Computer-Assisted Language Learning Electronic Journal (CALL-EJ), 25(1), 1-22. 2024

Investigating the Impact of Transparent Language Online on L2 Listening Outcomes

Muhammad Asif* (muhammad-asif@umt.edu.pk)
Department of English and Literary Studies, School of Liberal Arts,
University of Management and Technology, Lahore, Pakistan
https://orcid.org/0000-0003-0076-8717
*Corresponding author

Muhammad Sheeraz (m.sheeraz@iiu.edu.pk)
Department of English,
International Islamic University, Islamabad, Pakistan

Abstract

This study investigates the impact of Transparent Language Online (TLO) on L2 listening outcomes at the tertiary level. The participants (N = 1,000) were L2 learners who had diverse cultural, linguistic and technological backgrounds and displayed major disparities in their English language skills. They had minimal to no prior proficiency in English listening skills. Cambridge First Certificate in English (FCE) listening tests were used to collect data. The data gathered were statistically analyzed to address the research questions and examine statistical hypotheses. Using TLO, the learners reached the Common European Framework of Reference for Languages (CEFR) levels B1 intermediate or B2 upper intermediate in listening. No other skills were examined. The findings indicate that Transparent Language can be an effective Web-delivered platform for improving L2 listening.

Keywords: Listening proficiency; second language; technology integration; Transparent Language Online; Web-delivered language learning platform

Introduction

Listening, a challenging skill for second language (L2) learners and educators, plays a critical part in an individual's day-to-day life (Bozorgian et al., 2021; Cross & Vandergrift, 2018; Rost, 2013). It is a core part of L2 learning (Maftoon & Fakhri Alamdari, 2020; Vandergrift, 2007), a primary component of communicative competence (Richards, 2008) and an essential channel for language input and acquisition (Krashen, 1992). Vandergrift states that listening is the "least explicit of the four language skills, making it the most difficult to learn" (2004, p. 4). It occupies a greater part of an individual's time in communicative exercises and activities compared to reading, writing and speaking. Listening comprehension might appear quite straightforward to L1 speakers; however, it is a source of frustration and chaos for L2 learners.

Over the past decade, technological advancements in streaming services, social platforms and commercial media websites have transformed L2 listening materials and methodologies (Godwin-Jones, 2021; Hsieh, 2020; Hubbard, 2017). Although recognized as valuable resources by second language acquisition (SLA) and computer-assisted language learning (CALL) experts, these innovations remain underutilized in L2 listening research, particularly in real classroom settings at the tertiary level (Cárdenas-Claros, 2020). The identified research gap results from the insufficient attention given to L2 listening research in real classroom settings, particularly at the tertiary level. This persistence arises from a scarcity of empirical studies and the procedural complexity of L2 listening (Dalman & Plonsky, 2022; Lynch, 2011). This seems to be a major challenge for L2 educators and learners alike, as they may experience disappointment, unsatisfactory academic performance or lack of consideration in the L2 classroom.

Addressing this gap, instructional technology tools and Web-delivered language platforms emerge as potential solutions to expedite L2 listening both inside and outside the classroom (Asif et al., 2022). These technologies, deeply integral to our contemporary world, can change how learners connect, correspond, learn and are instructed in L2. Despite the massive popularity of commercial online language platforms, having millions of active users (Kukulska-Hulme et al., 2017; Loewen et al., 2019), their efficacy for L2 learning has received little consideration from researchers (Jiang et al., 2021; Loewen et al., 2020). One such Web-delivered language learning platform, which has been completely neglected and marginalized by researchers, is Transparent Language Online (TLO). Therefore, it is vital to examine TLO's efficacy for improving L2 listening.

The current research used the TLO platform as a supplementary means to train L2 learners in advanced technology tools, improve their L2 listening proficiency, and help them become self-directed learners. The results validate the efficacy of this platform, not only addressing the identified gap in L2 listening research but also providing profound insights into the potential of TLO for language development.

Literature review

This literature review explores the intersection between technology and language learning, specifically focusing on the synergy of blended learning and declaratively accelerated blended learning. It also explicates Ullman's declarative and procedural model (2001), reviews the role and significance of Web-delivered language learning platforms and prepares the ground for Transparent Language Online.

Blended learning and declaratively accelerated blended learning

Blended learning, a blend of real-time in-person classroom learning (human pedagogy) with the utilization of Web-based (online) resources, is the finest and most practical technique to teach and learn foreign languages (Al-Obaydi, 2021; Asif et al., 2022; Asif & Aslam, 2023; Dziuban et al., 2018; Madden et al., 2019; McCarthy, 2016). Researchers state that blended learning is better than either human (face-to-face) instruction or technology-oriented learning alone (Madden et al., 2019; Rivera, 2016). Madden et al. (2019) advise that "blended learning itself is a kind of bridge between traditional in-person learning and the modern movement of online education. It seeks to integrate the benefits of the two: the face-to-face (F2F) interaction and the

personal advising and mentoring of the former, with the scale, asynchrony and flexibility of the latter" (p. xii).

Declaratively accelerated blended learning (DABL), a specialized subtype of blended learning, is an ideal mix of technology and human instruction designed to accelerate declarative memory (Allen, 2016; Patchan et al., 2016). Grounded in the cardinal principles of human learning, DABL focuses on L1 and L2 acquisition by emphasizing meaningful, interactive and responsive contexts (Konishi et al. 2014). The most helpful theory underpinning DABL is Ullman's (2001) model, which asserts the separation of declarative and procedural elements in language learning. It posits the mental lexicon's connection to the declarative system and specific brain structures in temporal regions, while associating the mental grammar with the procedural system and structures in the basal ganglia and frontal cortex. The model's predictions on the neurocognitive basis of lexicon and grammar find support in various evidence, including psycholinguistic studies, analyses of developmental language disorders, neurological cases, and neurophysiological observations. Furthermore, DABL experience goes beyond theory, integrating technology-delivered activities to pre-load the "declarative" aspects of a lesson (grammar, culture and context), followed by live instruction focusing on "procedural" skill-building (tasks, role play and L2 conversation) using the language of the lesson (Ullman, 2001). This approach, blending theory, technology and human instruction, defines the essence of DABL.

To distinguish between blended learning and declaratively accelerated blended learning (DABL), it is essential to understand their foundational principles. While blended learning broadly integrates in-person and technology-based instruction (Madden et al., 2019; Rivera, 2016), DABL specifically targets and accelerates declarative memory, a critical element in language acquisition (Asif et al., 2022). Unlike the broad scope of blended learning, DABL's distinctiveness lies in its strategic focus on memory development. Moreover, DABL's adaptability allows it to be easily tailored to accommodate various curricula, student needs and environments (Conrad & Donaldson, 2012). It can deliver Web-delivered lessons anytime, anywhere, on any device and incorporates skill-building guidance that efficiently leverages declarative skills (Konishi et al. 2014; Kukulska-Hulme et al., 2017; Reinders & Benson, 2017).

In short, the integration of blended learning methodologies with the specialized framework of DABL presents a compelling and flexible approach in foreign language education. This strategic synthesis effectively combines technology-enhanced modalities with traditional pedagogical elements. The integration of both declarative and procedural components, supported by Web-delivered platforms such as Transparent Language Online (TLO), establishes a trajectory for accelerated language proficiency. This research-driven approach aligns with learners' expectations for an efficient and effective language education experience, demonstrating a logical and empirically grounded framework for language acquisition and skill development.

