawing, commonly referred to as zeichnen, thus different from "malen", the word used for both drawing and painting. Both could be an explanation for the German monolinguals sometimes not arranging drawing and painting stimuli next to each other like hypothesized. This finding is especially important for anyone wanting to research linguistic relativity and conceptual transfer on the basis of verb partial (in-)equivalence and as a reminder to take caution when mapping lexicogrammatical structures onto behaviors.

In spite of these findings, the statistically significant difference in painting and streichen stimuli separations in the studied sample is in support of the large body of literature about object categorization (e.g., Malt et al., 2004) and placement and direction encoding in verbs (van

Bergen & Flecken, 2017; Park & Ziegler, 2014). This finding thus provides further empirical support to Whorf's (1959) idea of linguistic relativity.

The analysis of differences in painting and streichen separation in the sampled German bilinguals showed significantly different behaviors from both the German and English monolinguals. German–English bilinguals separated painting and streichen stimuli more often than English monolinguals (correlating with the more frequent separation of the two by German speakers). This suggests that what has been referred to as conceptual transfer is occurring: bilinguals show signs of "cross-linguistic influence in the expression and interpretation of conceptual meaning" as Jarvis (2016, p. 608) describes. The German bilinguals that took part in this all had at least 10 years of experience learning English, a total self-reported average proficiency of 7.3 out of 10, and an average of 32.6 hours of exposure to English media per week. Consequently, the results suggest that even advanced learners of English are prone to categorize according to their native language's habits. This finding has significant implications for language teaching as conceptual meaning may not typically be addressed in teaching contexts.

However, the sampled bilinguals still separated painting and streichen stimuli less often than German monolinguals (correlating with the less frequent separation of the two by English speakers). This suggests that the German–English monolinguals also showed signs of what has been called cognitive restructuring (Wang & Wei, 2021; Park & Ziegler, 2014), perceiving differences more similarly to the target speech community. The stage at which this happens in a learner's journey is still unclear in any domain of conceptual transfer. Due to the sample size of the bilingual group, this study cannot make any claims toward this. Park and Ziegler (2014) also point out that a result like this may be interpreted as evidence that bilinguals do not maintain two separate categorization ways that can be accessed in each language environment but that categorization preferences can be influenced by habits from both speech communities at the same time.

CONCLUSIONS

In this study, I investigated the extension of linguistic relativity and conceptual transfer (Whorf, 1956; Bylund & Athanasopoulos, 2014; Park and Ziegler, 2014) to verb categorization. Recent psycholinguistic evidence suggests a close relationship between lexicogrammatical features and categorization behavior in certain aspects of verb perception like position ("put into standing position" vs. "put into lying position" in Dutch compared to English) and direction ("put in" vs. "put on" in Korean and English), which has been attributed to the concepts of linguistic relativity and conceptual transfer. However, this had not been shown for the lexical differences among drawing and painting. In this study, I investigated differences in categorization of different drawing- and painting actions as related to their lexical (in-)equivalence in German and English. In contrast to previous studies, I gave participants the option to arrange the stimuli on a scale that allowed for more than just an odd-one-out arrangement, not restricting the participants to one of two outcomes predefined by the researcher.

The results revealed significant differences in separations between two types of painting ("painting" and "streichen") stimuli among all three groups, English monolinguals, German monolinguals, and German bilinguals, with the bilinguals showing behaviors similar to both groups, thus falling in between the two other groups. This finding supports the research hypotheses. However, the study also revealed that separations between drawing and painting stimuli were not significantly different between groups. This could be due to a wide range of possible interpretations of the stimuli showing drawing- and painting actions.

Since this is the first study researching drawing and painting actions, it can serve as guidance for further research. More importantly, it adds to the growing perspective that linguistic relativity phenomena do not map onto lexicogrammatical cross-linguistic differences in as straightforward a way as sometimes assumed. In addition to providing evidence of linguistic relativity in painting related actions among German and English monolinguals, this paper contributes to the nascent fields of conceptual transfer (Jarvis, 2016) and cognitive restructuring (Wang & Wei, 2021; Park & Ziegler, 2014) showing evidence for both in German–English bilinguals.

