Assessing English language proficiency in Costa Rican high schools

Jose Fabián Elizondo-González, Ana Carolina González-Ramírez & Walter Araya-Garita
Universidad de Costa Rica, Costa Rica

This study examines the English language proficiency of Costa Rican high school students using data from the 2022 administration of a Costa Rican English language proficiency test (PELEx PDL-MEP), involving 66,842 participants across public and private schools. Statistical analyses included Welch's t-test to compare overall performance between public and private school students, Welch's ANOVA with post hoc Games-Howell tests to analyze differences across fourteen school types, and multilevel modeling analysis to estimate clustering effects based on educational regions. Results showed that private school students significantly outperformed public school students, with a mean score difference of 13.56 points (Cohen's d=1.78), placing private students predominantly in the B2 CEFR band and public students in the A2 band. Among public school categories, bilingual and specialized programs achieved the highest scores, approaching those of private schools, while adult education-focused options demonstrated the largest performance gaps. These findings highlight significant inequities in English proficiency across school types, emphasizing the need for targeted educational interventions and equitable resource distribution.

Keywords: bilingual education, CEFR, Costa Rica, language testing, public and private schools

Email address for correspondence: ana.gonzalezramirez@ucr.ac.cr

© The Author(s) 2026. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits the user to copy, distribute, and transmit the work provided that the original authors and source are credited. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Introduction

Costa Rica's education system and bilingual policy

The structure of Costa Rica's education system comprises pre-school education, elementary or primary education (grades 1 to 6), secondary education or high school (grades 7 to 11 or 12), and tertiary education. Elementary and secondary education services provided by both private and public institutions include instruction in a second or foreign language in adherence with the guidelines and study programs mandated by Costa Rica's Ministry of Education (Ministerio de Educación of Costa Rica or MEP, for its acronym in Spanish), which have evolved as modern times require more proficient speakers of foreign languages, especially English.

Over the past three decades, Costa Rica has pursued an ambitious agenda to expand bilingual education and strengthen English instruction at all school levels. Early steps included the introduction of foreign language programs in primary schools in 1994, followed by the establishment of Experimental Bilingual High Schools in 1995, which were designed to deliver high-quality academic education with added value in English as a second language (Dirección Regional de Educación Alajuela, 2004). The state's initiatives continued with programs such as the Plan Nacional de Inglés and Costa Rica Multilingüe, both launched in 2008, and the Alianza para el Bilingüismo (Alliance for Bilingualism or ABi, for its acronym in Spanish) created in 2018 (Consejo Superior de Educación et al., 2019; Decreto Ejecutivo, 2019). Most recently, the Política Educativa de Promoción de Idiomas (Educational Policy for Language Promotion) was adopted in 2021 to further improve and consolidate students' communicative skills in a second language (Rojas Díaz, 2021).

The ABi in particular embodies the country's most ambitious goals. Aligned with the Common European Framework of Reference for Languages (CEFR), it seeks to coordinate efforts across public and private institutions, reduce regional and institutional disparities, and raise the overall quality of English instruction. Its central objective is for Costa Rica to achieve a fully bilingual population by 2040, with high school graduates attaining independent command of English (Rojas Díaz, 2021). The envisioned benefits extend beyond education, encompassing improved employability, enhanced social mobility, and greater appeal to international investment (Alianza par

el Bilingüismo, 2018; Asociación Coalición Costarricense de Iniciativas de Desarrollo [CINDE], 2018; Vargas González et al., 2024).

Despite the aforementioned ambitious national policies, the reality of English instruction differs sharply across school types and regions. A full description of the multiple types of secondary schools in the country is given in Appendix I. International and private schools generally provide more intensive English programs, offering additional lessons, immersive environments, and access to native or highly proficient teachers. Public schools, by contrast, often struggle with shortages of qualified staff, inconsistent teacher proficiency —even though MEP requires all instructors to demonstrate C1 proficiency (Dirección General de Servicio Civil, 2025)— and fewer instructional hours. These disparities are particularly visible between well-resourced institutions in the Central Valley—the country's economic hub—and rural or peripheral schools, which frequently lack adequate materials, functioning facilities, computer labs, or reliable internet access (Castro Arias, 2021; Programa Estado de la Nación, 2021; Quesada Pacheco et al., 2023).

Broader systemic challenges further undermine equity. Teachers face heavy workloads, weak administrative support, and declining mental health (Chaves Pérez, 2021). In rural areas, single-teacher schools, common but rarely staffed for English, offer the subject in only about one quarter of cases (Castro Arias, 2021, p. 60). Technological gaps are also stark: many schools lack up-to-date digital tools, only 25% of public institutions have stable internet access, and few teachers have formal training in ICT or online pedagogy (Mojica Barquero, 2021; Programa Estado de la Nación, 2021, 2023). These conditions create learning environments that limit students' opportunities to reach national proficiency benchmarks.

The consequences are evident in higher education. Many students enter university with poor English skills, prompting publicly-funded universities, such as the University of Costa Rica (UCR) and the Instituto Tecnológico de Costa Rica (ITCR), to incorporate English language requirements into their programs to enhance students' competitiveness in the global job market. Expanding bilingual majors has also been recommended as a way to address the shortage of qualified bilingual teachers and to advance the ABi's long-term vision of a fully bilingual population by 2040 (Vargas González et al., 2024).

Language testing initiatives

Building on these policy initiatives, MEP also committed to evaluating the English proficiency of graduating high school students as part of the ABi's agreements. Initially, this exit requirement posed a challenge, since the international exams available in Costa Rica were costly and limited in supply, making them inaccessible to many students and financially unsustainable for the state. To address this gap, the Prueba de Dominio Lingüístico (Language Proficiency Test or PDL-MEP, for its acronym in Spanish), a specially designed national proficiency test, was introduced to assess and certify students' English skills. Developed to provide a more affordable and scalable alternative, the PDL-MEP was administered in 2019, 2021, and 2022 (Araya, 2021; Universidad de Costa Rica, 2021).

The PDL-MEP was developed by the Programa de Evaluación en Lengua Extranjera (Foreign Language Assessment Program or PELEx, for its acronym in Spanish) at the University of Costa Rica's School of Modern Languages, in collaboration with the Dirección de Gestión y Evaluación de la Calidad (Directorate of Quality Management and Evaluation or DGEC) at MEP. Administered as a mandatory graduation requirement for both public and private schools, the test marked a departure from previous models by being delivered digitally on a national scale. More importantly, it assessed students' proficiency through CEFR bands, representing the first use of this international framework in the design of a national language test in Costa Rica (Araya Garita et al., 2022). The test evaluated receptive skills—reading and listening comprehension—and aimed to provide not only certification but also system-level feedback on the effectiveness of English instruction. Although MEP did not schedule its administration in 2023, the University of Costa Rica donated 25,000 tests that were subsequently used in high schools across the country.

The PDL-MEP represented a significant departure from its predecessor, the Exámenes de Bachillerato, which MEP had administered from 1988 until their elimination in 2019. This exam was built on a linear, content-based curriculum and focused largely on reading comprehension and grammar-structure identification through multiple-choice questions (Bolaños Guerrero, 2011). Its minimum passing score of 70 out of 100 served as a graduation requirement for more than two decades (Observatorio Latinoamericano de Políticas Educativas, 2019). However, the exam's washback was

narrow and reductive: instruction often emphasized vocabulary memorization and isolated grammar drills at the expense of integrated communicative skills. As Altamirano Alvarado and Navarro Godínez (2016) observed, even when teachers' trimestral plans listed objectives across the four skills, in practice only reading was consistently taught, disconnected from listening, speaking, and writing. Acknowledging these limitations, MEP formally discontinued the Bachillerato exams in 2019.

Despite over two decades of national testing, English proficiency outcomes remained below expectations, since only about 10% of Costa Ricans reported speaking a second language, with even lower rates in rural areas (INEC, 2017; Education First English Proficiency Index, 2023). Within the school system, earlier assessments revealed that just 11% of high school graduates reached an intermediate–advanced level of English, while 65% remained at a basic level (Quesada Pacheco et al., 2023). More recent surveys showed modest progress, with 29% of students attaining at least B1 in 2019 and 36% in 2021 (Ministerio de Planificación Nacional y Política Económica, 2019). Still, the majority of graduates fell short of MEP's minimum B1 benchmark, underscoring the persistent gap between policy goals and actual student outcomes and highlighting the need for renewed initiatives such as the PDL-MEP.

Purpose of the study

This study examines the outcomes of the 2022 administration of the PDL-MEP. The analysis focuses on graduating high school students from both public and private institutions and evaluates how recent policy initiatives—particularly ABi and Costa Rica Multilingüe—are reflected in measurable proficiency outcomes. By comparing results across school types, the study seeks to determine the extent of progress toward national bilingualism goals and to assess the role of standardized testing as a mechanism for certifying individual achievement and monitoring system-level performance.

To address these aims, this exploratory study attempted to answer the following research questions:

RQ1: Are the overall language proficiency scores obtained by students from private high schools higher than the scores obtained by the students from public high schools?