Technology for teaching the declarative element of language

Declarative acceleration leverages technology to speedily present words, phrases and clauses using a suite of connecting and engaging sequences, games and learning activities (Konishi et al., 2014; Means et al., 2014). Because of its adaptability, personalization and speed, this technology more effectively instructs the declarative language element (Bolgün & McCaw, 2019; Radianti et al., 2020). Asif et al. (2022) state that a computer can illustrate numerous learning experiences in real time, detect, record and organize results, and consistently alter them until the learner stores a task/ assignment in their declarative memory (p. 279). Modern functional magnetic resonance imaging (fMRI) research, particularly Ullman's work, identifies specific memory

structures—declarative and procedural memory systems—in the human brain contributing to language learning (Ullman, 2001). Declarative memory retains factual information, while procedural memory learns skills and extends to the automatic and unconscious application of acquired knowledge (Pili-Moss et al., 2020; Suzuki et al., 2023).

Consequently, learners are closely associated with language learning through personal computers, laptops, the Web and mobile devices, thus eliminating spatio-temporal limitations (Du et al., 2022; Ko, 2017; Şad et al., 2020; Vadivel et al., 2022). An educator would be hard-pressed to reproduce these flexible, rapid, efficient and customized learning practices in a traditional classroom environment. Hence, via Web-delivered language learning platforms such as TLO, learners experience remarkable acceleration in language assimilation into declarative memory. The platform rapidly acquaints students with words and phrases using instructive games and varied, engaging activities about reading, writing, listening and speaking that start from the first lesson and proceed until the last. It has become a comprehensive solution that contrasts with other learning systems and applications and explicitly matches the requirements of students.

Effectiveness of commercial Web-delivered language learning platforms

Web-delivered language learning platforms, which frequently include learning applications within their systems, play a significant role in training language learners worldwide. Sacco, as cited in Vesselinov et al. (2019), says that the market for these suppliers is colossal and growing exponentially across the globe. Key providers include Transparent Language Online, Rosetta Stone, Duolingo, Busuu, and Babbel, with competitors such as Mango Languages, Berlitz, WeSpeke, Fluent Forever, Fluenz, Linguistica360, Rocket Languages, and Yabla, among others. These platforms provide a customized learning environment, allow constant input and increase the number of learning procedures due to their versatile and diverse configurations and learning formats. Despite their widespread popularity and millions of active users (Kukulska-Hulme et al., 2017; Loewen et al., 2020), there is a paucity of research on the efficacy of these platforms for second language (L2) acquisition (Jiang et al., 2020; Loewen et al., 2019; Loewen et al., 2020; Rachels & Rockinson-Szapkiw, 2018; Rosell-Aguilar, 2018).

The competitive nature of the industry, coupled with the commercial interests of these providers, has resulted in a scarcity of empirical data supporting the ambitious claims of rapid proficiency advancement made in their promotional materials (Loewen et al., 2020). Critics, including Robert DeKeyser, have expressed skepticism about the continual promises of applications and devices as ultimate solutions to learners' difficulties (Pearl, 2017). There have been various requests for research on the effectiveness of these platforms for L2 learners (Loewen et al., 2019; Loewen et al., 2020; Plonsky & Ziegler, 2016; Smith, 2017). Despite the extensive claims made by platforms like Babbel and Fluent Forever regarding rapid language acquisition, the lack of transparent definitions for terms like "fluency" and "in a short time" raises questions about the veracity of their assertions (Vesselinov et al., 2019). Claims about reaching specific proficiency levels within certain time frames lack empirical support, necessitating a thorough investigation into the effectiveness of these platforms (Jiang et al., 2021; Loewen et al., 2020).

Existing studies primarily focus on the time required to reach proficiency levels equivalent to conventional basic language courses, particularly in Spanish. Studies on commercial learning platforms, e.g., Babbel (Vesselinov and Grego, 2016), Busuu (Vesselinov & Grego, 2016), and Duolingo (Jiang et al., 2020; Krashen, 2014; Loewen et al., 2019; Pearl, 2017; Ratzlaff, 2015; Vesselinov and Grego, 2012), Rosetta Stone (Lord, 2015, 2016; Vesselinov et al., 2019), have shown notable improvements in L2 learners' grammar, vocabulary and receptive skills after some

usage of available online language platforms. However, researchers have raised concerns about the platforms' claims regarding improving writing and speaking skills (van Deusen-Scholl, 2015; Lord, 2016; Rodrigues & Vethamani, 2015).

Transparent Language: Theoretical grounding

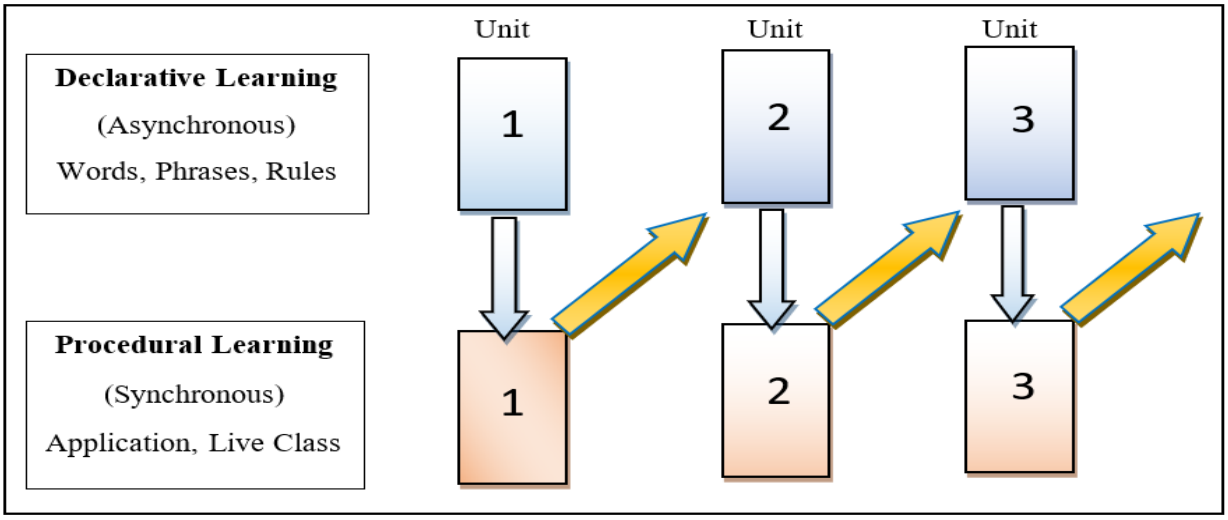
Transparent Language (TL) has undergone a transformative journey from its origins as LanguageNow, LanguagePro and 101 Languages during the CD/DVD era. Originally focused on Comprehensible Input, wherein the cursor's placement triggered dynamic displays of word meanings, sentence structures, grammatical descriptions and pronunciations, TL's innovation gained recognition from Krashen. Comprehensible Input (CI) works—just as almost any active engagement with another language works—but it was slow and uncertain. As TL's focus turned toward instantiating, maintaining and enhancing reliable language skills with the least amount of invested time, CI and many other methods that built proficiency came up short. However, recognizing the evolving dynamics of language learning and the need for more efficient approaches, TL has since transcended Krashen's framework.

