Colombian Eleventh Graders' English Performance and Motivation before, during, and after the COVID-19 Pandemic

Desempeño y motivación en inglés de estudiantes colombianos de undécimo grado antes, durante y después del COVID-19

Edwin Nel Viáfara Idrobo

Bachelor's Degree in Foreign Languages, English and French Foreign Languages, Universidad del Valle, Cali (Colombia).

viafara@correounivalle.edu.co

https://orcid.org/0009-0000-5223-3058

María Eugenia Guapacha Chamorro

PhD in Applied Linguistics. Associate Professor, School of Language Sciences, Universidad del Valle, Cali (Colombia).

maria.guapacha@correounivalle.edu.co

https://orcid.org/0000-0002-3769-331X

A B S T R A C T

This study investigated the impact of COVID-19 on eleventh graders' English performance and motivation in a private school. A mixed-methods approach was employed to identify four points: a) differences in students' performance on the Saber 11 English test across six cohorts (2017 to 2022); b) differences in students' performance on the school's in-house test across three cohorts (2020 to 2022); c) the relationship between both English tests; and d) students' perceptions of their learning and motivation during and after the pandemic. The results revealed that there were no statistically significant differences in the Saber 11 English test scores across cohorts; there were statistically significant differences in the in-house test scores, with the post-pandemic cohort having a higher performance; there were no strong relationships between both tests; and students perceived that their learning and motivation improved after the pandemic due to face-to-face interactions. The study concluded that the COVID-19 pandemic did not impact students' performance significantly; hence, the inferences made about students' performance are valid. Students' positive perception of their learning and motivation post-pandemic might have positively influenced their performance. It is also concluded that the school's in-house test did not have a strong predictive validity or correlation with the Saber 11 English test. The consequences of English tests on students and institutions are discussed.

Keywords: COVID-19, Saber 11 English test, in-house test, consequential validity, predictive validity, mixed methods research..

R E S U M E N

Este estudio investigó el impacto del COVID-19 en el desempeño y la motivación en inglés de estudiantes de undécimo grado en una escuela privada. Se empleó un enfoque de métodos mixtos para identificar: a) diferencias en el desempeño de los estudiantes en la prueba Saber 11 en inglés de seis cohortes (2017 a 2022); b) diferencias en el desempeño de los estudiantes en la prueba interna de la escuela de tres cohortes (2020 a 2022); c) la relación entre ambas pruebas de inglés; y d) las percepciones de los estudiantes sobre su aprendizaje y motivación durante y después de la pandemia. Los resultados revelaron que no hubo diferencias estadísticamente significativas en los puntajes de las pruebas de inglés Saber 11 entre cohortes; hubo diferencias estadísticamente significativas en los puntajes de las pruebas internas, en las cuales la cohorte pospandemia tuvo un mejor desempeño; no hubo relación entre ambas pruebas; y los estudiantes percibieron que su aprendizaje y motivación mejoraron después de la pandemia debido a las interacciones presenciales. El estudio concluyó que la pandemia de COVID-19 no afectó significativamente el desempeño de los estudiantes, por lo tanto, las inferencias sobre los desempeños son válidas. La percepción positiva de aprendizaje y motivación pospandemia puedo haber influido en los desempeños. También se concluyó que la prueba interna de la escuela no tiene una fuerte validez predictiva con relación a la prueba Saber 11. Se discuten las consecuencias de las pruebas de inglés en los estudiantes y las instituciones.
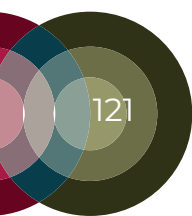
Palabras clave: COVID-19, prueba Saber 11 en inglés, prueba interna, impacto y validez predictiva, investigación con métodos mixtos.

Background

EFL Learners' Learning Performance before and during COVID-19

Before the pandemic, previous studies around the globe investigated factors influencing EFL learners' performance in the classroom, with a major focus on EFL university learners. For instance, shyness, peer pressure, anxiety, fear of making mistakes, and the lack of exposure to the target language hindered Saudi EFL university learners' performance (Rasheeda, 2020). In contrast, dynamic assessment approaches were positive in improving Indonesian EFL university learners' learning and performance (Partlo & Zahruni, 2020) and Iranian language institute learners (Estaji & Farahanynia, 2019). Mobile-game apps were also helpful to improve EFL university learners in Taiwan (Chen et al., 2019). It was also found that teachers' roles and students' autonomy are relevant factors for successful online teaching and learning (Salgado, 2015). Regarding EFL learners' performance in standardized testing, previous research investigated the validity of language tests (e.g., Borger, 2019) and test-takers' perceptions of the tests (e.g., Galikyan et al., 2019), highlighting the importance of the test task, context, conditions, and training in the validity of language tests and test-takers' performance.

With the surge of the COVID-19 pandemic in 2020 and the abrupt shift from traditional face-to-face learning to online learning, unprecedented challenges were faced by teachers and students (Phakiti & Isaacs, 2021), having notable consequences on students' personal, social, and academic lives. Issues and challenges included the closure of schools, the lack of essential tools for online learning, such as computers and internet access, the lack of accessibility of teaching materials, and device compatibility (Mahyoob, 2020; Pérez, 2021; Sanz et al., 2020). Other challenges included the lack of innovative assessment methods, students' lack of autonomy and technological skills, decreased teacher-student contact, and exam stress (Sandvik et al., 2022).

All these issues limited access to education across several educational and geographical contexts (Tao & Gao, 2022), suggesting the need to analyze the effectiveness of distance learning (Sanz et al., 2020) and assessment practices (Sandvik et al., 2022) in various subjects and scenarios.

In the Colombian context, prior to the pandemic, previous research documented the effectiveness of authentic resources and communicative and alternative methodologies, such as task-based language teaching (Castillo et al., 2017), flipped classrooms (Chivata & Oviedo, 2018), and virtual environments (Vega et al., 2017), in improving EFL learners' learning and performance in the classroom. Regarding performance in standardized tests, few studies investigated students' English performance in the national Saber 11 test. The Saber 11 English test is a high-stakes test for Colombian eleventh graders
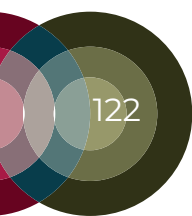
(last-year secondary school students), with passing or failing the test having positive or negative consequences on students' lives (e.g., access to higher education). Such studies are needed to shed light on the accurate effect of the pandemic on students' performance on the Saber 11 English test and their learning motivation.

Before the pandemic, previous research on the Saber 11 English test identified disparities between public and private schools in test results, with public school students showing a lower or basic level of English performance (Alonso et al., 2012, 2017; Gutiérrez & Mayora, 2021), confirming that bilingualism levels of Colombian secondary school students remain low (Díaz & Saldana, 2020). It was also found that computers and the Internet at home had a positive impact on English language performance in the Saber 11 test (Barrios et al., 2021).

During the pandemic, few studies compared the Saber 11 test results before and during COVID-19 to identify differences in students' performance. Findings in this regard are mixed. Díaz (2021) compared the Saber 11 test results before and during the pandemic (2019 and 2020) across five components (critical reading, math, English, social sciences, and sciences) in a public school in Cundinamarca. He found that there were no statistically significant differences in the performance of both cohorts, indicating no effect of the pandemic on students' performance. In contrast, Ballesteros and Gómez's (2022) study claimed that there was an effect of the pandemic on students' performance in Boyacá, as there was a downward trend in the scores during the pandemic. The authors also found that public school students performed lower than private school students. This study relied on descriptive statistics to claim differences in scores. However, there were no statistical analyses comparing scores before, during, and after the pandemic to provide more accurate insights into the pandemic's effect on this test.

EFL Learners' Perceptions of Learning and Motivation before and during COVID-19

Research on students' motivation and learning before the pandemic reported positive results in improving EFL language skills and language learning motivation through innovative methodologies, such as the flipped classroom model (Abdullah et al. 2019) and resources, such as YouTube (Alkathiri, 2019). Other studies investigated the relationship between intrinsic and extrinsic factors influencing language skills (e.g., Putra, 2017; Salehpour & Roohani, 2020).