RQ2: Are there significant differences in overall language proficiency scores among students in the private high school category (private and semi-private) and students in different public high school categories (e.g., academic, bilingual, virtual)[2]?

Literature review

The PDL-MEP as a national assessment

The development of national English assessments in Costa Rica must be understood within the broader policy landscape that prioritizes bilingualism as a national goal. Initiatives such as the Alianza para el Bilingüismo (ABi), the Política Educativa de Promoción de Idiomas (2019), and the adoption of CEFR-aligned curricula have positioned English proficiency as both an educational and economic priority (Alianza para el Bilingüismo, 2018; Rojas Díaz, 2021). Within this context, the Prueba de Dominio Lingüístico (PDL-MEP) was introduced not only to certify students' communicative competence but also to monitor the education system's effectiveness and hold institutions accountable for their outcomes. As Costa Rica's first CEFR-aligned national English test, the PDL-MEP aimed to integrate international standards with national curricular objectives while adapting to local delivery conditions (Araya Garita et al., 2022).

National assessments like the PDL-MEP must be considered within theoretical perspectives that recognize testing as both a technical and a social practice. McNamara (2000) highlights that test use inevitably carries ethical and political consequences, especially when scores inform judgments of educational quality and equity, making fairness, validity, and transparency central concerns (AERA, APA, & NCME, 2014; ILTA, 2024). Bachman (2009) emphasizes the complexity of second language testing, which requires expertise across linguistics, pedagogy, and psychometrics; Bachman and Palmer (1996) underline the obligation of developers to provide validity evidence for diverse populations in high-stakes contexts. More recent approaches, such as

[2] To see the breakdown of high school types across public and private institutions, see Appendix I.

O'Sullivan's (2021) Comprehensive Learning System (CLS), frame assessment as an integrated process that links curriculum, teaching, and assessment, rather than treating tests as isolated instruments. This perspective resonates with Dimova et al.'s (2020) call for locally grounded assessments and with Brown and Abeywickrama (2019) and Coombe et al. (2007), who stress the need for alignment with classroom realities. Viewed in this light, the PDL-MEP represents Costa Rica's attempt to situate national testing within international standards while adapting to local conditions through CEFR-linked reporting, curricular alignment, and adjustments to technological disparities (Araya Garita et al., 2022; Quesada Pacheco & Araya Garita, in press).

Building on these theoretical foundations, research on the PDL-MEP has examined its construct validity from multiple perspectives, including content, consequential, and logistical dimensions. Regarding content validity, Araya Garita et al. (2024) conducted an alignment study using the Angoff method to assess the correspondence between test items and CEFR descriptors. They found stronger agreement for the B1 level than for A1 and A2, where poor consistency among expert judges revealed gaps in measuring intended constructs. These findings point to the need for ongoing item review, calibration of subject-matter experts, and targeted development to strengthen the test's coverage of lower proficiency levels. Similar concerns about alignment were raised by Harsch and Hartig (2015), who showed that judges often apply varied criteria when interpreting CEFR levels, undermining the reliability of alignment procedures and raising questions about score interpretation. Content validity has also been addressed in another study by Araya Garita et al. (2022), which situated the PDL-MEP within a CEFR-aligned framework for assessing receptive skills. The authors emphasized the role of stakeholder input (teachers, policymakers, and test users) in shaping the validation process, consistent with Dimova et al.'s (2020) call for local relevance. They also highlighted the inclusion of broad communicative themes from Costa Rica's national curriculum, thereby enhancing the ecological validity of the test.

Evidence based on test consequences has been investigated in a University of Costa Rica (2021) report, which combined questionnaires, interviews, and psychometric data from 52 test takers. While the PDL-MEP was generally perceived as fair and useful, concerns emerged around differential fairness across test-taker profiles,

unintended negative effects such as test anxiety, and perceptions of misalignment with productive skills. These findings showed the importance of monitoring both intended and unintended consequences of test use.

Finally, Quesada Pacheco and Araya Garita (in press) examined the logistical framework developed by PELEx to ensure accessible and reliable administration. They described innovations such as online, hybrid, and offline delivery formats to address technological inequalities, digital monitoring systems, participant registries, support channels, and real-time assistance. Training programs for test administrators, field visits, and technical reviews were also implemented to guarantee stability and usability. By aligning with CEFR descriptors, maintaining consistency in item development, and safeguarding test-taker data, PELEx has attempted to combine technical robustness with stakeholder engagement and ethical oversight.

The administrations of the PDL-MEP between 2019 and 2022 also provided empirical evidence of Costa Rican students' performance across school types and modalities. The PDL-MEP was administered in 2019, 2021, and 2022. In 2019, according to Araya Garita (2021), 55,143 seniors took the test; 3,687 of them were enrolled in privately funded institutions. On the one hand, the results show that 54% of students from private schools on average obtained a B1 band based on the CEFR while 28% were placed in B2; only 3 students were A1 and 17% A2; only 24 students (1%) were C1. On the other hand, students from the public schools were mostly categorized as A2 (77%) and a minority of 19% in B1; only 2% in B2, and less than 2% in A1 and C1 (525 and 6 students, respectively).

The next PDL-MEP administration occurred in 2021, as the 2020 implementation was canceled due to the COVID-19 pandemic, which disrupted classroom instruction and non-essential educational activities. According to Quesada Pacheco et al. (2023), the 2021 PDL-MEP results revealed that most public high school students in Costa Rica scored at low proficiency levels (A1 and A2) in reading and listening comprehension, failing to meet MEP's B2-exit-profile expectations. Among public high school modalities, a detailed analysis by PELEx (2021) showed that 75% of students in academic high schools, 87% in night high schools, and 48.6% in technical high schools achieved A1 or A2 CEFR levels, which revealed a significant gap in meeting MEP's standards. In contrast but still within the public sector, 89% of students in

experimental bilingual high schools scored at B1 or above, a trend also observed in other public high school categories such as Spanish-English bilingual classrooms, scientific high schools, and humanistic high schools. These results contrast with those obtained in private high schools, where 10% of students scored at the A2 level, 23% at B1, 49% at B2, and 18% at C1, meaning that 90% achieved scores at B1 or above.

English proficiency outcomes across school models in Latin America

According to research carried out in other Latin American countries, high school students' English proficiency varies markedly by institutional model, particularly between public and private high schools. A comparative analysis examining standardized testing measures in Dominican Republic found that institutional type significantly influences English language outcomes (Tavarez Da Costa & Reyes Aria, 2021), with students attending private institutions consistently outperforming their public school peers. In Argentina, a privately funded national study reported a strong correlation between English attainment and private schooling, suggesting that students in private schools receive more comprehensive instruction and perceive English as a more integral part of their education (British Council, 2015). Similar findings emerged in Colombia, where Gómez Molina and Pérez Sua (2022) analyzed a decade of Saber 11 test data and found that private school students, especially in rural areas, consistently achieved higher English scores. The aforementioned authors attribute this superior performance to better instructional resources and a school culture more conducive to language learning—an advantage that public institutions often lack.

The distinction in English language outcomes is also evident within the public and private sectors themselves, underscoring differences among institutional subtypes. In Mexico, Eccius-Wellmann and Santana (2020) found that most students from private high schools scored at the B2 level on the CEFR scale, while those from public high schools were typically placed at B1. In Chile, Walczak et al. (2017) used national Simce Inglés test data to show that 63% of private school students reached at least B1 proficiency, compared to public and voucher-subsidized school students, who were more often categorized at A2/B1. These studies not only confirm significant differences between public and private sectors but also highlight the variation found within each sector. Factors such as socioeconomic status, geographic location, and

access to pedagogical resources may help explain why certain public school alternatives (e.g., bilingual or urban academic schools) may achieve better results than others. This emerging narrative positions institutional type as a key determinant of English language attainment across the region and informs ongoing policy discussions around equity and quality in language education (Elacqua et al., 2018). Taken together, these studies support the assertion that private school students generally attain higher English proficiency levels than public school students. They also suggest that distinctions within the public and private sectors—such as rural vs. urban schools and fully private vs. semi-private or voucher-based institutions—play a role in the variance of student English proficiency outcomes. This literature provides a foundation for addressing both of the present study's research questions regarding the impact of school type on English language proficiency.

Methods

Participants

This study used a secondary dataset of de-identified data that is publicly available through PELEx. Consequently, the anonymity of the participants was assured, and the study did not undergo an ethical review by an Institutional Review Board (IRB).

In 2022, 66,886 Costa Rican high school seniors were required to take the PDL-MEP. After a negligible number of missing values (0.05% for listening and 0.06% for reading) were identified using the "mice" program in R, these cases were deleted listwise, as their exclusion would not impact the statistical power or results. The final dataset comprised 66,842 students from 27 educational regions. Participation was mandatory for all public high school students, while private high schools could have also administered alternative standardized tests such as TOEIC, TOEFL, or IELTS to fulfill graduation requirements. This flexibility likely contributed to the underrepresentation of private school students in the dataset, which includes 61,834 public school students and 5,008 private school students. Notably, demographic variables such as gender, ethnicity, socioeconomic status, and age were not collected.