To fortify the theoretical foundation underpinning TL's effectiveness, the platform embraced Ullman's declarative and procedural memory model (2001). In contrast to Krashen's broader conceptualization, Ullman's model offered a more specific framework that resonated with TL's exploration of effective technology-enabled activities. Declaratively accelerated blended learning emerged as a pivotal concept within TL, representing a strategic shift towards a more comprehensive and adaptive language learning model. This innovative approach, which integrates technology with human instruction, underscores TLO's commitment to strategically supporting language instructors without aiming to replace them (Asif et al., 2022). The emphasis on combining pre-loaded declarative aspects with live procedural skill-building acknowledges the distinctive contributions of human interaction in the language learning process. Instructors play a pivotal role in optimizing the interweaving of these components, ensuring efficiency, flexibility, sustainability and comprehensive tracking and reporting of learners' progress.

DABL surpasses the capacities of live instructors in specific domains. Tech-delivered activities in TL, presenting "chunks" of lexicon, grammar and culture, facilitate focused learner action, repetition and ongoing reinforcement. While proficient instructors excel in guiding learners to integrate memorized chunks into skills, TL's systematic approach ensures precision and active memory retention over time. The integration of synchronous discussions, role-plays and in-language tasks (in a traditional setup) with asynchronous preparations utilizing the Web, portable or computer capabilities (setup anywhere, on any device, at any time) fosters a versatile and accessible learning environment.

The term "DABL" distinguishes TL's approach from other concepts like the "flipped classroom". Flipped means many things, but DABL specifically means deliberately driving tech-delivered declarative mastery at a strong pace and then interweaving that with aligned live-delivered procedural skill-building. This alignment ensures that both tech-delivered and live-delivered time are tightly focused on the same lesson, enhancing the effectiveness of the learning experience.

Figure 1
Declarative and procedural learning (Transparent Language, 2013)



DABL can be applied to any curriculum or program of instruction. If instruction is not available, the tech lessons also optimize independent learning. However, optimized independent learning is not as effective as optimized DABL learning. Transparent Language (TL) has developed a shared special-purpose digital platform for the authoring, dissemination and use of language materials: it is called Transparent Language Online (TLO) in commercial spaces and the US government offering is called the CL-150 platform.

Transparent Language Online (TLO) is a mastery-based, self-paced, controllable and interactive learning platform that offers instruction in over 115 world languages. Trusted by the US government, the US military, corporations, the medical community and over 12,000 schools and institutions worldwide, TLO takes a meticulous approach to address learners' shortcomings. The platform presents numerous learning encounters per minute, observes and records learning outcomes, and continually adjusts until learners commit material to declarative memory (Asif et al., 2022).

TLO's versatility is evident as it requires no installation and can be accessed via multiple devices, including smartphones and tablets. This platform allows instructors to prepare interactive listening, reading and vocabulary lessons; organize and manage classes through a portal; assign, track and report learners' activities; and design dialogue-oriented courses. In addition, it offers students various learning resources, including educational games, diverse listening activities, and automated tests to enable them to assess their command of L2 and system accessibility 24/7, so they can use this platform anywhere at a time that suits them. After completing a unit of study, students take a self-assessment test on TLO to determine whether they need to spend more time in that unit or are ready to move on to the next one. This iterative process ensures that the learning experience remains adaptive, tailored to individual needs and aligns with the principles of the DABL framework.

Research gap

The literature lacks specific insights into the effectiveness of declaratively accelerated blended learning (DABL) and Transparent Language Online (TLO) in developing L2 listening

proficiency, particularly in tertiary education. While existing studies often generalize the efficacy of language learning platforms, they lack a specific focus on these aspects. Therefore, there is a pressing need for a more targeted investigation to uncover precise insights into how DABL and TLO impact L2 listening outcomes.

Research questions

This investigation addresses the following research questions and hypotheses:

RQ1. How and to what extent does the integration of TLO with pedagogy impact on L2 listening outcomes at the tertiary level?
RQ2. To what extent is there a difference in the listening proficiency scores of L2 learners exposed to TLO instruction and those in a traditional classroom environment? Is this difference statistically significant?
RQ3. Is there any significant difference in L2 listening proficiency results when two sub-groups of TLO, i.e., U1CS and U2CS, are compared?

Hypotheses:

RQ1
$\mathbf{H_0}$. There is no significant impact of TLO on L2 listening outcomes at the tertiary level.
$\mathbf{H_a}$. There is a significant impact of TLO on L2 listening outcomes at the tertiary level.

RQ2
$\mathbf{H_0}$. There are no significant differences in the listening proficiency scores of L2 learners exposed to TLO instruction and those in a traditional classroom environment.
$\mathbf{H_a}$. There are significant differences in the listening proficiency scores of L2 learners exposed to TLO instruction and those in a traditional classroom environment.

RQ3
$\mathbf{H_0}$. There are no significant differences in L2 listening proficiency results when two sub-groups of TLO, i.e., U1CS and U2CS, are compared.
$\mathbf{H_a}$. There are significant differences in L2 listening proficiency results when two sub-groups of TLO, i.e., U1CS and U2CS, are compared.

Method

Participants

The participants of this study were university undergraduate English language learners (ELLs), $N = 1{,}000$, residing in Pakistan, who studied Functional English as their L2 in their first semester in the faculties of computer science (CS) and business studies (BS). They were equally distributed into two major groups, i.e., experimental ($n = 500$) and control ($n = 500$), and further sub-divided into four sub-classifications, i.e., U1CS ($n = 250$), U2CS ($n = 250$), U1BS ($n = 250$) and U2BS ($n = 250$). The experimental group comprised U1CS and U2CS, while the control group included U1BS and U2BS. Notably, participants self-reported minimal or no prior proficiency in

English listening skills. The experimental group received instruction through TLO, while the control group underwent traditional classroom instruction. This investigation used a quantitative experimental research method and a convenience sampling technique based on major and class/ university to select the participants (Riazi, 2016) and G*Power version 3.0.10 to determine the sample size (Faul et al., 2009).

Time on task

To address the potential discrepancies in 'time on task' between groups, particularly with the TLO platform, the study implemented monitoring mechanisms. The researchers recorded the time spent by each participant using TLO to ensure fairness and uniform engagement across both groups. This approach aimed to minimize the impact of varied time investments, engaging a more accurate assessment of the platform's efficacy.

Instruments and procedures

Cambridge First Certificate in English (FCE) listening standardized tests with Cronbach's alpha of .85 were used as the primary data collection instrument. The assessment items numbered 30, and the allocated time was 40 minutes for each one. These tests measured the listening proficiency of all students. The researchers administered four FCE listening tests, i.e., a pre-test, test 1, test 2, and a post-test, and gathered 4,000 observations from $N = 1{,}000$ students at university 1 (U1) and university 2 (U2). The local subsystem of the Learning Management System (LMS), developed by in-house designers based on the Moodle View Controller (MVC) design at the Department of Computer Science at University 1 (U1CS), facilitated test administration (Asif et al., 2022, p. 274). The pre-test was administered in week 1, tests 1 and 2 in weeks 6 and 11, respectively, and a post-test in week 16. Computer-based FCE listening assessments were simultaneously administered to the experimental group, i.e., U1CS and U2CS, and paper-based to the control group, i.e., U1BS and U2BS. Both groups undertook the same standardized tests in two different modes. This decision was made to ensure consistency within each group and control for potential biases introduced by differing test formats. The LMS subsystem performed automated checking on the experimental group, whereas the investigators manually checked control group assessments. Each correct answer was awarded one mark, each incorrect answer scored zero. Cambridge CEFR was used as an assessment benchmark to assess L2 language proficiency in listening skills.