During the pandemic, students worldwide encountered challenges in maintaining motivation and commitment to their studies due to the abrupt shift to online learning. While some students expressed
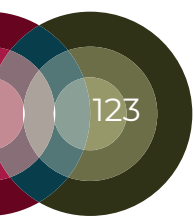
satisfaction with technology and software for online learning (Alwehebi, 2021), this shift impacted other students and teachers negatively (Chauri et al., 2021). Katz (2021) found that corrective feedback, positive comments, and personal thematic discourse were relevant in fostering positive relationships between teachers and students and in improving students' performance at a university in Hawaii that transitioned from face-to-face to virtual classes. Similarly, Yang et al. (2021) identified that learner autonomy, social interaction, teacher support, and a well-structured online learning environment were positive factors influencing Chinese EFL university students' emotions in an online learning environment. Chauri et al.'s (2021) study with EFL university students in Alcalá revealed students' positive motivation during online learning due to teachers' commitment to sustaining students' motivation during the pandemic. In contrast, Cui (2021) identified that Chinese EFL university students expressed difficulty maintaining motivation due to the lack of face-to-face interaction but had positive perceptions of teaching quality.

In Colombia, before the pandemic, learner motivation was also linked with authentic learning resources and innovative methodologies (e.g., Vega et al., 2017). During the pandemic, Valdivieso et al. (2020) identified that university students from different programs and semesters in Tunja perceived a negative impact of COVID-19 on their performance and motivation due to the lack of face-to-face interactions. Similarly, EFL public school teachers' perceptions of online teaching and learning yielded that the lack of face-to-face interaction, a limited access to technological resources, students' lack of motivation, and personal and family-related factors hindered online teaching (Garcia, 2021). However, there is a lack of studies investigating learners' motivation during and after the pandemic.

Overall, previous studies provide an understanding of the challenges and opportunities associated with EFL learning during the COVID-19 pandemic and the low levels of Colombian students' performance in the Saber 11 English test results before and during the pandemic. Previous studies also highligthed the significance of maintaining student motivation and engagement in online learning environments through positive feedback, effective interactions, teacher support, learning resources, and face-to-face interactional features in the virtual environment. Autonomy, social interaction, and well-structured online environments also play critical roles in students' emotions and motivation.

While these findings provide valuable information on EFL learners' performance and motivation, most of these studies were conducted with university students, whose motivation, learning processes, and performance might be different from EFL secondary school students. There is a notable gap in research investigating the Saber 11 English test results and learners' motivation before, during, and after the pandemic. It is relevant to extend research into high school contexts to understand the impact of the pandemic on this population.
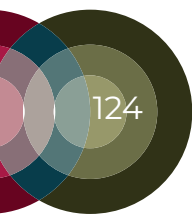
The Context: The Saber 11 English Test and the School's In-House English Test

In Colombia, the *Saber 11 English test* is a high-stakes examination, mandatory for higher education admission and used to evaluate quality of education. The test assesses eleventh graders' communicative competence in reading and language use. However, it does not evaluate oral and written skills, raising concerns about the comprehensiveness of this test in assessing students' language proficiency. The *Common European Framework of Reference* (CEFR) (Council of Europe, 2001) serves as a guideline to classify students' language performance into Basic user (pre-A1, A1, and A2 levels) and Independent user (B1 level). Eleventh graders must demonstrate a B1 level in the Saber 11 English test.

In the school where this study was conducted, eleventh graders also take an *in-house English test* aimed at familiarizing students with the Saber 11 English test and predicting their performance on this test. This internal test has been administered to eleventh graders since 2020; ninth graders also start taking it until they reach eleventh grade. The school's in-house English test is similar in structure to the Saber 11 test but differs in its administration and results reports. The internal test provides more detailed results, including breakdowns of components and competencies and students' strengths and weaknesses, which are not described in the Saber 11 test results. Table 1 summarizes both tests' features.

Table 1. Features of the Saber 11 English Test and the School's In-House Test

Features	Saber 11 English test	School's in-house English test
Components	Similarities	
	• Communicative competence: Explicit assessment (language use, reading, grammar, and vocabulary). • MCER guidelines: four levels (pre-A1, A1, A2, B1, B+). • Target Level: B1. • 55 questions. • Seven Components: vocabulary, pragmatics, communicative knowledge, grammar, reading comprehension and paraphrasing, text Interpretation, grammar and lexical proficiency.	
Administration	Similarities	
	• Required materials: pencil, eraser, sharpener. • Answer sheet.	
	Differences	
	• Paper-based test. • In-person test. • Registration via ICFES platform. • Administered at designated locations (universities, schools, etc.). • Supervised by designated ICFES staff. • Administered in the mid-school year.	• Online test in 2020 (in the pandemic). • In-person test. • Conducted by Los Tres Editores company. • Administered at school. • Supervised by an assigned teacher. • Administered one or two months before the Saber 11 test.
Test results delivery	Similarities	
	• Guidelines for result interpretation.	
	Differences	

	• Digital results on the ICFES website.
	• Five performance levels:
	  ○ Pre-A1 (0-47)
	  ○ A1 (48-57)
	  ○ A2 (58-67)
	  ○ B1 (68-78)
	  ○ B+ (79-100)
	• Overall analysis of performance:
	  ○ Learning levels and differences among students.
	  ○ Comparisons with reference groups.
		• Physical and digital (PDF) results.
		• Four performance levels:
		  ○ Unsatisfactory (0-40)
		  ○ Minimum (41-55)
		  ○ Satisfactory (56-70)
		  ○ Advanced (71-100)
		
		• Detailed analysis of performance:
		  ○ Analysis by area and student.
		  ○ Individual and group scores Numeric and graphical outputs (mean, standard deviation).
		  ○ Component results in percentages.
		  ○ Competence and component breakdown.
		  ○ Identification of strengths and weaknesses.

Source: own elaboration.
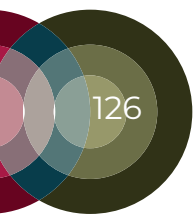
Conceptual framework: consequential validity and predictive validity

Since this study deals with test impact and consequences and how a test predicts students' performance on another test, the notions of consequential and predictive validities in language assessment are reviewed in this section.
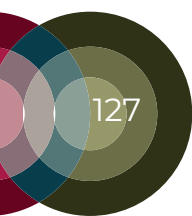
Consequential Validity

Consequential validity (or test impact) is concerned with the consequences of test scores on individuals and institutions (Messick, 1996; Bachman & Palmer, 2010; Brown & Abeywickrama, 2019; Green, 2014; Weir, 2005). For instance, the Saber 11 test can affect students' academic and job opportunities and institutions' rankings based on these test outcomes. The pandemic further complicated this scenario, introducing variables such as limited access to educational resources, changes in teaching practices, and potential disparities in the test's representativeness of students' actual performance.

To understand the broader effects of language assessments, it is necessary to differentiate between *impact* and *washback*. *Test impact* refers to the societal effects of tests, encompassing both intended and unintended consequences (Bachman & Palmer, 2010; Green, 2014), whereas *washback* focuses on the localized effects of assessments on teaching and learning practices (Bachman & Palmer, 2010; Brown & Abeywickrama, 2019). These concepts are relevant in the present study as they help discern how the Saber 11 English test and the school's in-house test impacted students and the school before, during, and after the COVID-19 pandemic.

Predictive Validity

Predictive validity is part of criterion-related validity, which is associated with criterion-referenced tests and concerned with how well a language test aligns with predetermined criteria and intended performance levels (Brown & Abeywickrama, 2019; Green, 2014; Weir, 2005). In the present study, the school's internal test is a criterion-referenced test, aligned with the Saber 11 English test's criteria and performance levels.