Sampling

The sampling technique attempted to collect the data from every single high school senior eligible for graduation within the Costa Rican territory in 2022. To achieve this, PELEx in conjunction with MEP administered tests in computer laboratories to students from high schools, prisons, and even hospitals (if students were bedridden) during the months of May, August, September, and October of 2022, under the attentive eyes of an assigned invigilator. Students took the PDL-MEP in computer laboratories that met the minimum connectivity and hardware requisites and whose assignment was provided by MEP authorities. Students in remote areas where these conditions were not met took the test in nearby schools or were temporarily provided with laptop computers.

Measures

In regard to the main purpose and construct of the PDL-MEP, Araya Garita et al. (2022) established that the test aims to assess the English proficiency level of high school students in regard to "both regional and global contexts that pertain to the socio-interpersonal, transactional, and academic domains, formally and informally while using as reference the descriptors of CEFR" (pp. 127-128). The test would enable participants to identify their level of reading and listening proficiency; however, the authors emphasize that, as it does not constitute an official certification, it would not be valid for purposes such as university admissions, visa applications, or employment selection.

In terms of the test specifications, this is a criterion-referenced, online, multiple-choice English proficiency test for the reading comprehension and listening comprehension skills. The test items have been developed by applied linguists, using as a foundation the bands from the Common European Framework of Reference for Languages (CEFR) and the guidelines provided in their 2020 Companion Volume (Council of Europe, 2020). As mentioned above, the objective of the test is to measure students' comprehension in their receptive language skills: reading and listening. Each skill test included 50 questions that needed to be answered within 60 minutes. For a more comprehensive description of the test see Araya Garita et al. (2022) and Quesada Pacheco and Araya Garita (in press).

Data analysis

To address the first research question, mean scores between two independent groups (public and private high school students) were compared. The null hypothesis stated that there is no significant difference in language proficiency scores between students in private and public high schools. The alternative hypothesis posited that students in private schools would outperform those in public schools. Although the normality assumption is a key requirement for t-tests, the large sample size (N= 66,842) ensures that the Central Limit Theorem applies, minimizing concerns about deviations from normality (Coladarci & Cobb, 2013). Levene's test revealed a violation of the homogeneity of variances assumption ($F(1,66840) = 52.408$, $p < .001$). Consequently, Welch's t-test, which is robust to unequal variances, was employed (West, 2021).

For the second research question, a one-way ANOVA was conducted to compare scores across fourteen distinct high school types, including two private high school types (private and semi-private) and twelve public high school types (academic day, night academic high schools, virtual education high schools (Colegio Virtual Marco Tulio Salazar), CONED, experimental bilingual high schools, bilingual classrooms, technical high schools, CINDEA, IPEC, scientific high schools, prison inmate education, and humanistic high schools). Although the school types differed in student body composition, selection process, and instructional model, they were all subject to the national English curriculum and were administered the same CEFR-aligned proficiency test. This provided a basis for comparing average proficiency scores across school types under a common assessment framework. The purpose of these comparisons was to examine whether outcome patterns varied across institutional contexts and to support the broader use of results for system-level monitoring and future decisions related to policy, resource allocation, and research on equity in language education.

The following hypothesis was tested: there are no statistically significant differences in overall language proficiency scores among students attending different types of high schools. Although the assumption of normality is a key requirement for ANOVA, the Central Limit Theorem ensures that normality is approximately met for groups with large sample sizes. For smaller groups, Shapiro-Wilk tests revealed some deviations from normality ($p < 0.05$); however, ANOVA is robust to moderate departures from

normality (Coladarci & Cobb, 2013). Levene's test indicated a significant violation of the homogeneity of variances assumption ($F(13,66828) = 258.91$, $p < 0.001$). Consequently, Welch's ANOVA was employed, as it accounts for unequal variances while comparing multiple independent groups.

Although neither school nor student identifiers were available, students were nested within 27 educational regions. To explore the potential impact of this clustering, a multilevel model with region as a random effect was estimated post hoc. The results closely mirrored those of the original ANOVA, supporting the robustness of the findings. A detailed summary of this supplemental analysis is provided in Appendix II.

Out of 66,842 participants, 1,295 respondents (2%) were identified as outliers using the interquartile range criterion. Outliers were retained to ensure the analysis reflects the full range of variability in participants' scores. To assess the robustness of the results, all statistical tests were run twice—once with and once without outliers—and the findings were consistent across both conditions.

Data cleaning and analyses were primarily conducted using R, while IBM SPSS was employed for specific analyses related to the second research question.

Results

RQ1: Are the overall language proficiency scores obtained by students in private high schools higher than the scores obtained by the students in public high schools?

To compare overall performance, Table 1 presents descriptive statistics for test scores, averaging listening and reading comprehension results, for students enrolled in private and public high schools.

Table 1. Comparison of overall test scores obtained by students from private and public high schools

Private high schools									
n	M	95% CI	Mdn	Min	Max	SD	Skew	Kurt	SE
5,008	38.08	[37.86, 38. 31]	40.5	13	50	8.04	-0.9	-0.08	0.11
Public high schools									
n	M	95% CI	Mdn	Min	Max	SD	Skew	Kurt	SE
61,834	24.52	[24.46, 24. 58]	22.5	5	49.5	7.58	0.98	0.4	0.03

Note. *n* = sample size; *M* = mean; 95% CI= 95% confidence interval; *Mdn* = median; *Min* = minimum; *Max* = maximum; *SD* = standard deviation; *Skew* = skewness; *Kurt* = kurtosis; SE = standard error.

To address the first research question, a Welch Two Sample t-test was conducted to compare the overall mean scores between private and public high school students in the current sample (N= 66,842). Results indicated that private high school students (n= 5,008, M= 38.08, SD= 8.04) scored significantly higher on average than public high school students (n= 61,834, M= 24.52, SD= 7.58); t (5753.2) = 115.34, p <.001. Based on Plonsky and Oswald (2014) , the effect size (d = 1.78, 95% CI [1.75, 1.81]) is considered large, providing strong evidence to reject the null hypothesis.

The histograms in Figure 1 and the scatterplots in Figure 2 visually highlight this difference. As shown in Figure 1, private high school students' scores are concentrated toward the higher end of the scale, following a slightly left-skewed distribution. In contrast, public high school students' scores are clustered toward the lower end, forming a more right-skewed distribution. Similarly, the scatter plot in Figure 2 illustrates that private high school students predominantly occupy the first quadrant, achieving higher scores on both listening and reading tasks. Conversely, public high school students are largely grouped in the third quadrant, with lower scores below the 50% threshold of the overall score (24.52 out of 50 points).

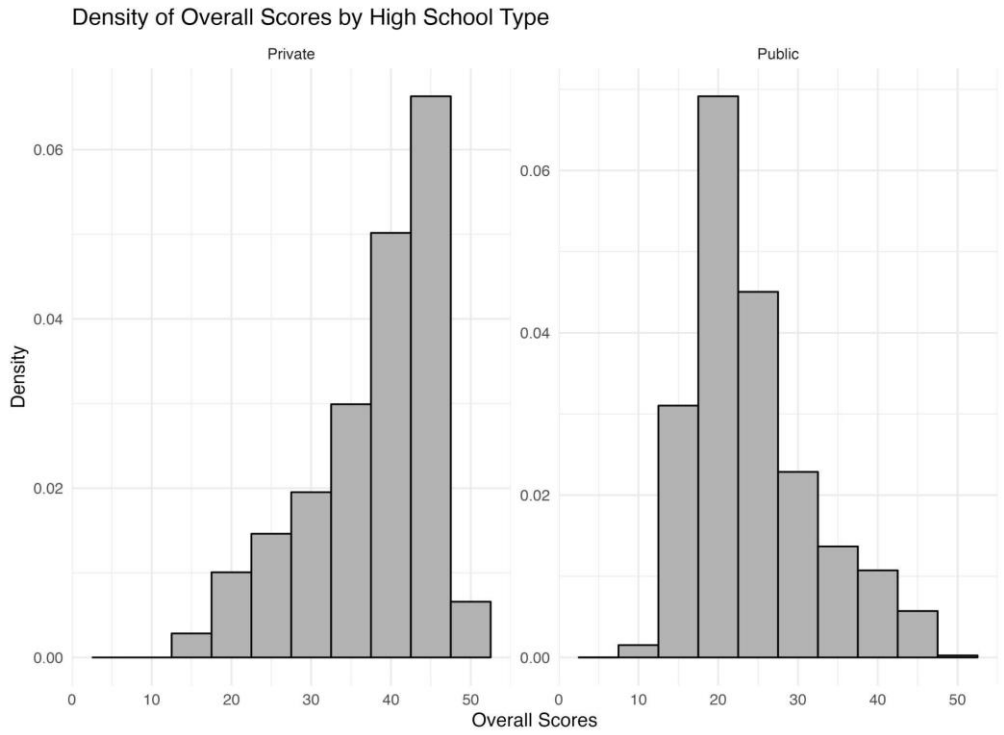


Figure 1. Comparison of overall scores for students enrolled in private and public high schools