Data collection

The data collection process commenced when the investigators received approval from the authorities at U1 and U2. The data collection process continued throughout a 16-week semester, during which time participants attempted four listening tests. To thank the students for their worthy contribution to this research project, the Deans/ Directors of the four faculties rewarded students for their active participation. They were rewarded for their efforts while attempting different assessments as graded tasks. After students attempted the tests, the LMS teams shared the listening data with the investigators and deleted them from their local system. The data gathered were organized in Microsoft Excel 2019 and converted to IBM SPSS 25.0 files. Responses to the question items were graded and coded based on different variables, such as age group, class,

section, faculty and university, followed by normality tests, and subjected to several statistical apparatuses to address the research questions and trial statistical hypotheses.

Data analysis

Descriptive statistics were calculated to answer the research questions and conduct statistical analysis in two ways: (1) frequency examination in percentage form and (2) mean, median, standard deviation and confidence intervals using IBM SPSS 25.0. The researchers used the codebook for descriptive statistics to identify missing values and remove unfinished assessments, test items and participants who did not participate in any test or venture stage. After calculating and examining descriptive statistics, the investigators applied inferential statistical tests to the data gathered to acknowledge or dismiss hypotheses H_0 and H_a.

Results

This section reports the results in a question-by-question format. Research question (RQ1) addresses how and to what extent integrating TLO with pedagogy impacts on L2 listening outcomes at the tertiary level. This question focuses on the sub-classifications of the experimental and control groups, i.e., U1CS, U1BS, U2CS and U2BS. RQ2 concentrates on reporting significant differences in the proficiency scores of L2 learners exposed to TLO instruction (experimental group) and those in a traditional classroom environment (control group). This allows the investigators to compare the scores of the two groups and highlights whether TLO delivers better results in terms of students' learning than traditional modes of learning. Last, RQ3 explores any significant differences in L2 listening proficiency results when two sub-groups of TLO, i.e., U1CS and U2CS, are compared. The study also focuses on several variables: mode of instruction, test type, department, university and result. Result is the only dependent continuous variable; the rest are independent categorical variables. Each variable focused on works out differently in different situations and how they influence L2 listening is investigated. The researchers gathered 4,000 observations (listening data) from $N = 1000$ participants, concentrating on three RQs. They applied descriptive statistics and the most appropriate inferential statistical analysis to the data collected to answer these RQs and assess statistical hypotheses. Descriptive statistics is a significant apparatus for quantitatively describing and summarizing data's key features (Baffoe-Djan & Smith, 2019).

Impact of TLO on L2 listening outcomes (RQ1)

These descriptive data, obtained from listening, based on the variables 'result' for 'mode of instruction' and the grouping variable 'test type', answer RQ1. Table 1 shows the *mean* of result by mode of instruction. The table also presents an exploration of the experimental and control groups at U1 and U2 among the four listening assessments.

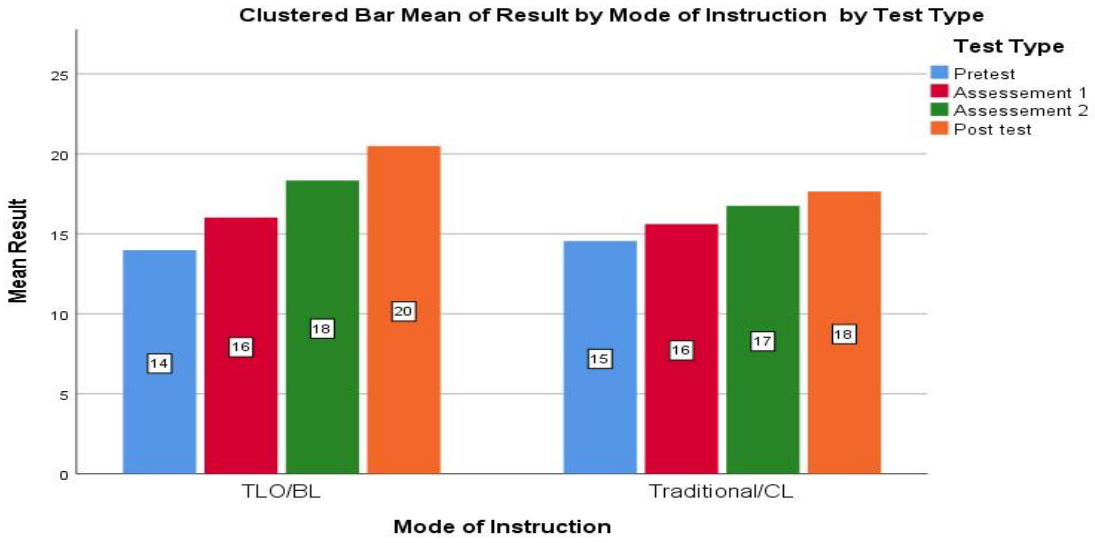
Table 1

Descriptive statistics: Mean of result by mode of instruction

Mode of Instruction	Mean	95% CIs for Mean: Lower Bound	95% CIs for Mean: Upper Bound	Median	Variance	Std. Deviation	Skewness	Kurtosis
TLO	17.20	17.02	17.39	17.00	17.732	4.211	.047	-.234
Traditional	16.14	15.99	16.29	16.00	12.109	3.480	-.035	-.094

As shown in Figure 2, the bar chart presents a clear comparison between the two groups.

Figure 2
Descriptive statistics: Mean of result by mode of instruction and test type



The investigators applied ANOVA test to the variable 'mode of instruction' to answer RQ1, check its hypotheses, and validate the results gained from descriptive statistics. As shown in Table 2, ANOVA shows that examining two variables, 'result' and 'mode of instruction', generated substantial outcomes following an F-Distribution, with an F of 75.447 at p-value = 0.05.

Table 2
Analysis

	Sum of Squares	df	Mean Square	F	Sig.
Between groups	1,125.721	1	1,125.721	75.447	.000
Within groups	59,652.630	3,998	14.921		
Total	60,778.351	3,999			

Now the question may arise whether variances were assumed to be equal. To answer this query, the researchers used a test of homogeneity of variance to check this assumption. As shown in Table 3, Levene Statistic shows that the variances for the variable 'mode of instruction' were not equal; nonetheless, they showed highly significant results.

Table 3
Result of test of homogeneity of variance

	Levene Statistic	df1	df2	Sig.
Based on mean	76.216	1	3,998	.000
Based on median	74.516	1	3,998	.000
Based on median and with adjusted df	74.516	1	3,897.141	.000
Based on trimmed mean	75.811	1	3,998	.000

Since the test of homogeneity of variance confirmed that the experimental and control groups had unequal variances and they violated the strict assumption of classical ANOVA, the researchers applied a Welch-ANOVA test due to its insensitivity to unequal variances. As shown in Table 4, Welch-ANOVA generates notable results, with F statistic having a value of 75.447**.