Criterion-related validity encompasses two forms: *concurrent* and *predictive*. *Concurrent validity* assesses the extent to which test results align with relatively recent performances, while *predictive validity* gauges a test's ability to forecast future performance (Green, 2014; Brown & Abeywickrama, 2019). Notably, predictive validity is relevant in the present study because the school's in-house test

is administered two months before the Saber 11 English test to familiarize students with it and predict their performance on the national test.

In sum, the notions of consequential validity and predictive validity provide an understanding of how test scores influence individuals and institutions, and how well criterion-referenced tests predict students' performance on another test. Therefore, the present study concentrates on students as the primary stakeholders who are directly impacted by the tests and the pandemic.

Research questions

Quantitative Research Questions

- RQ1. Are there any differences in the performance of six cohorts of students on the Saber 11 English test from 2017 to 2022?

- RQ2. Are there any differences in the performance of three cohorts of students on the school's internal test from 2020 to 2022?

- RQ3. Is there any relationship between the school's internal test scores and the Saber 11 English test scores from 2020 to 2022, and what do these results indicate regarding the internal test's predictive validity?

Qualitative Research Question

- RQ4. What are the students' perceptions regarding learning and motivation during and after the COVID-19 pandemic?

Methods

Research Design

This mixed-methods study (Creswell & Creswell, 2018) facilitated an understanding of the impact of COVID-19 on eleventh graders' performance and motivation from quantitative and qualitative perspectives. Specifically, the quantitative component involved analyzing test scores across cohorts and conditions (before, during, and after the pandemic), while the qualitative phase explored students' perceptions of their learning, performance, and motivation during and after the pandemic collected
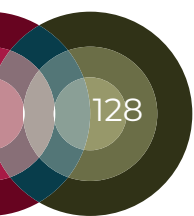
through a questionnaire. The study distinguishes among three periods: the pre-pandemic period (2017-2019), the pandemic period (2020-2021), and the post-pandemic period (2022). The school resumed full in-person teaching in 2022 when more COVID-19 restrictions were lifted in Colombia. Consequently, these periods highlight shifts in learning modalities, teaching resources, and overall school dynamics.

Research Site and Participants

The research site is a private school in Cali, Colombia, serving to socioeconomic strata 1 and 2 (lower class). This school was chosen based on research convenience, accessibility, and purpose; specifically, we had access to data regarding students' performance and permission from the stakeholders to conduct this study. One of the researchers is an instructor in this school.

The school is equipped with an English classroom with internet, projectors, and audio to support task-based language teaching, though instruction still emphasizes grammar. However, limited class time (two hours weekly) and focus on grammar and reading constrain communicative skills, which may affect performance on the Saber 11 test. Eleventh graders are not explicitly prepared for this exam but take external mock exams and, recently, in-house tests designed to simulate it.

The study comprises six cohorts (2017–2022, N=166) for Saber 11 scores, three cohorts (2020–2022, N=71) for internal test scores, and qualitative data from 16 eleventh graders in 2022 (ages 15–18; 10 girls, 6 boys). These cohorts are categorized as pre-pandemic (2017–2019), during (2020–2021), and post-pandemic (2022).

Pre-pandemic (2017–2019): There was no dedicated English classroom or technology; teaching relied on textbooks, photocopies, and Spanish. In 2019, a TV-equipped classroom was added. Students took the pre-ICFES but not the in-house test.

During the pandemic (2020–2021): Classes transitioned to an online format via Microsoft Teams, utilizing textbooks, computers, or smartphones at home. Internet issues, poor sound, and distractions limited learning; by other hand, participation was low, and Spanish predominated. Eventually, in-person classes resumed gradually in 2021. Both the Saber 11 and in-house tests were administered under safety protocols.

Post-pandemic (2022): Students returned to full-time in-person classes. The English classroom was equipped with Internet, TV, and a computer, in addition to an autonomous work platform. Significantly,
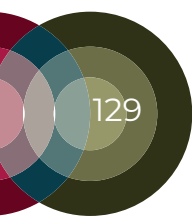
the school community valued the return. Eleventh graders took the pre-ICFES, while teachers reinforced Saber 11 content. Lastly, the in-house test was also administered in person.

## Data Collection Instruments and Analyses

Saber 11 English Test Scores

The Saber 11 test scores from six cohorts (2017-2022) were analyzed using a one-way independent measures ANOVA (Pallant, 2016). This statistical method helped to identify differences in English performance across six years, classified into three periods (before, during, and after the pandemic), thereby allowing us to identify any test impact due to pandemic conditions. The data met the assumptions for performing ANOVA, including a continuous dependent variable (Saber 11 scores), more than two independent groups (cohorts), independence of observations, and homogeneity of variances. Normal distribution was achieved in four of the six datasets, with both ANOVA and Kruskal-Wallis tests yielding similar statistical results.

The School's in-House English Test Scores

The school's in-house English test scores were used to perform two analyses. In the first analysis, test scores from 2020 to 2022 were compared using a one-way independent measures ANOVA to identify any statistical differences in students' performance during and after the pandemic. Assumptions for ANOVA were met, including a continuous dependent variable (internal test scores), more than two independent groups (cohorts), and independence of observations. Although a normal distribution was observed in only two of the three datasets, ANOVA and Kruskal-Wallis tests showed similar statistical results.

Regarding the second analysis, correlations were performed between the Saber 11 English test scores and the school's in-house test results across three cohorts (2020 to 2022) to identify any relationship between the tests and assess the school test's predictive validity. The results of Spearman's rho correlations are reported.

The quantitative analyses were performed using Excel and the free statistical software *JASP*. These analyses provided insights into the tests' impact on students' performance before, during, and after the pandemic to enable valid inferences regarding test scores.
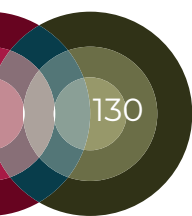
Perceptions Questionnaire

A qualitative survey comprising 13 open-ended questions was administered in Spanish to cohort 2022 (N=16) in October 2022. Conducted in the school's computer lab, it focused on students' learning, performance, motivation, and test perceptions during and after the pandemic, thereby providing insights into test validity. The survey, which was designed by the researchers, reviewed by an expert, and piloted with two students, was chosen over interviews due to time and access constraints.

The questionnaire comprised four sections:

I. Learning overall and across skills during/after COVID-19 (five questions).

II. Assessment and performance on Saber 11 (four questions).

III. Language learning resources (one question).

IV. Motivation and related factors (three questions).

Subsequently, data were analyzed through content analysis (Krippendorff, 2013), following Creswell and Creswell's (2018) steps: preparing and organizing responses, coding with in vivo and color-coded labels (positive/negative, during/after pandemic), generating themes, and presenting results with quotes, tables, and graphs.

Validity, Reliability and Ethics

Validity and reliability ensure the accuracy and relevance of the instruments used (Creswell & Creswell, 2018). Validity was established by implementing the following measures: The instruments were designed to answer the research questions and to assess the impact of COVID-19 on students' English performance and perceptions; also survey questions were revised and piloted; ANOVA, correlations, and content analysis were used to ensure accurate analysis of relationships and impacts; and the sample size was adequate for statistical and qualitative analyses. Reliability was ensured through consistent data collection methods and repeated analyses conducted by both researchers. Strict ethical standards were followed, especially given the participation of minors. Informed consent was obtained from students and the school principal. Anonymity was maintained by using codes instead of names; participation was voluntary, with no risk of harm. Ultimately, the research benefits the school by providing valuable insights into English performance, as the principal authorized the study and expressed interest in its outcomes.
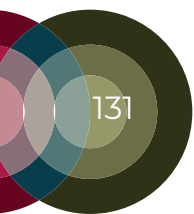
Findings

RQ1. Differences in the Saber 11 English Test Scores from 2017 to 2022

The Saber 11 English test results of six cohorts of eleventh graders (N = 166) were compared to identify any statistically significant differences before (2017 to 2019), during (2020 and 2021), and after the pandemic (2022). With a maximum score of 100, the test categorizes performance into Basic User (pre-A1 (0-47), A1 (48-57), and A2 (58-67)) and Independent User (B1 (68-78) and B+ (79-100)).