REFERENCES

Athanasopoulos, P. (2007). Interaction between grammatical categories and cognition in bilinguals: The role of proficiency, cultural immersion, and language of instruction. *Language and Cognitive Processes*, *22*, 689–699.

Bassetti, B., & Cook, V. J. (2011). Relating language and cognition: The second language user. In V. J. Cook & B. Bassetti (eds.), *Language and bilingual cognition*, pp. 143–190. Psychology Press.

Boroditsky, L. (2001). Does language shape thought? English and Mandarin speakers' conceptions of time. *Cognitive Psychology*, *43*, 1–22.

Bylund, E., & Athanasopoulos, P. (2014). Language and thought in a multilingual context: The case of isiXhosa. *Bilingualism: Language and Cognition*, *17*(2), 431–441.

Cook, V., Bassetti, B., Kasai, C., Sasaki, M., & Takahashi, J. A. (2006). Do bilinguals have different concepts? The case of shape and material in Japanese L2 users of English. *International Journal of Bilingualism*, *10*(2), 137–152.

Gullberg, M. (2009). Reconstructing verb meaning in a second language: How English speakers of L2 Dutch talk and gesture about placement. *Annual review of cognitive linguistics*, *7*(1), 221–244.

Jarvis, S. (2016). Clarifying the scope of conceptual transfer. *Language Learning*, *66*(3), 608–635.

Koster, D., & Cadierno, T. (2018). Is perception of placement universal? A mixed methods perspective on linguistic relativity. *Lingua*, *207*, 23–37.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325–343.

Lupyan, G. (2016). The centrality of language in human cognition. *Language Learning*, *66*(3), 516–553.

Malt, B. C., Sloman, S. A., & Gennari, S. P. (2003). Universality and language specificity in object naming. *Journal of Memory and Language*, *49*(1), 20–42.

Nicoladis, E., & Gao, H. H. (2021). How bilinguals refer to Mandarin throwing actions in English. International *Journal of Bilingualism*. *Special Issue*, 1–18.

Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, 'n'roll: The representation of motion in language and cognition. *Cognition*, *84*(2), 189–219.

Park, H. I., & Ziegler, N. (2014). Cognitive shift in the bilingual mind: Spatial concepts in Korean-English bilinguals. *Bilingualism: Language and Cognition*, *17*(2), 410–430.

Pavlenko, A. (2009). Conceptual representation in the bilingual lexicon and second language vocabulary learning. In A. Pavlenko (Ed.), *The Bilingual Mental Lexicon: Interdisciplinary Approaches* (pp. 125–160). Multilingual Matters.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking". In J. Gumpertz, & G. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge University Press.

Stepanova, O., & Coley, J. (2002). The green eyed monster: Linguistic influences on concepts of envy and jealousy in *Russian and English. Journal of Cognition and Culture*, *2*(4), 235-262.

Van Bergen, G., & Flecken, M. (2017). Putting things in new places: Linguistic experience modulates the predictive power of placement verb semantics. *Journal of Memory and Language*, *92*, 26-42.

Wang, Y., & Wei, L. (2021). Cognitive restructuring in the multilingual mind: language-specific effects on processing efficiency of caused motion events in Cantonese English–Japanese speakers. *Bilingualism: Language and Cognition*, *24*(4), 730–745.

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings* (edited by J. B. Carroll). MIT Press.

LIST OF TITLES OF SECOND LANGUAGE STUDIES GRADUATE WORK AT THE UNIVERSITY OF HAWAIʻI AT MĀNOA, 2023–2024

MILANG SHIN

Department of Second Language Studies, University of Hawaiʻi at Mānoa

In previous editions of *Second Language Studies,* we provided annotated summaries of each available graduate work. However, because many students choose to delay their work appearing publicly, we have opted to list the titles of all work instead and to include the relevant permanent link to the available works on ScholarSpace (https://scholarspace.manoa.hawaii.edu). We hope this new format makes it easier for readers to access the works that are available at the time of publishing.