Table 4
Welch-ANOVA: Robust tests of equality of means

Result	Statistic[a]	df1	df2	Sig.
Welch	75.447	1	3860.919	.000

Asymptotically F distributed.

To the researchers' surprise, Welch-ANOVA produced the same result, F value = 75.447, as classical ANOVA.

Thus, the findings for descriptive statistics, simple ANOVA, a test of homogeneity of variance and Welch-ANOVA support hypothesis H_a for RQ1, validating a significant impact of TLO on L2 listening outcomes at the tertiary level.

Differences in the listening proficiency scores of l2 learners (RQ2)

Descriptive statistics focus on the variable 'result' for 'test type', taking 'department' as a grouping variable, answer RQ2, and examine statistical hypotheses. Table 5 presents the *mean* of result by test type. The table presents an overview of descriptive statistics for L2 learners who showed continuous improvement in their listening outcomes when the researchers compared the *means* from pre-test to post-test. For instance, mean score at the pre-test level was 14.27, the highest point being 19.07, with a slight increase in standard deviation from 3.378 to 3.469.

Table 5

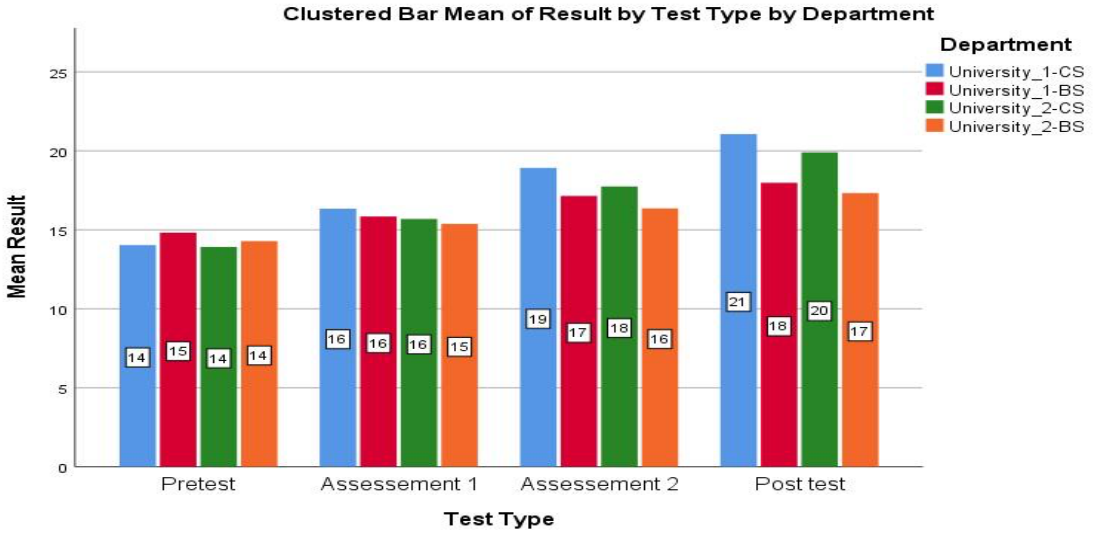
Descriptive statistics: Mean of result by test type

Test Type	Mean	95% CIs for Mean		Median	Variance	Std. Deviation	Skewness	Kurtosis
		Lower Bound	Upper Bound					
Pre-test	14.27	14.06	14.47	14.00	11.412	3.378	.171	.021
Assessment 1	15.81	15.59	16.03	16.00	12.495	3.535	.082	.029
Assessment 2	17.54	17.33	17.76	17.00	11.862	3.444	.160	.053
Post-test	19.07	18.85	19.28	19.00	12.031	3.469	.065	.183

The graphical representation of results obtained in Table 5 illustrates that L2 learners in the experimental group (U1CS and U2CS) demonstrated consistently improved results and higher listening outcomes compared to learners in the control group (U1BS and U2BS). Figure 3 further validates these findings.

Figure 3

Descriptive statistics: Mean of result by test type and department



Regression analysis was the most appropriate tool for researchers to apply at this stage, taking 'test type' as an independent variable and 'result' as a response variable. The RQ2 investigation concerns the statistically significant relationship between the two understudy variables for the experimental and control groups. There were four different test types for which *mean* scores differed from each other. The regression analysis (see Table 6) produced highly significant results to support this assumption.

Table 6

Regression analysis: ANOVA[a]

Model		Sum of squares	df	Mean square	F	Sig.
1	Regression	13,018.525	1	13,018.525	1,089.788	.000[b]
	Residual	47,759.826	3,998	11.946		
	Total	60,778.351	3,999			

Dependent variable: Result
Predictors: (Constant), test type

It was also observed that not only was the regression model significant, but it also proved that the coefficients produced by this model are very significant under students' t-distribution, as shown in Table 7.

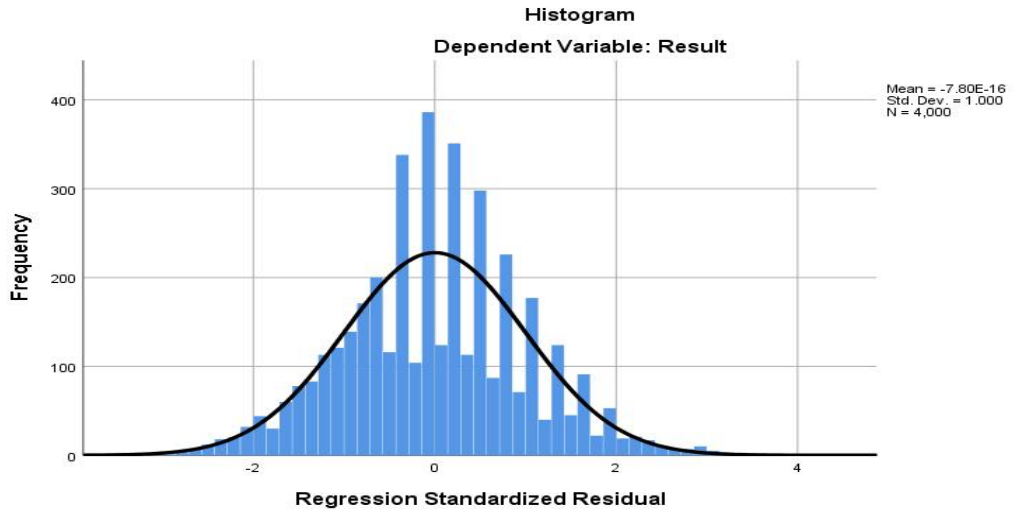
Table 7
Regression analysis: Coefficients[a]

Model		Non-standardized coefficients		Standardized coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	12.637	.134		94.407	.000
	Test Type	1.614	.049	.463	33.012	.000

Dependent variable: Result

Figure 4 presents the standardized regression residual and a histogram for the dependent variable 'result'. The histogram shows that the regression standardized residual satisfies the normality assumption.

Figure 4
Regression standardized residual



Considering the findings of the stated tests, the study accepts hypothesis H_a for RQ2, proving there are significant mean differences in the listening proficiency scores of L2 learners exposed to TLO instruction.