The descriptive statistics in Table 2 reveal consistently low performance levels (between 46.7 and 51.1 points) in the Saber 11 English test across the six cohorts (2017 to 2022). These levels categorize students as Basic Users, regardless of the pandemic periods, indicating limited proficiency in using the language effectively in real communicative situations. Notably, the average scores demonstrated a slightly upward trend during (2021) and after (2022) the COVID-19 pandemic, reaching the highest mean in the post-pandemic (2022). The standard deviation decreased during and after the pandemic, highlighting more consistent scores in these periods.

Table 2. Descriptive Statistics Saber 11 English Test Scores across Cohorts (2017-2022)

	Before COVID-19			During COVID-19		After COVID-19
	2017	2018	2019	2020	2021	2022
Sample (n)	29	36	29	22	32	18
Mean	47.5	47.4	46.7	47.0	48.7	51.1
Standard Deviation	9.6	10.1	10.7	5.85	9.9	7.1
Skewness	1.81	0.58	0.03	-0.24	1.01	-0.70
Kurtosis	6.33	0.79	-0.73	0.43	1.97	-0.40
Shapiro-Wilk	.858	.968	.979	.955	.923	.927
P-value of Shapiro-Wilk	.001	.384	.812	.402	.025	.174
Minimum	32.0	30.0	27.0	34.0	31.0	37.0
Maximum	84.0	76.0	69.0	58.0	79.0	62.0

Source: own elaboration.

Then, to identify statistically significant differences in the Saber 11 English test scores across the six cohorts and periods, a one-way independent-subjects ANOVA was performed. As shown in Table 3, the results indicate that there were no statistically significant differences in students' performance in the Saber 11 English test before, during, and after the pandemic [F (5, 160) = .607, p = .695, η^2 = .019 (small effect size)]. Given that there were no significant differences among cohorts, no post hoc tests were performed. Overall, these findings suggest that the pandemic did not significantly impact student performance on the Saber 11 English exam.

Table 3. One-Way ANOVA on Saber 11 English Test Scores
(2017-2022)

ANOVA results—Saber 11 English test scores (2017-2022)						
Cases	Sum of Squares	df	Mean Square	F	p	η^2
Cohort	267.627	5	53.525	0.607	.695	.019
Residuals	14114.885	160	88.218			

Note. Type III Sum of Squares.
Source: own elaboration.

RQ2. Differences in the School's Internal Test Scores from 2020 to 2022

Initially, the internal test was introduced to the students in 2020. The performance of three cohorts of eleventh graders (N = 71) was compared across three years, during (2020-2021) and after the pandemic (2022), to identify any differences in scores. The descriptive statistics in Table 4 demonstrate consistently low performance (between 39.0 and 59.6 points) in the internal test results across the three years (2020 to 2022), classifying these students as Basic Users. Notably, students started at the pre-A1 level during the initial Covid period (2020), with subsequent improvements to levels A1 and A2 in 2021 and 2022.

Table 4. Descriptive Statistics In-House English Test
Scores across Cohorts (2020-2022)

Descriptive statistics		
	During COVID-19		After COVID-19
	2020	2021	2022
Sample (n)	22	32	17
Mean	39.0	57.5	59.6
Standard Deviation	9.92	21.5	11.6
Skewness	0.03	-0.33	-1.92
Kurtosis	-1.39	-0.95	4.82
Shapiro-Wilk	0.93	0.94	0.82
P-value of Shapiro-Wilk	0.15	0.08	0.00
Minimum	23.0	11.0	24.0
Maximum	54.0	87.0	73.0

Source: own elaboration.

The results of both the Saber 11 English test and the school's internal test revealed similar levels of performance, which still fall short of reaching the B1 level required (between 68 and 78 points). The 2022 cohort, having had more experience with the internal exam since ninth grade, achieved the highest score, while the 2020 cohort, taking the internal exam for the first time that year, recorded the lowest score. This indicates that test familiarity influences students' performance on this internal test.

To identify statistically significant differences in scores across cohorts, a one-way independent subject ANOVA was performed. As shown in Table 5, there were statistically significant differences among the three cohorts during and after the pandemic [$F_{(2, 68)} = 10.280$, $p = .001$, $\eta^2 = .232$ (large effect size)], suggesting a distinct impact of COVID-19 on students' performance.

Table 5. One-Way ANOVA on School's In-House Test Results across Cohorts (2020-2022)

ANOVA results—school's in-house test scores (2020–2022)						
Cases	Sum of Squares	df	Mean Square	F	p	η^2
Cohort	5642.312	2	2821.156	10.280	.001	.232
Residuals	18661.576	68	274.435			

Note. Type III Sum of Squares.
Source: own elaboration.

Given the observed disparities across the three cohorts, post hoc tests were conducted to identify specific where differences occurred. As illustrated in the post hoc analysis in Table 6, statistically significant differences were revealed among the cohorts. In particular, the 2020 cohort demonstrated markedly lower performance than the 2021 and 2022 cohorts ($p < .001$), whereas no statistically significant difference was found between the 2021 and 2022 cohorts ($p = .908$). These results suggest that the 2020 cohort might have faced challenges during the initial pandemic period or lacked familiarity with the internal test, factors which likely impacted their performance. In contrast, the 2021 and 2022 cohorts exhibited comparable performances.

Table 6. Post Hoc Test on School's In-House Test Scores

Post Hoc comparisons						
Cohorts		Mean Difference	SE	T	Cohen's d	ptukey
2020	2021	-18.472	4.588	-4.026	-1.115	< .001
	2022	-20.556	5.350	-3.843	-1.241	< .001
2021	2022	-2.085	4.972	-0.419	-0.126	0.908

Note. P-value adjusted for comparing a family of 3.
Source: own elaboration.

RQ3. Relationship between both English Tests and Results of the Internal Test's Predictive Validity

Since it is administered two months apart, the internal test helps eleventh graders familiarize themselves with the Saber 11 English test. Specifically, the comparison of scores between the internal test and the Saber 11 test, as shown in Table 7, reveals that the 2020 cohort performed better in the Saber 11 test (M = 47.0) than in the internal test (M = 39.0). Conversely, the subsequent cohorts of 2021 and 2022 performed better in the internal test (2021: M = 57.5; 2022: M = 59.6) compared to the Saber 11 English test (2021: M = 48.7; 2022: M = 50.7). This outcome indicates that students' familiarity with the school's internal test influences their performance positively.

Table 7. Descriptive Statistics In-House Test and Saber 11 Test Scores (2020-2022)

Descriptive statistics						
	During COVID-19				After COVID-19	
	Internal 2020	Saber11 2020	Internal 2021	Saber11 2021	Internal 2022	Saber11 2022
Sample (n)	22	22	32	32	17	17
Mean	39.0	47.0	57.5	48.7	59.6	50.7
Standard Deviation	9.92	5.85	21.5	9.99	11.65	7.25
Skewness	0.03	-0.24	-0.33	1.01	-1.92	-0.60
Kurtosis	-1.39	0.43	-0.95	1.97	4.82	-0.50
Shapiro-Wilk	0.93	0.95	0.94	0.92	0.82	0.93
P-value of Shapiro-Wilk	0.15	0.40	0.08	0.02	0.00	0.29
Minimum	23.0	34.0	11.0	31.0	24.0	37.0
Maximum	54.0	58.0	87.0	79.0	73.0	62.0

Source: own elaboration.

Furthermore, to identify any relationship between test scores and assess the predictive validity of the school's in-house test, three separate correlations were performed (see Table 8). The results of the Spearman's rho test yielded no statistically significant correlation between the internal test and Saber 11 test scores for the 2020 (r = -0.223; p = .318) and 2022 (r = 0.154; p = .555) cohorts, indicating no relationship between these tests. Conversely, the 2021 cohort exhibited a significant negative correlation (r = -0.379; p = .032), suggesting an inverse relationship between the scores. In other words, an

increase in one test score corresponds to a decrease in the other. Despite placing students in the same proficiency level, the internal test lacks strong predictive validity for Saber 11 English performance.