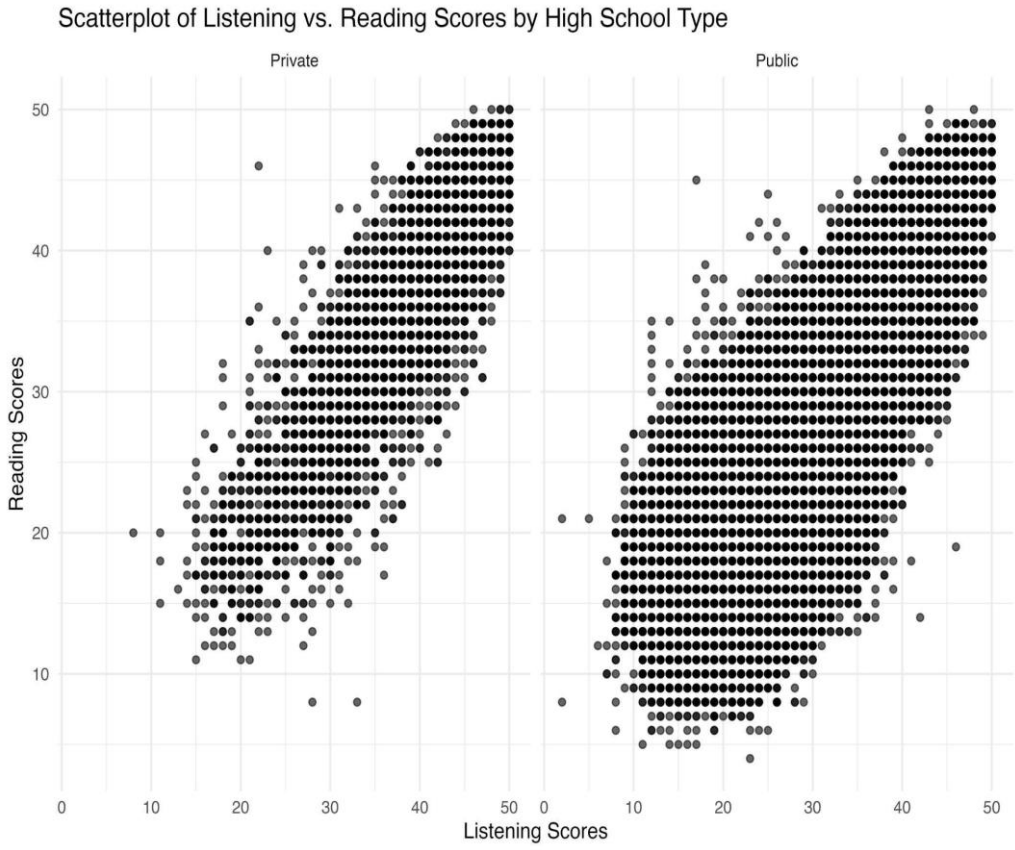


Figure 2. Comparison of reading and listening scores for students enrolled in private and public high schools

These findings underscore a clear performance disparity between private and public high school students, with private school students achieving an average score that is 13.56 points higher than their public school counterparts (38.08 vs. 24.52). This substantial difference, supported by a large effect size (d=1.78), reflects a marked contrast in outcomes and raises questions about the potential underlying factors contributing to this gap.

RQ2: Are there significant differences in overall language proficiency scores among students in the private high school category (private and semi-private) and students in different public high school categories (e.g., academic, bilingual, virtual)?

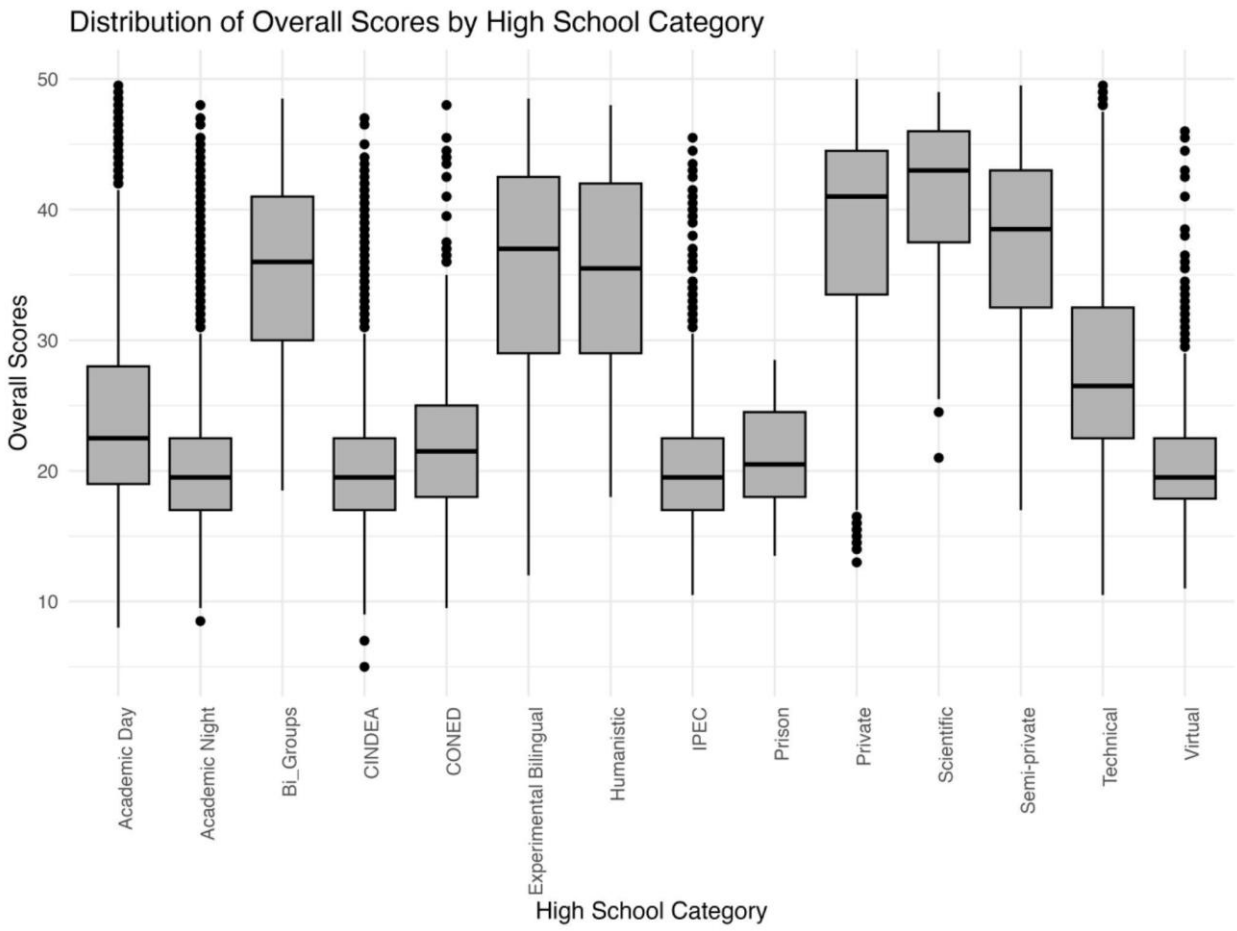
Across the diverse high school types in private and public high schools, a detailed comparison of overall scores—calculated as the average of listening and reading comprehension totals—is summarized in Table 2.

Table 2. Descriptive statistics of different high school categories in descending order based on high school type and mean scores

Type	Category	*n*	*M*	*95% CI*	*SD*	*Mdn*	*Min*	*Max*	*Skew*	*Kurt*	*SE*
Private	Private	3997	38.26	[38.00, 38.52]	8.29	41	13	50	-0.97	-0.02	0.13
Private	Semi-private	1011	37.39	[36.96, 37.82]	6.9	38.5	17	49.5	-0.54	-0.48	0.22
Public	Scientific	197	41.08	[40.23, 41.93]	6.04	43	21	49	-0.97	0.17	0.43
Public	Humanistic	109	35.62	[34.25, 36.97]	7.2	35.5	18	48	-0.14	-1.12	0.69
Public	Bil. Exp.	2157	35.5	[35.15, 35.85]	8.31	37	12	48.5	-0.46	-0.84	0.18
Public	Bil. Class	201	35.23	[34.22, 36.24]	7.26	36	18.5	48.5	-0.37	-0.71	0.51
Public	Technical	14567	27.79	[27.68, 27.91]	7.22	26.5	10.5	49.5	0.59	-0.37	0.06
Public	Ac. Day	27459	24.32	[24.23, 24.40]	7.18	22.5	8	49.5	1.04	0.65	0.04
Public	CONED	660	22.01	[21.59, 22.43]	5.5	21.5	9.5	48	1.14	2.48	0.21
Public	Prison	42	20.86	[19.60, 22.11]	4.03	20.5	13.5	28.5	0.25	-1.03	0.62
Public	Virtual	480	20.75	[20.29, 21.21]	5.12	19.5	11	46	1.82	4.96	0.23
Public	IPEC	1166	20.42	[20.13, 20.71]	5.1	19.5	10.5	45.5	1.51	3.55	0.15
Public	Ac. Night	6433	20.37	[20.25, 20.49]	4.95	19.5	8.5	48	1.53	3.96	0.06
Public	CINDEA	8363	20.08	[19.98, 20.17]	4.57	19.5	5	47	1.15	2.86	0.05

Note. *n* = sample size; *M* = mean; 95% CI= 95% confidence interval; *SD* = standard deviation; *Mdn* = median; *Min* = minimum value; *Max* = maximum value; *Skew* = skewness; *Kurt* = kurtosis; *SE* = standard error of the mean.