Investigating L2 listening proficiency outcomes between U1CS and U2CS (RQ3)

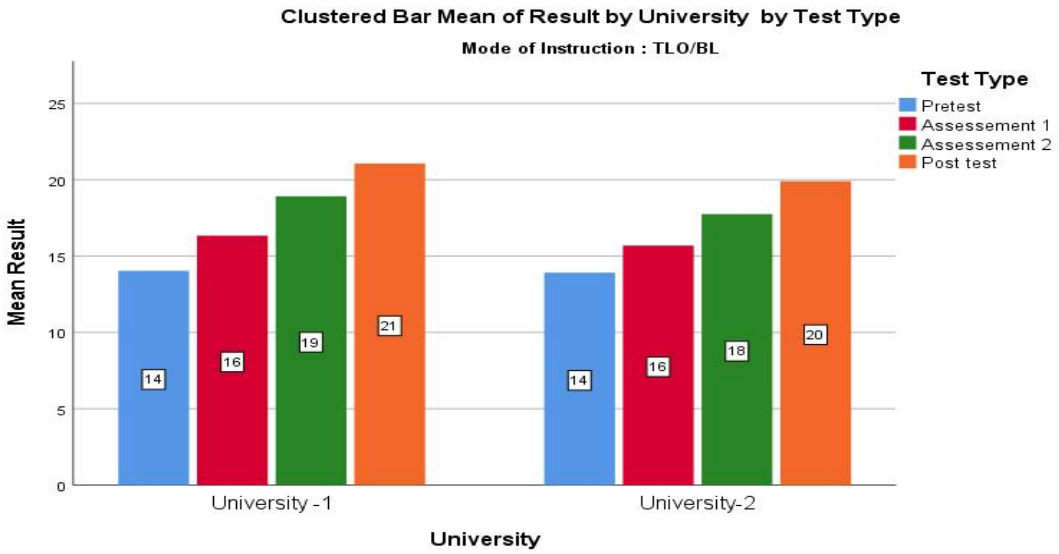
The most common descriptive measures for the *mean* of 'result' of two groups taking 'test type' as a grouping variable were obtained to answer RQ3. Two sub-groups of the experimental group (U1CS and U2CS) were selected. Table 8 shows that U1CS generated higher *mean* scores than U2CS, despite using the same mode of instruction, TLO, but in a different setup. There, the researchers observed another significant descriptive, i.e., 'median'. U1CS produced a median score of 18, while for U2CS, it was 17.

Table 8
Descriptive statistics: Mean of result by university by test type

University	Mean	95% CIs for Mean: Lower bound	95% CIs for Mean: Upper bound	Median	Variance	Std. deviation	Skewness	Kurtosis
U1 (CS)	17.59	17.29	17.89	18.00	23.724	4.871	-.022	-.559
U2 (CS)	16.81	16.60	17.02	17.00	11.457	3.385	-.120	-.463

Figure 5 supports Table 8 and proves that U1CS shows better listening outcomes than U2CS when the means of both groups from pre-test to post-test were compared.

Figure 5
Descriptive statistics: Mean of result by university and test type



RQ3 examines the *mean* difference between U1CS and U2CS. The researchers applied an independent sample t-test to investigate the difference in *mean* and check the equality of variance. A Levene's test, see Table 9, shows that the variables had equal variances. This test presents significant results, and an independent sample t-test gives substantial results.

Table 9
Independent samples test

		Levene's test for equality of variances		t-test for equality of means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean difference	Std. error difference	95% Confidence interval of the difference	
									Lower	Upper
Result	Equal variances assumed	134.544	.000	4.137	1998	.000	.776	.188	.408	1.144
	Equal variances not assumed			4.137	1781.418	.000	.776	.188	.408	1.144

The researchers used Pearson's correlation between U1CS and U2CS to confirm the findings of the independent sample t-test. Readers may question the degree of strength and the level of interdependence between the two groups. The strength level between U1CS and U2CS is .377, as shown in Table 10, which proposes a 37.87% degree of strength, which is positive. When the results of one sub-group increase, the outcomes of another group will, in like manner, increase.

Table 10
Pearson correlations between the two sub-groups of TLO

		U1CS	U2CS
U1CS	Pearson correlation	1	.377**
	Sig. (2-tailed)		.000
	N	1000	1000
U2CS	Pearson correlation	.377**	1
	Sig. (2-tailed)	.000	
	N	1000	1000

**. Correlation is significant at the 0.01 level (2-tailed).

While evaluating the disparity between U1CS and U2CS, the researchers utilized Cohen's d test to measure the effect size of the observed mean difference. The descriptives analysis for U1CS and U2CS yielded a Cohen's d value of 0.04, indicating a notably small effect according to Cohen's established criteria:

- d=0.2 indicates a small effect,
- d=0.5 conveys a medium effect, and
- d=0.8 signifies a large effect.

This small effect size, indicative of minimal variability within the sample, is a crucial point of consideration. It is essential to recognize that while the t-test establishes statistical significance in the mean difference, the practical implications may be restricted due to the small effect size. This underscores the precision and reliability of the study but necessitates a thoughtful interpretation of the observed differences. Thus, the inclusion of Cohen's d test serves to enrich the insights gained from the t-test, providing a nuanced understanding of the extent of distinctions between U1CS and U2CS.

As a result, the independent samples t-test and Pearson correlations not only verify the findings of descriptive statistics but also validate hypothesis H_a, indicating a significant difference in L2 listening proficiency results when U1CS and U2CS are compared. U1CS delivers better listening outcomes than U2CS.

This is evident from the outcomes of the RQ1, RQ2, and RQ3 explorations, that L2 learners demonstrated outstanding listening proficiency results, improved academic accomplishments and increased confidence in using learning technologies.

Discussion

This study, which explores the impact of Transparent Language Online (TLO) on L2 listening outcomes, significantly contributes to the existing literature on declaratively accelerated blended learning (DABL) and TLO. The investigation involved a diverse group of participants (N = 1,000) with varying cultural, linguistic and technological backgrounds, displaying significant disparities in their English language skills. The participants reported minimal or no prior proficiency in English listening skills. The study utilized Cambridge First Certificate in English (FCE) listening tests to collect data, and the findings were statistically analyzed to address the research questions and corresponding hypotheses.

Thorough examination of the impact of TLO on L2 listening outcomes (RQ1) through descriptive and inferential statistics showed a significant influence on the participants. Differences in the listening proficiency scores of L2 learners (RQ2) and investigating L2 listening proficiency scores for U1CS and U2CS (RQ3) were analyzed using regression analysis, ANOVA, independent sample t-tests and Pearson correlations. Guided by the TLO platform, participants achieved Common European Framework of Reference for Languages (CEFR) levels B1 intermediate or B2 upper intermediate in listening, providing empirical evidence of TLO's effectiveness in improving L2 listening.

The theoretical foundation of this study, primarily grounded in Ullman's (2001) model, aligns with the principles of declaratively accelerated blended learning (DABL). This model highlights the strategic separation of declarative and procedural aspects in language learning. The alignment of this theoretical approach with DABL principles resonates with the broader framework of blended learning, which seeks to bridge the gap between traditional and online education, as emphasized by Madden et al. (2019). Focused on L2 listening proficiency, a vital component of communicative competence (Richards, 2008), the current investigation explores how TLO influences this specific linguistic domain.