Table 8. Spearman's rho Correlations on Internal and Saber 11 Test Scores (2020-2022)

		Saber 11 scores 2020
Internal test scores 2020	Pearson's r	-0.221
	p-value	0.322
	Effect size (Fisher's z)	-0.225
	Spearman's rho	**-0.223**
	p-value	**0.318**
	Effect size (Fisher's z)	-0.227
		Saber 11 scores 2021
Internal test scores 2021	Pearson's r	-0.158
	p-value	0.387
	Effect size (Fisher's z)	-0.160
	Spearman's rho	**-0.379**
	p-value	**0.032**
	Effect size (Fisher's z)	-0.399
		Saber 11 scores 2022
Internal test scores 2022	Pearson's r	-0.015
	p-value	0.954
	Effect size (Fisher's z)	-0.015
	Spearman's rho	**0.154**
	p-value	**0.555**
	Effect size (Fisher's z)	0.155

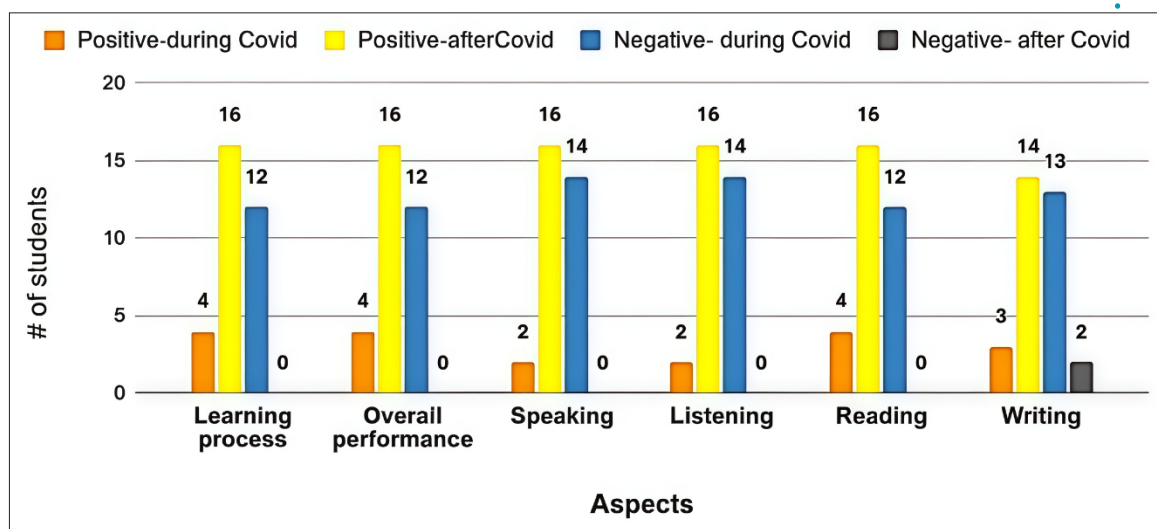**Source:** own elaboration

RQ4. Students' Perceptions of Learning, Performance and Motivation during and after the Pandemic

In total, 16 students from the 2022 cohort shared their perceptions of learning, performance, and motivation during and after the COVID-19 pandemic.

Perceptions of Learning overall and across Language Skills during and after COVID-19

Figure 1 illustrates students' positive and negative perceptions regarding their learning process and language skill development (speaking, listening, reading, and writing) during and after the COVID-19 pandemic.


Figure 1. Students' Positive and Negative Perceptions of Learning and Performance across Language Skills during and after COVID-19

Source: own elaboration

Across all categories, positive perceptions increased significantly after the pandemic (yellow bars). The highest number of students (*n* =16) expressed positive views after COVID-19 in every skill, indicating a perceived improvement in learning conditions or outcomes post-pandemic. Most negative responses were associated with the COVID-19 period (blue bars). This suggests that students experienced more challenges in language learning during the pandemic, particularly in listening (14 negative responses) and writing (13).

Overall, students perceived their English learning experiences more positively after the pandemic, especially in speaking, listening, and reading. However, the pandemic period was associated with higher dissatisfaction, particularly in receptive and productive skills like listening and writing. These
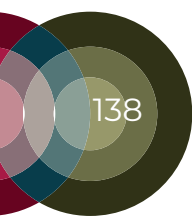
results highlight the negative impact of pandemic-related disruptions and the recovery of student confidence in language learning once regular instruction resumed.

For instance, S1 reported that her speaking ability was affected by the pandemic:

> After the pandemic, I felt I could interact more with the teacher and classmates. I always like to participate and talk in class, but I felt I did not do it so much during the online classes because they were short.

For other students, writing and listening were affected by online teaching during the pandemic. S13 expressed the following: "*We did not practice writing and listening so much. We had to do grammar and reading activities and textbook activities. Sometimes we watched some videos.*" The lack of writing practice during the pandemic reflected the constraints of the learning modality and the teaching methodology, whereas speaking was enhanced due to the face-to-face interactions.

Perceptions of Assessment and Learning Resources during and after COVID-19 and Performance on Saber 11 Test

All students indicated that the teacher's feedback, the assessment activities, and the learning resources were more varied before and after COVID-19. During the pandemic, they perceived that feedback was more general since the teacher could not focus on individual students. Moreover, most students expressed that, as there was no active student participation during the pandemic, there were very few feedback opportunities. Most of the class and assessment activities focused on reading, vocabulary, and grammar. In the post-pandemic period, there were additional opportunities to be assessed through role plays, listening activities, and class participation. The following quotes reflect students' perceptions of assessment during and after the pandemic.

> S11: "*I feel that we had to complete many worksheets and readings during the pandemic. Everything was about reading and practice activities.*"

> S9: "*After the pandemic, the teacher did the activities he used to do before, such as role plays, songs, videos, movies, and presentations.*"

> S6: "*I think that the reading activities were useful during the pandemic because we practiced for the test, but sometimes they were monotonous. After the pandemic, the games, videos, and presentations were more interesting and useful for our motivation and learning.*"

Concerning performance on the Saber 11 test, while a few students (n = 6) anticipated a good performance on this high-stakes test, many students (n = 10) indicated they felt insecure about their results. Potentially, this lack of confidence in their performance might have been influenced by the scarcity of interactions during the pandemic, even though interaction and additional practice opportunities increased after the pandemic. For instance, S2 reported that she "*likes English, so [she] felt that the test results would be positive, regardless of the pandemic.*" In contrast, S16 expressed concerns about her performance: "*I am not sure whether I will have very good results on the test. Although we practiced English in class during and after the pandemic, I do not feel very confident about my results.*" Collectively, these quotes reflect that, beyond the pandemic factor, individual variables, such as personal interest and learning motivation, might have also influenced students' test results.

Perceptions of Motivation during and after COVID-19

Figure 2 presents students' perceptions of motivation and related psychosocial factors during and after the COVID-19 pandemic, covering five aspects: motivation, physical/emotional state, teacher interaction, classmate communication, and class participation.