To better understand these overall test score distributions, Figure 3 shows box plots representing each high school category's overall score distribution, which would allow for a more readily understandable comparison of test results across all groups. Several groups have smaller variances than others; for example, students in prisons show the least variance, performing very similarly within their group. Other groups seem to have very large within-group variances, such as academic night schools, CINDEA, and IPEC, which are all high school types that accommodate adult student populations.



Note. Outliers were retained to reflect the full range of variability, as discussed, with no significant impact on the findings.

Figure 3. Distribution of high school categories and overall scores

To answer the second research question, an analysis of variance (ANOVA) was conducted to examine differences in overall scores across groups. The Welch-corrected ANOVA results indicated a significant main effect of group membership on overall scores, Welch's F (13, 1304.624) = 2803.550, p < .001, suggesting substantial differences in mean scores between groups. The null hypothesis is rejected.

The effect size, eta-squared (η^2= 0.352, 95% CI [0.347, 0.357]), indicates that 35.2% of the variance in overall scores is explained by group membership. While the interpretation of eta-squared benchmarks is debated (Norouzian & Plonsky, 2018; Richardson, 2011), this value represents a substantial effect size, capturing a significant portion of the explained variance.

Given the violation of the homogeneity of variances assumption, post hoc comparisons using tests that assume equal variances, such as Tukey's HSD or the Bonferroni correction, would not be appropriate. Instead, the Games-Howell test was chosen for post hoc comparisons since it does not rely on the assumption of equal variances and is robust to violations of this assumption, making it a suitable choice in this scenario (Games & Howell, 1976).

Using private high schools as the reference group, all pairwise post hoc comparisons are statistically significant at the .005 level. Table 3 summarizes the pair-wise mean differences in ascending order and their p-values.

Table 3. Overall mean difference comparison between private high schools and other categories

Type	Category	Mean difference	95% CI	p-value
Public	Scientific	-2.82***	[-4.35, -1.30]	< 0.001
Private	Semi-private	0.87*	[0.02, 1.72]	0.04
Public	Humanistic	2.64*	[0.23, 5.04]	0.018
Public	Bilingual Exp.	2.76***	[2.02, 3.51]	< 0.001
Public	Bilingual Class.	3.03***	[1.23, 4.82]	< 0.001
Public	Technical	10.47***	[9.98, 10.95]	< 0.001
Public	Acad. Day	13.94***	[13.47, 14.40]	< 0.001
Public	CONED	16.25***	[15.41, 17.10]	< 0.001
Public	Prison	17.40***	[15.15, 19.65]	< 0.001
Public	Virtual	17.51***	[16.61, 18.41]	< 0.001
Public	IPEC	17.84***	[17.17, 18.51]	< 0.001
Public	Acad. Night	17.89***	[17.40, 18.37]	< 0.001
Public	CINDEA	18.18***	[17.71, 18.66]	< 0.001

Note. *Significant at the 0.05 level, **Significant at the 0.01 level, ***Significant at the 0.001 level.

Table 3 illustrates that students in scientific high schools score the highest, outperforming students in private high schools, as well as all other high school types regardless of their private or public model.

While the mean difference for semi-private high school scores was statistically significant (mean difference= 0.87, p= 0.04), this result warrants caution due to its proximity to the significance threshold, suggesting that private and semi-private high schools may perform indistinguishably in another sample.

The humanistic, bilingual experimental, and bilingual classroom high school categories exhibited significantly lower mean differences from private schools, although these differences were the smallest among all public school types. In contrast, the high school types with the largest mean differences from private high schools were those focused primarily on adult education, including CONED, prison schools, virtual high schools, IPEC, academic night high schools, and CINDEA.

It is important to interpret these results cautiously, particularly for categories with smaller sample sizes, such as prison (n= 42) and humanistic high schools (n= 109).

Discussion

RQ1: Are the overall scores obtained by students in private high schools higher than the scores obtained by students in public high schools?

Although all high school categories receive English instruction as part of the *Educación General Básica* (Core General Education Subjects or EGB, for its acronym in Spanish) curriculum, as mandated by MEP (MEP, 2023), the results demonstrate a substantial and statistically significant gap in overall language proficiency scores between private and public high school students, a trend observed in previous administrations of the test. On average, private school students scored 13.56 points higher than their public school peers (38.08 vs. 24.52), with a large effect size (d= 1.78), placing private high school students on average in the B2 CEFR band and public high school students in the A2 band, using PELEx PDL's scoring scale, aligning with the findings in Walczak et al. (2017), Eccius-Wellmann and Santana (2020), Araya Garita (2021), PELEx (2021), and Quesada Pacheco et al. (2023).

These findings not only confirm the rejection of the null hypothesis but also highlight a considerable disparity in performance between both education models that may evidence differences in resource assignment and availability, curriculum quality and its implementation, or other structural factors also highlighted in the Latin American region, such as the perceived usefulness of the language for the students (Argentina) or the access to an educational culture that is more conducive to learning (Colombia). The observed performance differences may stem from misalignments within these components—such as variations in curriculum quality or access to instructional

support—highlighting the need for systemic reform guided by a cohesive philosophy of learning that stems from stakeholders' discussions on education quality and accessibility (Elacqua et al., 2018).

As highlighted by Araya Garita (2019), substantial efforts by the public sector are still needed to reach the country's ambitious goal of bilingualism and overcome the gaps between high school categories. On the one hand, the current average number of lessons offered in the public education system seems insufficient to advance students to higher proficiency levels. In addition, public high schools have access to limited resources and employ teachers whose working conditions are worsened by a lack of technological training and inconsistent English levels. Prior research has further shown that rural schools are at an even greater disadvantage, offering fewer English lessons, relying on outdated or inadequate facilities, and often lacking stable internet connectivity (Castro Arias, 2021; Programa Estado de la Nación, 2021, 2023; Quesada Pacheco et al., 2023). Although our study did not include an analysis on school location, it is reasonable to speculate that the proficiency gaps observed here may be compounded by such geographic inequalities. Future studies should examine the interaction between school type and geographic region to better understand how the urban-rural divide shapes English learning opportunities. On the other hand, the higher proficiency levels obtained by private schools indicate that the conditions offered yield more positive results: increased allocation of resources, more comprehensive English instruction, better qualified teachers, and more immersive learning environments. The evidence suggests that relegating English to 'just another subject' has not resulted in the expected outcomes, echoing findings from Argentina (British Council, 2015).

The implementation of Comprehensive Learning System (CLS) principles (O'Sullivan, 2021) could help mitigate disparities by fostering a more integrated and equitable educational ecosystem across school models, one that reconciles testing practices with local sensitivity and contextualization (Dimova et al., 2020). Such disparities raise important questions about the broader implications for achieving educational equity and identifying the factors driving performance divides. To ensure fairness and transparency, experts stress that standardized test results should be shared responsibly and ethically, with stakeholders informed before policy decisions that

impact programs and accountability are made (Elacqua et al., 2018; McNamara, 2000). Moreover, these results must be interpreted within the context in which they occur, since ecological validity (Dimova et al., 2020) is essential for understanding their causes, implications, and meaning. The CLS framework highlights these principles by encouraging a shift from isolated interventions toward holistic solutions that address structural inequalities in education (O'Sullivan, 2021).

In line with this integrative vision, foreign language instruction based on CEFR principles has encouraged Costa Rican teachers to design student-centered environments that foster autonomy and active learning (Quesada Pacheco & Araya Garita, in press). Clearer learning goals now guide lesson planning, enabling continuous monitoring of student progress and tailored instruction when necessary. Classroom practices promote active participation and a more realistic use of language, thus reinforcing learner autonomy and engagement, which are central to CEFR's learner-centered philosophy. Assessment practices have also evolved to include meaningful, contextualized tasks that integrate all skills and rely on performance-based rubrics, such as those applied in the PDL-MEP. These approaches produce more authentic measures of student proficiency, enhancing construct validity and offering actionable insights that guide instruction and support learning (Quesada Pacheco et al., 2023).