Moreover, this research contributes to the existing literature by redirecting attention to TLO's efficacy in enhancing L2 listening outcomes. While other language learning platforms,

including Babbel (Vesselinov and Grego, 2016), Busuu (Vesselinov & Grego, 2016), Duolingo (Jiang et al., 2021; Krashen, 2014; Loewen et al., 2019; Pearl, 2017; Ratzlaff, 2015; Vesselinov and Grego, 2012), and Rosetta Stone (Lord, 2015, 2016; Vesselinov et al., 2019), have been studied, the unique attributes of TLO in improving listening proficiency were previously overlooked. This research adds empirical evidence to the ongoing discourse regarding the effectiveness of commercial language platforms.

Addressing observed disparities in results between groups, TLO's adaptability and versatility emerge as key factors. These differences are attributed to the dynamic nature of the platform, accommodating diverse learner needs. This also supports Madden et al.'s (2019) emphasis on the flexibility of blended learning in catering to various learning preferences and needs. This research goes beyond the mere identification of differences and offers a detailed analysis of the factors contributing to variations between these groups. This was in line with Asif et al. (2022), who stated that TLO's adaptability to varying curricular demands and the self-paced nature of the platform are identified as significant contributors to observed disparities. The insights provide actionable information for educators and platform developers to tailor instructional approaches effectively.

Furthermore, the pedagogical implications drawn from the empirical findings form a solid foundation for practical recommendations. The researchers also advocate for the integration of a self-paced autonomous listening course within TLO. This suggestion is intricately tied to the study's results, ensuring a logical extension of the findings. Learners are encouraged to adjust their expectations in accordance with the learning curve, emphasizing TLO's adaptability to individual learning paces.

As a result, the findings not only substantiate the overall positive impact of TLO on L2 listening outcomes but also provide specific insights into the achieved proficiency levels and factors contributing to varied results between groups. These insights enrich the discussion, offering a robust foundation for educators and researchers to build upon.

Conclusion

This study underscores the significant impact of Transparent Language Online (TLO) on L2 language pedagogy, particularly in enhancing learners' listening outcomes and confidence in using Web-based language learning platforms. Despite notable findings, this investigation does, without a doubt, have certain limitations that may impact L2 listening outcomes. Focused on university undergraduate English Language Learners (ELLs) studying Functional English, the research conducted four listening assessments in two different setups, i.e., computer-based and paper-based settings, in observed environments that may have affected L2 listening outcomes, as well as learners' confidence, because of linguistic, pedagogical, technical or technological constraints. Next, due to scarce human labor, inadequate resources, limited funds and spatio-temporal constraints, this research was delimited to listening skills and ignored other skills, i.e., reading, writing and speaking. Despite these limitations, this investigation may be viewed as helpful in facilitating a different approach to L2 language pedagogy and learning.

This study suggests some potential guidelines, directions and propositions for future researchers and practitioners regarding Web-delivered language learning platforms. A diversified participant background, including K-12, graduate and post-graduate levels, and exploring diverse domains, particularly for L2 learners, could enhance our understanding of the impact of online learning platforms. The expansion of the study to include both public and private-sector

institutions, along with the exploration of additional variables such as gender, age, sections, time points, and constructs, would contribute to a more comprehensive understanding of learners' engagement. Furthermore, the study recommends extending the evaluation of TLO beyond English to assess its efficacy in over 115 foreign languages, varying setups and different educational levels. This approach would enrich our understanding of TLO's effectiveness across diverse linguistic and educational contexts.

The effectiveness of TLO in enhancing learners' proficiency scores and confidence is evident. Learners became better acquainted with the TL system and foreign language by using different educational games, interactive and engaging learning activities and drills pertaining to integrated language skills from scratch until completion of a given task. TLO proved to be a comprehensive solution compared to other online language learning platforms. It explicitly synchronized with learners' requirements and assisted them by increasing their confidence in using a foreign and/or additional language. This research confirms the effectiveness of TLO for L2 listening outcomes at the tertiary level.

A rerun of the current research could deepen our comprehension of the impact of computerized broadcasts, e-assessments, Web-based glossaries and forums on different skills, such as reading, writing and speaking and different variables, including vocabulary and culture. The results affirm the objectives of the current project. Descriptive and inferential statistical tests produced highly significant results and verified and accepted hypothesis H_a for RQs 1–3, which substantiates that TLO positively impacted L2 listening outcomes and helped learners achieve higher proficiency scores. The efficient implementation of TLO in L2 pedagogy can increase learners' knowledge and impart confidence in using it in diverse circumstances.

References

Al-Obaydi, L. H. (2021). Humanistic learning elements in a blended learning environment: A study in an EFL teaching context. *Interactive Learning Environments*, *29*(2), 1–14. https://doi.org/10.1080/10494820.2021.1919717

Allen, M. W. (2016). *Michael Allen's guide to e-learning: Building interactive, fun and effective learning programs for any company*. John Wiley & Sons.

Asif, M., Sheeraz, M., & Sacco, S. J. (2022). Evaluating the impact of technological tools on the academic performance of English language learners at tertiary level: A pilot investigation. *Pegem Journal of Education and Instruction*, *12*(1), 272–282. https://doi.org/10.47750/pegegog.12.01.28

Asif, M., & Aslam, A. (2023). Exploring the significance of blended learning models for L2 educators in the developing world : A review of current literature. *Pakistan Languages and Humanities Review*, *7*(4), 135–147. https://doi.org/doi.org/10.47205/plhr.2023(7-IV)12

Baffoe-Djan, J. B., & Smith, S. A. (2019). Descriptive statistics in data analysis. In *The Routledge handbook of research methods in applied linguistics* (pp. 398–414). Routledge.

Bolgün, M. A., & McCaw, T. (2019). Toward a neuroscience-informed evaluation of language technology. *Computer Assisted Language Learning*, *32*(3), 294–321. https://doi.org/10.1080/09588221.2018.1516675

Bozorgian, H., Fallahpour, S., & Alinasab Amiri, M. (2022). Listening for young-adult EFL learners: Metacognitive intervention through L1. *International Journal of Listening*, *36*(3), 207-220. https://doi.org/10.1080/10904018.2021.1923499

Cárdenas-Claros, M. S. (2020). Spontaneous links between help option use and input features

that hinder second language listening comprehension. *System*, *93*, 102308.

Conrad, R.-M., & Donaldson, J. A. (2012). *Continuing to engage the online learner: More activities and resources for creative instruction*. John Wiley & Sons.

Cross, J., & Vandergrift, L. (2018). Metacognitive listening strategies. In *The TESOL encyclopedia of English language teaching*, Vol 8, (pp. 1–5). John Wiley & Sons. https://doi.org/10.1002/9781118784235.eelt0589

Dalman, M., & Plonsky, L. (2022). The effectiveness of second-language listening strategy instruction: A meta-analysis. *Language Teaching Research*, 13621688211072981.