Figure 2. Students' Positive and Negative Motivation and
Engagement during and after COVID-19
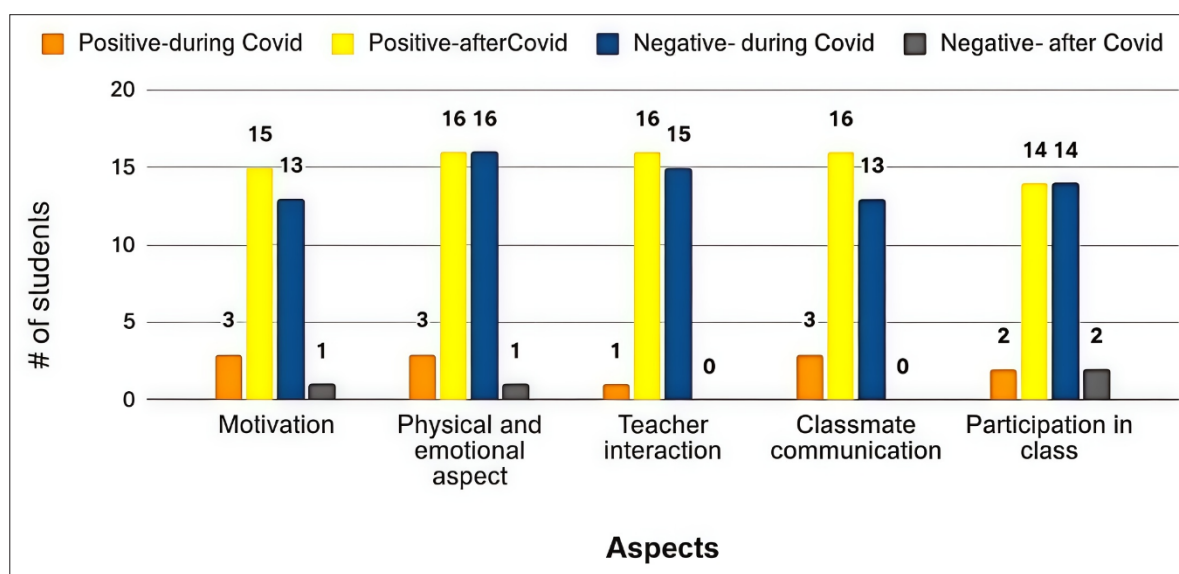
Source: own elaboration
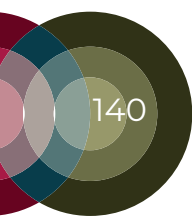
Across all categories, positive perceptions after COVID-19 (yellow bars) were consistently high (15–16 students), indicating a strong recovery in students' motivation and social engagement. Teacher interaction received unanimous positive feedback after the pandemic (16 students), highlighting its critical role in post-pandemic re-engagement. Negative during-Covid responses (blue bars) were also consistently high (13–16 students), suggesting students felt significant declines in motivation, emotional well-being, and social interaction during the pandemic. In particular, the physical and emotional aspects and teacher interaction had 16 negative responses each, showing major struggles in these areas.

In general, students overwhelmingly viewed their motivation, emotional well-being, and social engagement as significantly improved after the pandemic, particularly in terms of teacher and peer interaction. Conversely, during COVID-19, they experienced strong declines in these areas, especially in emotional health and participation. Therefore, these results underscore the vital importance of in-person interaction for student well-being and motivation in language learning contexts and their potential influence on test performance.

To illustrate, some quotes capture students' perceptions. Several students indicated they preferred in-person classes because there were more learning advantages and higher motivation. For example, student S13 mentioned the following:

> To me, face-to-face learning has had more advantages than disadvantages. I have been more motivated to learn. I wasn't motivated in virtual learning because I fell asleep and didn't have resources. Also, other personal factors didn't allow me to learn easily.

Likewise, student S16 expressed that "*during the pandemic, there were many factors in virtual settings that prevented [him] from developing the language; for example, external noises, headaches, or sight problems caused by the computer screen.*" Finally, for student S7, face-to-face interactions were preferred as he could interact with other students: "*I prefer face-to-face learning because I learn from watching and interacting with others and not just by listening behind a computer screen.*"

In summary, the students' answers reflect the perceived inadequacy of remote communication during the pandemic and the importance of emotional support post-pandemic. While the quantitative results suggested that the pandemic did not affect students' performance on the English tests, the qualitative results indicate students' positive attitudes towards English language learning and motivation post-pandemic.
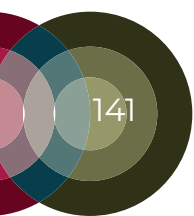
Discussion

This mixed-methods research investigated the impact of the COVID-19 pandemic on eleventh graders' performance and motivation in a private school in Cali, Colombia. Specifically, it investigated differences in the Saber 11 English test and the school's in-house English test scores before, during, and after the pandemic. It also investigated the relationship between both English tests and gathered insights into eleventh graders' perceptions of learning, performance, and motivation. For this purpose, the study analyzed the Saber 11 English test scores of six cohorts (2017 to 2022), the school's in-house English test scores of three cohorts (2020 to 2022), and the perceptions of 16 students who experienced the pandemic and post-pandemic periods. These analyses informed the consequential validity (test impact) of both English tests on students and the school, as well as the predictive validity of the school's in-house English test during those periods.

Regarding differences in test scores, the results from the Saber 11 and the school's in-house English tests revealed that students' performance across all cohorts fell within the basic level (between pre-A1 and A2 levels), regardless of the pandemic period. These results are consistent with previous research reporting that most students from lower socioeconomic strata show low proficiency levels in the Saber 11 English test (Alonso et al., 2012, 2017; Gutiérrez & Mayora, 2021) and that the bilingualism level of secondary school students remains low (Díaz & Saldana, 2020). Also, the results suggest that students from low socioeconomic backgrounds require better exposure to the target language to reach the B1 level (independent user). Therefore, there is a need for a comprehensive and sustained approach to enhancing English teaching and learning.

Notably, the results of both English tests showed a slight increase in average scores in the post-pandemic period (2022). To determine differences in test scores and the potential impact of COVID-19 on students' performance, statistical tests were performed. The results yielded no statistically significant differences in students' performance before, during, and after COVID-19 on the Saber 11 English test, thus aligning with previous research reporting no statistically significant differences in scores before and during the pandemic (Díaz, 2021). This suggests that the pandemic did not significantly affect students' performance on the Saber 11 English test in this school's context; hence, the interpretations of test scores are valid.

In contrast, there were statistically significant differences in the school's in-house test scores, with the 2020 cohort displaying the lowest level of performance. This result could be attributed to this cohort's lack of familiarity with the internal test and the potential impact of the pandemic that year. This cohort was introduced to this test for the first time in 2020, in contrast to the 2021 and 2022 cohorts who were exposed to this test in earlier grades. Additionally, changes and challenges generated by
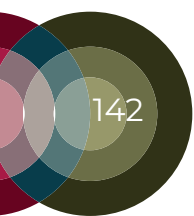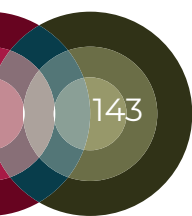
the pandemic, such as online learning and technological issues, might have adversely affected this specific cohort (2020).

The absence of a significant impact of the pandemic on students' Saber 11 English test performance might be associated with the school's readiness to implement strategies to cope with the challenges introduced by the pandemic. For instance, the school equipped classrooms with technological resources, including projectors, computers, and internet access, during and after the pandemic. Hence, students' increased use of technological tools during the pandemic had a beneficial impact on academic performance and learning (Díaz, 2021; Perez, 2021). Also, students started to attend school gradually in 2021; therefore, they were not affected by a predominant period of online instruction.

A further contributing factor might be the introduction of the internal test in 2020. This test helped students prepare for and familiarize themselves with the Saber 11 English test. Particularly, the 2022 cohort received more training and preparation for the national test since 2020; therefore, their scores were not greatly affected by the pandemic, as hypothesized. This familiarity feature, displayed by the 2022 cohort, might be the key factor in explaining why there were statistically significant differences in students' performance on the school's internal test. Unlike the 2022 cohort, the 2020 cohort faced initial difficulties attributed to the pandemic's immediate impact, lack of experience with the internal test, and resource constraints.