RQ2: Are there significant differences in overall language proficiency scores among students in the private high school category (private and semi-private) and students in different public high school alternatives (e.g., academic, bilingual, virtual)?

At the 0.05 alpha level, the null hypothesis is rejected. It is worth mentioning that both the scientific and humanistic high schools are two-year programs designed for a handful of pre-selected, top-performing high school students from around the country, who are then selectively admitted after passing an entrance test designed for these specialized high schools (see Appendix I). Whereas scientific schools offer five 45-minute lessons per week, humanistic schools have three English lessons per week. In spite of this difference between them, these two school categories are top performers in the public school category and their students obtain higher scores than those in the rest of the sample, regardless of high school type, as seen in Table 3.

Because of this group's specific features and the small number of students enrolled in these programs, comparing these high schoolers' results with other public or private high school categories is not only challenging but it might result in unreliable and non-generalizable conclusions.

Consequently, the five-year public high school categories that perform the closest to the private and semi-private high school models are the experimental bilingual high schools and bilingual classrooms; each of which offer ten 30-minute English lessons per week, which stands out among the English lessons offered in all other public school categories. A three-point difference separates them from the private and semi-private high school scores (Table 3), and though significant, it is still a small difference that shows that experimental bilingual high schools and bilingual classrooms may be close to bridging the gap between public and private high school foreign language proficiency attainment in Costa Rica; this trend was also observed in Quesada Pacheco, Araya Garita, and Fallas Godínez (2023).

Public high school categories that focus on adult education seem to evidence the largest gaps in foreign language proficiency when compared to private high schools. A possible reason for this could be rooted in how dissimilar these adult populations might be in comparison to other students in private categories, which normally cater to students who are not in need of employment, are 13-18 years of age, and might potentially enjoy their social and cultural capital and resources, leading to more positive outcomes in foreign language acquisition (Gómez Molina & Pérez Sua, 2022). Adding to this already disadvantageous circumstance, English instruction in these adult-oriented schools ranges between two to four lessons per week, which does not seem to provide as much foreign language input as it is necessary to boost the learning of a second language among such a group of students. This adult population could have also met with an assessment system they are not used to taking, for they might have expected an evaluation similar to that of the previous *Bachillerato exam* (a rote learning, reading-focused test), as explained by Bolaños Guerrero (2011) or they may not be as technologically literate as their access to computer labs and IT trained educators is scarce. However, since no demographic data was collected for this study, analyses using covariates or other dependable variables could not be carried out, thus restricting and limiting the interpretability of our results.

Conclusions

This study has evidenced the existing gap in students' English proficiency across school models and categories in Costa Rica, which not only highlights the profound deficiencies in the country's English education system, but also outline the enormous challenges that lie ahead in order to achieve the national bilingualism goals by 2040 (ABi). First, the study showed a statistically significant difference in English performance between students from public and private high schools. Most students from private institutions scored at the B2 level, whereas the majority of public high school students were placed in the A2 level. These results align with those obtained in other Latin American countries, which suggest differences in available resources, curriculum quality, and structural factors as possible reasons. To reduce the inequalities between school models, an increase of the number of English lessons per week might be strongly advisable, as this results in greater exposure and practice opportunities for students. Along with more English lessons, there is an urgent need to provide learners with more appropriate study areas: language-immersive environments that include improved physical facilities, access to updated digital resources, and English teachers who are motivated and trained to teach with such tools.

Different public school categories also showed differences between them. Although the results of students from scientific and humanistic high schools were higher than those of students from private institutions, these results should be taken with a grain of salt. First, they offer different numbers of English lessons per week —three and five, respectively. Secondly and perhaps more importantly, their students undergo strict and selective admission processes, which results in smaller groups of few but exceptional students. The results of two other categories, experimental bilingual high schools and bilingual classrooms, are closer to those from private high schools and lend themselves to be used as examples of good practices. They show the positive effect of more English lessons per week (ten lessons) and have even higher academic demands that their students are expected to meet. The poor results obtained from adult-targeted institutions seem to indicate this population may benefit from more English lessons, a teaching method that is better suited to their specific needs and

capabilities, and training in technological literacy skills, albeit digital resources are available.

The findings of this study highlight the importance of integrating curriculum, instruction, and assessment within a Comprehensive Learning System (O'Sullivan, 2021), to promote equitable and effective language learning. PDL-MEP exemplifies how large-scale assessments, when aligned with CEFR standards and supported by robust stakeholder engagement, can function as diagnostic tools that inform instructional decisions and guide learning processes. The stronger performance of students in Experimental Bilingual and Bilingual Classrooms demonstrates the positive impact of increased English exposure and higher academic expectations within an integrated curriculum. In contrast, the varied outcomes across public high schools underscore the limitations of fragmented systems that lack alignment across educational components. These results affirm the need for continued investment in cohesive educational models that combine curriculum, pedagogy, and assessment to ensure fair and meaningful learning outcomes for all students.

Limitations and recommendations

Some high school categories have relatively small sample sizes, which might put a constraint on data interpretation and generalizability. However, some of these types of high schools are expected to have smaller sample sizes due to their nature: scientific and humanistic high school types, for example, are extremely selective in their admission process, leading to an understandably and consistently small number of participants. Other categories also have small sample sizes but are less consistent. Students taking this language test from prison may vary from year to year, and it is not expected that tens of thousands do so regularly. In such cases when sample size is not large enough for group comparisons, researchers may consider just reporting the descriptive statistics for those groups and using the rest for their hypothesis testing. In this way, these high school categories would still be acknowledged but would not compromise the reliability and validity of the claims made from the statistical tests.

The lack of student demographic information limits the scope of the investigation and restricts the types of analysis that can be done and the kind of inferences that can be made from the results obtained. It is recommended that test administrators collect

more detailed participant demographic information in future test administrations in order for researchers to develop more complex methodologies and provide nuanced interpretations of their findings after accounting for both controlled and extraneous variables.

The only two private high school models reported here are private and semi-private ones. However, there are other alternatives, such as the technical private high school. Future studies could attempt at collecting information from these other private high school types, which may contribute to creating a broader understanding of our Costa Rican educational landscape.

Though MEP implemented a new foreign language curriculum in 2016, in conjunction with nation-wide policies to promote English-Spanish bilingualism (Azofeifa Murillo, 2019), eight years later the gap between private and public foreign language education has not been bridged yet. It would be revealing to replicate this study in 2028, when the curricular changes have completed a full educational cycle (from kindergarten to high school) to help determine the efficacy of said policy and curricular changes.

As experts suggest, results from standardized tests such as PDL-MEP should be shared with all stakeholders before educational policies are set in place and as these are being implemented. Doing so could help to see more expeditious changes in educational programs and enhanced school accountability (Elacqua et al. 2018), which could in turn not only result in greater effectiveness of EFL programs, but also ensure increased transparency and fairness in the learning and assessment processes.

In regard to the educational system itself and, more specifically, instruction of English as a foreign language, some general recommendations can be provided to improve high school students' proficiency: more English lessons should be provided in schools where their number does not reach ten lessons per week; this could be achieved by either creating additional English initiatives or by expanding the reach of bilingual programs to other institutions. English instructors should be not only required to demonstrate their English proficiency and pedagogical knowledge but also given training opportunities and better working conditions, as this would likely improve their teaching practice and students' learning environments. Digital training is a must for faculty and learners, but this should go hand in hand with funding or alliances to

provide schools with the technological tools, such as internet connectivity and computer labs, that would allow for more positive outcomes. Adult learners may benefit not only from more English lessons, but also from having a specialized English assessment that adjusts more closely to their skills and needs.

Author disclosures

The authors are researchers at PELEx (Universidad de Costa Rica), developer and administrator of the PDL-MEP. PELEx's involvement was limited to providing a de-identified dataset; it had no role in the study's design, analysis, or interpretation. No other conflicts are declared.

The study was conducted independently with no financial support.

Author Contributions:

- Jose Fabián Elizondo-González: Conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing-original draft, writing-review & editing.
- Ana Carolina González-Ramírez: Conceptualization, investigation, project administration, resources, supervision, validation, visualization, writing- original draft, writing-review & editing.
- Walter Araya-Garita: Conceptualization, investigation, project administration, resources, supervision, validation, visualization, writing- original draft, writing-review & editing.

All authors read and approved the final manuscript.