Graduation Term	Student Name	Degree	Title of Scholarly Paper (AGC, MA) or Dissertation (PhD)
Fall 2023	Alzahrani, Raed	PhD	*Investigating the effect of motivational design on Saudi university students' motivation and L2 writing performance: An experimental mixed methods design using Keller's ARCS model* https://hdl.handle.net/10125/107904
Fall 2023	Choe, Ann	PhD	*A comparative analysis of L2 service-learners' interactional practices in community-based activities and the classroom*

			https://hdl.handle.net/10125/107901
Fall 2023	Le, Hoa	PhD	*Developing technology-mediated TBLT for beginning Vietnamese: Insights from action research* https://hdl.handle.net/10125/107878
Fall 2023	Yagi, Junichi	PhD	*Co-operative instruction in music and sports: Language(s), body, and objects* https://hdl.handle.net/10125/107864
Spring 2024	Ngyuen, Ha	PhD	*At an epistemological crossroads: International students' sense of belonging at a university striving to be a Hawaiian place of learning* https://hdl.handle.net/10125/108421
Spring 2024	Nishizawa, Hitoshi	PhD	*Perceptual adaptation to foreign accents by second language learners* https://hdl.handle.net/10125/108479
Summer 2024	Matsutani, Yuka	PhD	*Institutional policies and tutoring practices at a university writing center*
Summer 2024	Zenker, Fred	PhD	*The processing and acceptability of gapped vs. resumptive relative clauses in first and second language English*
Fall 2023	Chau, Khiet	MA	*A systematic review of AR-assisted and VR-assisted language learning*

Fall 2023	Joo, Ye won (Maria)	MA	*A Korean EFL teacher's agency and identity construction process: A discourse analysis approach* https://hdl.handle.net/10125/108504
Fall 2023	Kunogi, Mayuho	MA	*Reading graphic novels in English as a foreign language for social-emotional learning: A case study with Japanese EFL learners*
Fall 2023	Mase, Chihiro	MA	*Social factors in heritage language maintenance in Hawai'i* https://hdl.handle.net/10125/108508
Fall 2023	Matera, Lilou	MA	*An exploratory case study examining the reported affordances of pop culture media for second language learning*
Fall 2023	Nakamura, Yukie	MA	*Code-switching, repair, and correction in everyday child-parent interaction*
Spring 2024	Au, Kristine	MA	*Taking foreign language beyond the classroom and into everyday life: Non-heritage language learners of Korean in the foreign language context* https://hdl.handle.net/10125/108503
Spring 2024	Chun, Young	MA	*The effectiveness of textual enhancement in captioned foreign language media on grammar acquisition* https://hdl.handle.net/10125/108042

Spring 2024	Liu, Zhengrui	MA	*The impact of language proficiency on job satisfaction and motivation among hotel concierge staff*
Spring 2024	McCarrey, Anne	MA	*Multilingualism for equitable education: A critical analysis of Hawai'i's education policy*
Spring 2024	Owada, Aya	MA	*A comparative study of ChatGPT-supported and human-authored texts: Japanese high school students' creative writing* https://hdl.handle.net/10125/108501
Spring 2024	Villa, Lentina	MA	*Exploring the correlation between age at arrival, length of residence, and speaking proficiency: An analysis of immigrant language acquisition in the Netherlands*
Spring 2024	Wilkinson, Anastasia	MA	*Language maintenance and identity of 3rd generation Russian heritage language speakers in the US* https://hdl.handle.net/10125/108502
Summer 2024	Arakawa, May	MA	*L1 English, L2 Japanese users' L3 Korean writing processes*
Summer 2024	Lee, Sohyeon	MA	*Linguistic tools in L2 writing assessment: A Matthew effect among South Korean high school EFL students*
Fall 2023	Murata, Minori	AGC	*Fostering FL learner agency in COIL classes: Navigating power dynamics through diverse interactional patterns*