Concerning the relationship between the Saber 11 and the school's in-house test scores, students' performance was slightly higher on the in-house test. The correlational analysis indicated no consistent relationship in scores between both English tests. This suggests that the in-house test lacks strong predictive validity. Although the school's in-house English test helps eleventh graders to familiarize themselves with the structure of the Saber 11 English test, the in-house test does not accurately predict students' actual performance. The fact that the students' average scores were slightly higher than the Saber 11 test scores supports this claim. Other interpretations might be related to students' unfamiliarity with the Saber 11 test administration conditions, which might induce stress in students, impacting their performance. In contrast, students are familiar with the school's facilities and context, the teachers who administer and proctor the in-house test, and its structure. Additionally, preparation and personalized feedback from the institution allow targeted improvement, contrasting with the more generalized feedback from the national exam.

The results of students' perceptions of their learning, performance, and motivation during and after the pandemic revealed preferences for in-person learning. The return to face-to-face instruction was perceived positively by the students, as it improved interaction with the teacher and peers, content understanding, task delivery, development of language skills, and provided more learning opportuni-

ties, support, and confidence. These findings highlight the importance of effective interaction and feedback strategies in fostering positive relationships between teachers and students (Katz, 2021). Students' perceptions of improvement after the pandemic aligned with the slight improvement in average English test scores. This suggests that students' perceptions might provide insights into their strengths and weaknesses regarding learning and motivation. Although the improvement in performance was not statistically significant, small but positive changes are relevant for teachers and students.

Physical and emotional aspects were also perceived as improved after the pandemic, thus highlighting the importance of students' motivation, interaction, and emotional and physical well-being in the educational scenario (e.g., Cui, 2021; Chauri et al., 2021). The students' negative motivation during the pandemic was associated with stress, anxiety, isolation, resource limitations, and lack of interactions, confirming previous studies' findings on the challenges posed by the pandemic, including technical issues, difficulties in social interaction, a lack of feedback, and negative emotions (e.g., Cui, 2021; Chauri et al., 2021; Mahyoob, 2020).

Ultimately, these findings suggest that the end of the pandemic represented an improvement in students' emotional state and positive attitudes towards learning, which may have contributed to better performance on the tests. These improvements reflect students' adaptation to face-to-face learning, emphasizing the pivotal role of motivation in academic success and access to technological resources, particularly during disruptive changes in the educational environment.

Regarding the consequential validity of the Saber 11 and the school's in-house English test, both positively impact students and the institution by reflecting language abilities, providing feedback, and identifying areas for improvement. The in-house test also familiarizes students with testing procedures.

However, the in-house test shows weak predictive validity. While it often places students at low proficiency levels, these do not align numerically with Saber 11 scores. This misalignment, influenced by curriculum focus and test familiarity, may create false expectations. Thus, further score comparisons are needed to assess the in-house test's effectiveness in predicting Saber 11 performance.

Conclusions, implications, limitations and further research

This research concludes that the COVID-19 pandemic had a minimal impact on students' performance on the Saber 11 English test and the school's in-house English test. Also, face-to-face interactions foster students' learning and motivation, as indicated by students' perspectives. Although the pandemic influenced the use of new information and communication technologies in the virtual space and transformed traditional education, it did not significantly affect students' performance on this
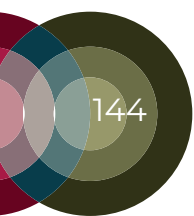
important English test. Potentially, the transitional phase towards dynamic and flexible methodologies during virtual teaching might have also allowed access to diverse pedagogical resources adaptable to the digital environment, thereby enhancing the students' experience.

The study also highlights institutions' and teachers' efforts to promote interaction, address technical challenges, and improve educational quality and equity during the pandemic. Crucially, it emphasizes the role of exams in shaping decisions, teaching, and infrastructure, noting that English performance must be understood in context. In this sense, schools and policymakers should analyze test scores across pre-, during, and post-pandemic periods to assess the pandemic's impact and adjust curricula, teaching, and assessment strategies. Communicative methods, technological tools, and continuous test preparation are recommended to strengthen proficiency and test-taking skills.

Key findings include:

- Comparing the Saber 11 and in-house test results revealed performance changes and post-pandemic improvements (e.g., 2022 cohort).

- Significant differences in in-house test scores (notably in 2020) were linked to pandemic conditions, test familiarity, and access to technology.

- Both tests provide valuable feedback for students and teachers to guide decision-making.

- The in-house test (introduced in 2020) aided Saber 11 preparation but lacks strong predictive validity.

- Students' perceptions confirmed slight post-pandemic performance gains, largely due to renewed in-person interaction.

The mixed methods approach provided insights into the impact of COVID-19 on EFL secondary school students' performance and motivation. The study contributes a unique perspective on the pandemic's influence in a specific educational context, emphasizing the need for contextualized analyses of test scores. Although the present study comprised several cohorts before, during, and after the pandemic, future research might involve a larger sample size and a broader timeframe for quantitative analysis and detailed interviews with a small sample of participants. In the present study, the scores of the 2023 cohort were not available when this study was conducted. Further research is also needed across regions and public and private schools to make comparisons and identify differences in impact based on sociodemographic and contextual factors. Finally, future research might explore the relationship between national and internal tests, delving into students' perceptions of learning in different environments.
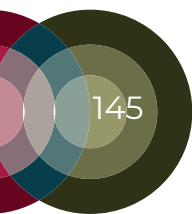
References

Abdullah, M. Y., Hussin, S., & Ismail, K. (2019). Investigating the effects of the flipped classroom model on Omani EFL learners' motivation level in English speaking performance. *Education and Information Technologies*, *24*, 2975-2995. https://doi.org/10.1007/s10639-019-09911-5

Alkathiri, L. A. (2019). Students' perspectives towards using Youtube in improving EFL Learners' motivation to speak. *Journal of Education and Culture Studies*, *3*(1), 12-30. http://dx.doi.org/10.22158/jecs.v3n1p12

Alrasheedi, S. (2020). Investigation of factors influencing speaking performance of Saudi EFL learners. *Arab World English Journal (AWEJ)*, *11*, 66-77. https://dx.doi.org/10.24093/awej/vol11no4.5

Alonso, J., Casasbuenas, P., Gallo, B., Torres, G. (2012). *Bilingüismo en Santiago de Cali: Análisis de los resultados de las Pruebas Saber 11 y Saber Pro*. Universidad Icesi. https://www.icesi.edu.co/centros-academicos/images/Centros/cienfi/libros/Bilinguismo_en_Santiago_de_Cali.pdf

Alonso, J., Díaz, D., Martínez, D., Mayora, C., Moreno, L., & Ochoa, M., Roldan, B. (2017). *Bilingüismo en Santiago de Cali: Análisis de la evolución de los resultados de las Pruebas Saber 11*. Universidad Icesi. https://www.icesi.edu.co/centros-academicos/images/Centros/cienfi/libros/Saber11_libroFINAL2.pdf

Alwehebi, K, A. (2021). The effect of online EFL classes during the Covid 19 pandemic: Evaluating the performance of university students and teachers in Saudi Arabia. *Review of International Geographical Education (RIGEO)*, *11*(5), 990-1006. DOI:10.48047/rigeo.11.05.95

Bachman, L. F., & Palmer, A. S. (2010). Language assessment in practice: Developing language assessments and justifying their use in the real world. Oxford University Press.