ORCID iDs

Jose Fabián Elizondo-González https://orcid.org/0000-0003-4819-0213

Ana Carolina González-Ramírez https://orcid.org/0000-0003-1937-6465

Walter Araya-Garita https://orcid.org/0000-0002-5340-6384

References

Alianza para el Bilingüismo. (2018). *SITEAL*. https://siteal.iiep.unesco.org/bdnp/3109/acuerdo-apertura-alianza-bilinguismo

Altamirano Alvarado, J., & Navarro Godínez, M. (2017). Into the reading methodology of an EFL teacher: A way to improve the students' academic achievement for the National-High-School-Graduation Examination. *Revista De Lenguas Modernas*, (25). https://doi.org/10.15517/rlm.v0i25.27694

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Araya Garita, W. (2021). Dominio lingüístico en inglés en estudiantes de secundaria para el año 2019 en Costa Rica. *Revista de Lenguas Modernas*, (34), 1–21. https://doi.org/10.15517/rlm.v0i34.43364

Araya Garita, W., Elizondo González, J., & González Ramírez, A. (2022). Un acercamiento al constructo de la prueba de dominio lingüístico del idioma inglés desarrollada por la Universidad de Costa Rica para el Ministerio de Educación de Costa Rica. *Estudios de Lingüística Aplicada, 0*(75), 119-143. doi: https://doi.org/10.22201/enallt.01852647p.2022.75.1013

Araya Garita, W., Fallas Godínez, J., & Valdelomar Miranda, N. (2024). Reflecting on the CEFR alignment to the assessment context: Overview of the standard-setting process. [Conference presentation]. Conference on Democracy and Language Education: Teaching and Assessing for Social Justice, Barranquilla, Colombia.

Asociación Coalición Costarricense de Iniciativas de Desarrollo [CINDE]. (2018). *Costa Rica education facts: English a national priority*. https://www.cinde.org/en/essential-news/costa-rica-declares-english-teaching-a-national-priority

Azofeifa Murillo, M. (2019, February 11). *Más de 5.300 estudian inglés gracias a Alianza para el Bilingüismo (ABi)*. Ministerio de Educación Pública. Gobierno

de Costa Rica. https://www.mep.go.cr/noticias/mas-5300-estudian-ingles-gracias-alianza-bilingueismo-abi

Bachman, L. F. (2009). Preface. In L. Cheng & A. Curtis (Eds), *English language assessment and the Chinese learner* (pp. x – xii). Routledge.

Bachman, L.F., & Palmer, A.S. (1996). *Language testing in practice*. Oxford University Press.

Bolaños Guerrero, F. (2011). Análisis multivariado de las calificaciones de bachillerato en Costa Rica en el año 2006. *InterSedes*, *8*(15).

British Council. (2015, May). *English in Argentina: An examination of policy, perceptions and influencing factors*. https://www.teachingenglish.org.uk/sites/teacheng/files/English%20in%20Argentina.pdf

Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson Education.

Castro Arias, K. (2021). Challenges and possible solutions for English teaching and learning in the Province of Guanacaste, Costa Rica. In Zúñiga, J.P. (Ed.), *English language teaching in Costa Rica: Facing challenges* (pp. 28-43). Centro de Investigación y Docencia en Educación (CIDE). http://hdl.handle.net/11056/23947

Chaves Pérez, J.D. (2021). Reconsidering English language learning in Costa Rica: Exploring teachers' well-being and its importance, challenges, and recommendations in the wake of a new era. In Zúñiga, J.P. (Ed.), *English language teaching in Costa Rica: Facing challenges* (pp. 28-43). Centro de Investigación y Docencia en Educación (CIDE). http://hdl.handle.net/11056/23947

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum. https://doi.org/10.4324/9780203771587

Coladarci, T., & Cobb, C. D. (2013). *Fundamentals of statistical reasoning in education* (4th ed.). Wiley.

Consejo Superior de Educación, Ministerio de Educación Pública, & Política Educativa de Promoción de Idiomas. (2019, March 19). *Política Educativa de Promoción de Idiomas: Hacia una Costa Rica bilingüe*.

https://ddc.mep.go.cr/sites/all/files/ddc_mep_go_cr/archivos/politica_educativa_para_la_promocion_de_idiomas.pdf

Coombe, C. A., Folse, K. S., & Hubley, N. J. (2007). *A Practical guide to assessing English language learners*. University of Michigan Press.

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Companion volume, Council of Europe, Strasbourg. https://rm.coe.int/common-european-framework-of-reference-for-languages- learning-teaching/16809ea0d4

Decreto Ejecutivo N° 41858-MTSS-MEP-MICITT-MCSP. (2019). Declaratoria de interés público de la alianza para el bilingüismo (ABi). http://www.pgrweb.go.cr/scij/Busqueda/Normativa/Normas/nrm_texto_completo.aspx?param1=NRTC&nValor1=1&nValor2=89499&nValor3=117519&strTipM=TC#up

Dimova, S., Yan, X., & Ginther, A. (2020). *Local language testing: Design, implementation, and development*. Routledge. https://doi.org/10.4324/9780429492242

Dirección General de Servicio Civil. (2025). *Manual descriptivo de especialidades docentes inglés*. https://www.dgsc.go.cr/ts_clase_docente/Especialidades%20Docentes/Ingles.pdf

Dirección de Gestión y Evaluación de la Calidad. (2019, December 10). *Bachillerato culmina su ciclo*. Ministerio de Educación Pública. https://mep.go.cr/noticias/bachillerato-culmina-su-ciclo

Dirección Regional de Educación Alajuela. (2004, August). *Liceo Experimental Bilingüe de Grecia*. Ministerio de Educación Pública. https://drea.mep.go.cr/circuito-10/liceo-experimental-bilingue-de-grecia

Eccius-Wellmann, C., & Santana, J. C. (2020). Variables that influence language proficiency in students from public and private high schools in Mexico. *MexTesol Journal, 44*(2). https://mextesol.net/journal/index.php?page=journal&id_article=20210

Education First English Proficiency Index. (2023). *EF EPI EF English Proficiency Index A Ranking of 113 Countries and Regions by English Skills*. Official Document online. www.ef.com/epi

Elacqua, G., Iribarren, M. L., & Santos, H. (2018). *Private schooling in Latin America: Trends and public policies* (Technical note No. IDB-TN-01555). Inter-American Development Bank, Education Division, Social Sector. https://publications.iadb.org

Games, P. A., & Howell, J. F. (1976). Pairwise multiple comparison procedures with unequal n's and/or variances: a Monte Carlo study. *Journal of Educational Statistics*, *1*(2), 113-125. https://doi.org/10.3102/10769986001002113

Gómez Molina, A. N., & Pérez Sua, L. M. (2022). *The gap between public and private education according to the score obtained the English area for the ICFES Saber 11 tests over the last 10 years (2010-2020)*. Repository ECCI. https://repositorio.ecci.edu.co/handle/001/3093

Harsch, C., & Hartig, J. (2015). What are we aligning tests to when we report test alignment to the CEFR? *Language Assessment Quarterly, 12*(4), 333–362. https://doi.org/10.1080/15434303.2015.1092545

International Language Testing Association (ILTA). (2024). *ILTA guidelines for practice*. https://www.iltaonline.com/page/ILTAGuidelinesforPractice

McNamara, T. (2000). *Language testing*. Oxford University Press.

Ministerio de Educación Pública (MEP). (2023, August). Compendio de ofertas y servicios del Sistema Educativo Costarricense. https://www.mep.go.cr/sites/default/files/2024-09/A17-CompendioOfertasServiciosMEP.pdf

Ministerio de Planificación Nacional y Política Económica (MPNP). (2019). Plan Nacional de Desarrollo y de Inversión Pública. Gobierno de Costa Rica. https://observatorioplanificacion.cepal.org/sites/default/files/plan/files/Costa%20Rica%20PNDIP%20%202019-2022.pdf

Mojica Barquero, A. (2021). Development of listening comprehension skills in 8th graders of the bilingual experimental high school in Santa Cruz by the promotion of local belonging during online learning. In J. P. Zúñiga, & J. I. Barrantes (Eds.), *English language teaching in Costa Rica: Reflections on*

emergent challenges. División de Educología de la Universidad Nacional, Costa Rica. http://hdl.handle.net/11056/21301

Norouzian, R., & Plonsky, L. (2018). Eta-and partial eta-squared in L2 research: A cautionary review and guide to more appropriate usage. *Second Language Research*, *34*(2), 257-271. https://doi.org/10.1177/0267658316684904

Observatorio Latinoamericano de Políticas Educativas. (2019, September 22). *Bachillerato se despide tras 30 años de aplicación*. Observatorioeducacion.org. https://observatorioeducacion.org/noticias/bachillerato-se-despide-tras-30-anos-de-aplicacion

O'Sullivan, B. (2021). *The Comprehensive Learning System*. British Council. https://www.britishcouncil.org/sites/default/files/the_comprehensive_learning_system_new_layout.pdf

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Programa Estado de la Nación. (2021). *Octavo estado de la educación*. CONARE-PEN. https://estadonacion.or.cr/wp-content/uploads/2021/09/Educacion_WEB.pdf

Programa Estado de la Nación. (2023). *Noveno estado de la educación*. CONARE-PEN. https://estadonacion.or.cr/wp-content/uploads/2023/08/EE-2023-Book-DIGITAL.pdf

Quesada Pacheco, A., & Araya Garita W. (in press). Large-scale online English language testing in Costa Rica: The pioneering PELEx experience. In T. Clark & E. Bruce (Eds), *Post-pandemic remote online language testing*. Routledge

Quesada Pacheco, A., Araya Garita, W., & Fallas Godínez, J. A. (2023). *La enseñanza y aprendizaje del inglés en la secundaria pública costarricense del siglo XXI: innovaciones, brechas y desafíos*. Repository CONARE. https://repositorio.conare.ac.cr/handle/20.500.12337/8517

Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135-147. https://doi.org/10.1016/j.edurev.2010.12.001

Rojas Díaz, K. (2021, April 20). *Costa Rica fortalece su ruta hacia el bilingüismo con Política Educativa de Promoción de Idiomas*. Ministerio de Educación Pública (MEP). https://www.mep.go.cr/noticias/costa-rica-fortalece-su-ruta-hacia-bilingueismo-politica-educativa-promocion-idiomas-0#:~:text=Costa%20Rica%20fortalece%20su%20ruta%20hacia%20el,de%20lenguas%20extranjeras%2C%20LESCO%20y%20lenguas%20ind%C3%ADgenas.