Ballesteros-Alfonso, A. L., & Gómez-Velasco, N. Y. (2022). Desigualdad de resultados pruebas Saber-11 antes y durante la pandemia covid-19 (2014-2021). *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, *20*(3), 1-23. https://dx.doi.org/10.11600/rlcsnj.20.3.5189

Barrios Aguirre, F., Forero, D. A., Castellanos Saavedra, M. P., & Mora Malagón, S. Y. (2021). *The Impact of computer and internet at home on academic results of the Saber 11 national exam in Colombia. SAGE Open*, *11*(3), 215824402110408. https://doi.org/10.1177/21582440211040810

Borger, L. (2019). Evaluating a high-stakes EFL speaking test: Teachers' practices and views. *Moderna språk*, *113*(1), 25-57. https://doi.org/10.58221/mosp.v113i1.7621

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

Castillo Losada, C. A., Insuasty, E. A., & Jaime Osorio, M. F. (2017). The impact of authentic materials and tasks on students' communicative competence at a Colombian language school. *Profile Issues in Teachers Professional Development*, *19*(1), 89-104. http://dx.doi.org/10.15446/profile.v19n1.56763
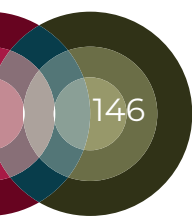
Chauri Galván, B., García Hernández, S., & Fernández-Gil, M. J. (2021). Teaching English online during the national lockdown: students' perceptions and experiences at a Spanish University. Íkala, *Revista De Lenguaje Y Cultura*, *26*(3), 603–621. https://doi.org/10.17533/udea.ikala.v26n3a08

Chen, C., Liu, H., & Hong-Bin, H. (2019). Effects of a mobile game-based English vocabulary learning app on learners' perceptions and learning performance: A case study of Taiwanese EFL learners. *ReCALL: The Journal of EUROCALL*, *31*(2), 170-188. https://doi.org/10.1017/S0958344018000228

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Chivata, Y. P., & Oviedo, R. C. (2018). EFL Students' Perceptions of Activeness During the Implementation of Flipped Learning Approach at a Colombian University. *Gist: Education and Learning Research Journal*, (17), 81-105. Retrieved from https://dialnet.unirioja.es/servlet/articulo?codigo=6758706

Creswell, J. W., & Creswell, J. D. (2018). *Research design: qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage.

Cui, Y. (2021). Perceived learning outcomes and interaction mode matter: students' experience of taking online EFL courses during covid-19. *English Language Teaching*, *14*(6), 84-95. https://doi.org/10.5539/elt.v14n6p84

Díaz Pinzón, J. E. (2021). Resultados comparativos de la prueba saber 11 antes y durante la pandemia por COVID-19 en una institución educativa de Soacha, Cundinamarca, Colombia. *Revista Repertorio De Medicina Y Cirugía*. *30* https://doi.org/10.31260/RepertMedCir.01217372.1259

Díaz Laínes, G., & Santana Sanabria, G. P. (2020). Programa nacional de bilingüismo y el nivel de desempeño en inglés de los estudiantes en Colombia. Armonización pedagógica. *Perspectivas e investigaciones sobre el futuro de la educación*, 241-258. https://www.researchgate.net/publication/342026203

Estaji, M., & Farahanynia, M. (2019). The immediate and delayed effect of dynamic assessment approaches on EFL learners' oral narrative performance and anxiety. *Educational Assessment*, *24*(2), 135-154. https://doi.org/10.1080/10627197.2019.1578169

Galikyan, I., Madyarov, I., & Gasparyan, R. (2019). Student test takers' and teachers' perceptions of the TOEFL junior® standard test. *ETS Research Report Series*, *2019*(1), 1-15. https://doi.org/10.1002/ets2.12264

García, X. C. (2021). *Perceptions of EFL teachers on how to teach English in a public school during the quarantine for Covid 19*. [Master's thesis, Universidad Pedagógica Nacional]/ http://hdl.handle.net/20.500.12209/16521

Green, A. (2014). *Exploring language assessment and testing language in action*. Routledge.

Gutiérrez Duque, D. M., y Mayora Pernía, C. A. (enero-abril, 2021). Variables predictoras del desempeño escolar en exámenes estandarizados de inglés: Evidencias desde el examen de Estado en Colombia. *Revista Virtual Universidad Católica del Norte*, (62), 33-62. https://www.doi.org/10.35575/rvucn.n62a3

Katz, S. (2021). Rapport in the foreign language classroom: from face-to-face to online in times of pandemic. Íkala, *Revista de Lenguaje y Cultura*, *26*(3), 485-511. https://doi.org/10.17533/udea.ikala/v26n3a02
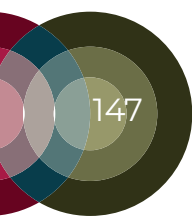
Krippendorff, K. (2013). *Content analysis: an introduction to its methodology* (3rd ed.). Sage.

Mahyoob, M. (2020). Challenges of e-learning during the COVID-19 pandemic experienced by EFL learners. *Arab World English Journal*, *11* (4), 351-362. https://dx.doi.org/10.24093/awej/vol11no4.23

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, *13*(3), 241256. https://doi.org/10.1177/026553229601300302

Pérez, A. P. (2021). *Aprendizaje del inglés en medio de la pandemia del COVID-19: Experiencias y emociones percibidas en estudiantes del nivel intermedio del Centro de Idiomas de la Universidad Peruana Unión, Lima-2020. Un enfoque mixto*. [Trabajo de grado]. Universidad Peruana Unión. https://repositorio.upeu.edu.pe/handle/20.500.12840/4516

Phakiti, A., & Isaacs, T. (2021). Classroom assessment and validity: Psychometric and edumetric approaches. *The European Journal of Applied Linguistics and TEFL 10*(1): 3–24. https://bd.univalle.edu.co/scholarly-journals/classroom-assessment-validity-psychometric/docview/2516291021/se-2

Pratolo, B. W., & Zahruni, N. A. (2020). Dynamic assessment effect on speaking performance of Indonesian EFL learners. *International Journal of Evaluation and Research in Education*, *9*(3), 778-790. http://doi.org/10.11591/ijere.v9i3.20466

Putra, A. S. (2017). The correlation between motivation and speaking ability. *Channing: Journal of English Language Education and Literature*, *2*(1), 36-57. https://doi.org/10.30599/channing.v2i1.87

Salehpour, G., & Roohani, A. (2020). Relationship between intrinsic/extrinsic motivation and L2 speaking skill among Iranian male and female EFL learners. *Bellaterra Journal of Teaching & Learning Language & Literature*, *13*(1), 43-59. https://doi.org/10.5565/rev/jtl3.803

Salgado G., E. (2015). Diálogo y aprendizaje percibido en estudiantes de modalidad virtual: Abordaje cualitativo en un programa universitario en Costa Rica. *Revista Iberoamericana de Educación a Distancia*, *18* (1), 191-211. https://n2t.net/ark:/13683/ptfg/efg

Sandvik, L. V., Svendsen, B., Strømme, A., Smith, K., Aasmundstad Sommervold, O., & Aarønes Angvik, S. (2023). Assessment during COVID-19: Students and teachers in limbo when the classroom disappeared. *Educational Assessment*, *28*(1), 11-26. https://doi.org/10.1080/10627197.2022.2122953

Sanz, G., Gil, A., Marnet, B., Berlanga, L., Tordesillas, M., Pérez Ruiz, S., & Gómez, S. (2020, October). Students' performance in French subjects in the COVID-19 confinement. In *Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 853-858). https://doi.org/10.1145/3434780.3436690

Tao, J., & Gao, X. (2022). Teaching and Learning Languages Online: Challenges and Responses. *System*, *107*, 102819. https://doi.org/10.1016/j.system.2022.102819

Valdivieso, M., Burbano, V. y Burbano, A. (2020). Percepción de estudiantes universitarios colombianos sobre el efecto del confinamiento por el coronavirus, y su rendimiento académico. *Espacios 41*(42), 269-281. https://revistaespacios.com/a20v41n42/a20v41n42p23.pdf
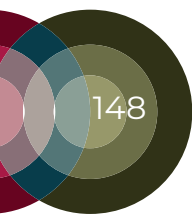
Vega-Carrero, S., Pulido, M., & Ruiz-Gallego, N. E. (2017). Teaching English as a second language at a university in Colombia that uses virtual environments: a case study. *Revista Electrónica Educare*, *21*(3), 1-21. https://doi.org/10.15359/ree.21-3.9

Weir, C. J. (2005). *Language testing and validation: an evidence-based approach*. Palgrave Macmillan.

Yang, Y., Gao, Z., & Han, Y. (2021). Exploring Chinese EFL learners' achievement emotions and their antecedents in an online English learning environment. *Frontiers in Psychology*, *12*, 722622 https://doi.org/10.3389/fpsyg.2021.722622