Tavarez Da Costa, P., & Reyes Aria, F. (2021). *Of standardized student measurements and tests in the Dominican Republic* (ED611783). ERIC. https://files.eric.ed.gov/fulltext/ED611783.pdf

Universidad de Costa Rica. (2021). *Informe de Resultados Pruebas de Dominio Lingüístico (PDL) y Test for English of Young Learners (TEYL)*. Escuela de Lenguas Modernas. https://cdn.observador.cr/wp-content/uploads/2022/01/Resultados-PDL-y-TEYL-Nota-te%CC%81cnica-para-prensa-1-comprimido.pdf

Vargas González, M. F., Noguera Salgado, N., & Maroto Vargas, A. P. (2024). Carreras de enseñanza bilingüe: Un paso más en la ruta hacia la Costa Rica bilingüe. *Actualidades Investigativas en Educación*, *24*(2), 1–26. https://doi.org/10.15517/aie.v24i2.57146

Walczak, A., Harrison, G., Muratorio, M., Flores, C., Brunner, S., & Docherty, C. (2017). Which factors affect English language attainment? A study of school students in Chile. *Cambridge English: Research Notes, 65*, 51–65. https://www.cambridgeenglish.org/Images/368332-research-notes-65.pdf

West, R. M. (2021). Best practice in statistics: Use the Welch t-test when testing the difference between two groups. *Annals of Clinical Biochemistry, 58*(4), 267–269. https://doi.org/10.1177/0004563221992088

Appendix I

Secondary Education in Costa Rica (As stated by MEP, 2023)

This type of education in Costa Rica has two stages or cycles:

III Stage / Cycle: Considered the first stage of secondary education, it teaches children in seventh, eighth, and ninth grades, who are around age 13 to 15. Special emphasis is placed in the study of Spanish, Science, Mathematics, Social studies, English, and Civic Education, which together are considered a set called *Educación General Básica* (Core General Education Subjects or EGB, for its acronym in Spanish),

Diversified Education: These educational establishments offer EGB subjects and what they call "materias complementarias" (supplementary subjects), which differ according to the specific type of high school. It teaches students around age 15 to 18 and should last 3 years maximum.

Diversified education institutions can be classified according to their schedule, the available specializations, and the subjects that they offer. According to the latter, there are academic high schools, technical high schools, and other options. These alternatives are briefly described below. Only those types of education mentioned in this paper have been explained here.

Academic High Schools

Traditional Academic High Schools (Colegios Académicos Tradicionales): These serve as the foundation upon which modifications for different types of institutions are based. Students receive three weekly English lessons.

Experimental Bilingual High Schools/Public Bilingual High Schools (Liceos Experimentales Bilingües/Liceos Bilingües Públicos): These institutions provide high quality education. They emphasize English as a second language by offering ten weekly lessons, taught by two teachers. Additionally, each school can choose specific subjects to be taught entirely in English.

Costa Rican Scientific High Schools (Colegios Científicos Costarricenses): These are considered more rigorous than other high schools that aim to prepare students for

higher education. They only offer instruction for tenth and eleventh graders. Candidates must take an admission test and students must pass all subjects each year or transfer to another institution. Students receive five weekly English lessons (45 minutes each).

Humanistic High Schools (Colegios Humanísticos, also called Pre-University Institutions): Designed for top-performing students, who must take an entrance exam, with an interest in literature, philosophy, history, arts, and culture. These institutions promote innovative methodologies, impactful social projects, research, and outreach. Students receive three weekly English lessons. They only offer instruction for tenth and eleventh graders.

Spanish-English Bilingual Classrooms (Secciones Bilingües Español-Inglés): These are tailored English classes within some academic high schools and which are available to only selected students. The curriculum includes EGB subjects with an emphasis on English and English Literature, amounting to ten weekly English lessons.

Marco Tulio Salazar Virtual High School (Colegio Virtual Marco Tulio Salazar, scheduled for technical closure in 2023): This institution employs a pedagogical approach based on self-learning techniques and uses a hybrid learning model that combines distance and in-person education. The program includes three weekly English lessons.

Technical High Schools

Professional Technical Education (Educación Técnica Profesional, also called "Colegios Técnicos"): The academic curriculum offers technology workshops and English for Specific Purposes (ESP). The program lasts three years and includes a technical branch with a specialized curriculum: industrial, commercial and services, or agricultural. It places special emphasis on mastering at least one second language with subjects such as Technical English. Students receive 2 to 4 English lessons per week according to the specialization they are enrolled in.

Integrated Adult Education Centers (CINDEA - Centros Integrados de Educación de Adultos): These institutions cater to individuals who have not completed primary or secondary education or seek technical training. Programs are flexible and

accommodating to students' interests, learning pace, and participation possibilities. CINDEA include academic, technical, and artistic branches, with second language offerings varying by institution, up to a maximum of five lessons per cycle.

Other Educational Options

Night Academic High Schools (Colegios Académicos Nocturnos): In Diversified Education, the offer includes Human Development courses. Students receive five weekly English lessons, which include conversational English classes.

Community Education Professional Institutes (IPEC - Institutos Profesionales de Educación Comunitaria): For individuals aged 15 and older who have not completed primary or secondary education. These institutions combine formal and non-formal education with a focus on lifelong learning, providing free courses and technical training programs. The Diversified Education component requires five hours of a second language per week.

National State Distance High School (CONED - Colegio Nacional Estatal a Distancia): This distance learning secondary education type came to be through a partnership between MEP and UNED (National Distance Education University), and it uses tutoring and virtual resources that address the needs, interests, aspirations, and self-learning capacities of the target population. The curriculum includes four English lessons per level.

Appendix II

Although the original analysis was not designed as a multilevel study, we later explored whether nesting by educational region (n = 27) might influence the results. This additional model was not intended to replace the primary analysis but was conducted as a follow-up to check whether accounting for regional clustering would alter the observed group differences.

To assess the potential impact of clustering, a model with a random effect for educational region was estimated. The intraclass correlation coefficient (ICC = 0.115) indicated that a modest portion of the variance in scores was attributable to region-level differences. The corresponding design effect (DEFT = 16.90), based on the ICC and average cluster size, suggested that ignoring the nested structure could influence standard errors.

We then estimated a multilevel model including school type as a fixed effect and region as a random intercept. The results closely mirrored those of the original ANOVA, both in terms of statistical significance and the relative differences across types (see Table A1). Because no individual-level predictors were available and school type reflects a contextual characteristic, the multilevel model served as a robustness check rather than a primary analytic strategy.

Table A1. Multilevel model estimates for school type based on educational region clustering

Parameter	Coefficient	SE	95% CI	z	p
(Intercept)	36.68	0.37	[35.95, 37.41]	98.40	< .001
Scientific	3.98	0.48	[3.04, 4.92]	8.32	< .001
Humanistic	-0.76	0.64	[-2.02, 0.49]	-1.19	0.233
Semi-private	-1.37	0.23	[-1.83, -0.92]	-5.92	< .001
Experimental Bilingual	-1.87	0.18	[-2.22, -1.52]	-10.36	< .001
Bilingual Groups	-2.86	0.48	[-3.79, -1.93]	-6.02	< .001
Technical	-9.64	0.12	[-9.88, -9.41]	-80.06	< .001
Academic Day	-13.07	0.11	[-13.29, -12.84]	-115.04	< .001
CONED	-15.17	0.28	[-15.71, -14.63]	-54.71	< .001
CINDEA	-16.06	0.13	[-16.32, -15.80]	-121.84	< .001
IPEC	-16.21	0.23	[-16.66, -15.77]	-71.75	< .001
Academic Night	-16.50	0.14	[-16.77, -16.23]	-119.64	< .001
Virtual	-17.50	0.32	[-18.13, -16.87]	-54.70	< .001
Prison	-18.62	1.02	[-20.61, -16.63]	-18.30	< .001

Random Effects

SD (Intercept: Region) = 1.85

SD (Residual)= 6.53

In sum, accounting for educational regional clustering did not alter the conclusions. For clarity and consistency, we report the ANOVA results as the main analysis and include this multilevel extension to confirm the stability of the findings.