Du, X., Zhang, M., Shelton, B. E., & Hung, J. L. (2022). Learning anytime, anywhere: A spatio-temporal analysis for online learning. *Interactive Learning Environments*, *30*(1), 34–48. https://doi.org/10.1080/10494820.2019.1633546

Dziuban, C., Graham, C. R., Moskal, P. D., Norberg, A., & Sicilia, N. (2018). Blended learning: the new normal and emerging technologies. *International Journal of Educational Technology in Higher Education*, *15*(1). https://doi.org/10.1186/s41239-017-0087-5

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Godwin-Jones, R. (2021). Evolving technologies for language learning. *Language Learning & Technology*, *25*(3), 6–26.

Hsieh, Y. (2020). Effects of video captioning on EFL vocabulary learning and listening comprehension. *Computer Assisted Language Learning*, *33*(5-6), 567-589.

Hubbard, P. (2017). Foundations of computer-assisted language learning. In J. B. Son, & S. Windeatt (Eds.), *Language teacher education and technology: Approaches and practices* (pp. 153–167). Bloomsbury Publishing.

Jiang, X., Rollinson, J., Plonsky, L., Gustafson, E., & Pajak, B. (2021). Evaluating the reading and listening outcomes of beginning-level Duolingo courses. *Foreign Language Annals*, *54*(4), 974–1002. https://doi.org/10.1111/flan.12600

Jiang, X., Rollinson, J., Plonsky, L., & Pajak, B. (2020). Duolingo efficacy study: Beginning-level courses equivalent to four university semesters. *Duolingo Research Report*, *2020*, 11. https://www.duolingo.com/efficacy

Ko, M. H. (2017). Learner perspectives regarding device type in technology-assisted language learning. *Computer Assisted Language Learning*, *30*(8), 844–863. https://doi.org/10.1080/09588221.2017.1367310

Konishi, H., Kanero, J., Freeman, M. R., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Six principles of language development: Implications for second language learners. *Developmental Neuropsychology*, *39*(5), 404–420. https://doi.org/10.1080/87565641.2014.931961

Krashen, S. (1992). The input hypothesis: An update. *Linguistics and language pedagogy: The state of the art*, 409-431.

Krashen, S. (2014). Does Duolingo "trump" university-level language learning? *International Journal of Foreign Language Teaching*, *9*(1), 13–15. http://sdkrashen.com/content/articles/krashen-does-duolingo-trump.pdf

Kukulska-Hulme, A., Lee, H., & Norris, L. (2017). Mobile learning revolution: Implications for language pedagogy. In *The handbook of technology and second language teaching and learning* (pp. 217–233). Wiley Online Library. https://onlinelibrary.wiley.com/doi/10.1002/9781118914069.ch15

Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL*, *31*, 293–311. https://doi.org/10.1017/S0958344019000065

Loewen, S., Isbell, D. R., & Sporn, Z. (2020). The effectiveness of app-based language instruction for developing receptive linguistic knowledge and oral communicative ability. *Foreign Language Annals*, *53*(2), 209–233. https://doi.org/10.1111/flan.12454

Lord, G. (2015). "I don't know how to use words in Spanish": Rosetta Stone and learner proficiency outcomes. *Modern Language Journal*, *99*(2), 401–405. https://doi.org/10.1111/modl.12234_3

Lord, G. (2016). Rosetta Stone for language learning: An exploratory study. *IALLT Journal of Language Learning Technologies*, *46*(1), 1–35. https://doi.org/10.17161/iallt.v46i1.8552

Lynch, T. (2011). Academic listening in the 21st century: Reviewing a decade of research. *Journal of English for Academic Purposes*, *10*(2), 79-88.

Madden, A. G., Margulieux, L. E., Kadel, R. S., & Goel, A. K. (2019). *Blended learning in practice: A guide for practitioners and researchers*. The MIT Press.

Maftoon, P., & Fakhri Alamdari, E. (2020). Exploring the effect of metacognitive strategy instruction on metacognitive awareness and listening performance through a process-based approach. *International Journal of Listening*, *34*(1), 1–20. https://doi.org/10.1080/10904018.2016.1250632

McCarthy, M. (Ed.). (2016). *The Cambridge guide to blended learning for language teaching*. Cambridge University Press.

Means, B., Bakia, M., & Murphy, R. (2014). *Learning online: What research tells us about whether, when, and how*. Routledge. https://doi.org/10.4324/9780203095959

Patchan, M. M., Schunn, C. D., Sieg, W., & McLaughlin, D. (2016). The effect of blended instruction on accelerated learning. *Technology, Pedagogy and Education*, *25*(3), 269–286. https://doi.org/10.1080/1475939X.2015.1013977

Pearl, M. (2017). *Are Duolingo users actually learning anything useful?* https://www.vice.com/en/article/ezxyyz/are-duolingo-users-actually-learning-anything-useful

Pili-Moss, D., Brill-Schuetz, K. A., Faretta-Stutenberg, M., & Morgan-Short, K. (2020). Contributions of declarative and procedural memory to accuracy and automatization during second language practice. *Bilingualism: Language and Cognition*, *23*(3), 639–651. https://doi.org/10.1017/S1366728919000543

Plonsky, L., & Ziegler, N. (2016). The CALL-SLA interface: Insights from a second-order synthesis. *Language Learning & Technology*, *20*(2), 17–37. https://www.lltjournal.org/item/10125-44459/

Rachels, J. R., & Rockinson-Szapkiw, A. J. (2018). The effects of a mobile gamification app on elementary students' Spanish achievement and self-efficacy. *Computer Assisted Language Learning*, *31*(1–2), 72–89. https://doi.org/10.1080/09588221.2017.1382536

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers and Education*, *147*. https://doi.org/10.1016/j.compedu.2019.103778

Ratzlaff, N. (2015). *A cross-comparison and efficacy study of Duolingo and an entry-level German 1A course.* Fresno State University. https://scholarworks.calstate.edu/concern/theses/7d278v29r

Reinders, H., & Benson, P. (2017). Research agenda: Language learning beyond the classroom. *Language Teaching*, *50*(4), 561–578. https://doi.org/10.1017/S0261444817000192

Riazi, A. M. (2016). *The Routledge encyclopedia of research methods in applied linguistics*. Routledge. https://doi.org/10.4324/9781315656762

Richards, J. C. (2008). *Teaching listening and speaking* (Vol. 35, Issue 4). Cambridge University Press.

Rivera, J. H. (2016). Science-based laboratory comprehension: An examination of effective practices within traditional, online, and blended learning environments. *Open Learning: The Journal of Open, Distance, and e-Learning*, *31*(3), 209–218. https://doi.org/10.1080/02680513.2016.1208080

Rodrigues, P. D., & Vethamani, M. E. (2015). The impact of online learning in the development of speaking skills. *Journal of Interdisciplinary Research in Education*, *5*(1), 2232–180.

Rosell-Aguilar, F. (2018). Autonomous language learning through a mobile application: A user evaluation of the busuu app. *Computer Assisted Language Learning*, *31*(8), 854–881. https://doi.org/10.1080/09588221.2018.1456465

Rost, M. (2013). *Teaching and researching listening*. Routledge.

Şad, S. N., Özer, N., Yakar, Ü., & Öztürk, F. (2020). Mobile or hostile? Using smartphones in learning English as a foreign language. *Computer Assisted Language Learning*, 1–27. https://doi.org/10.1080/09588221.2020.1770292

Smith, B. (2017). Technology-enhanced SLA research. In *The handbook of technology and second language teaching and learning*. (pp. 444–458). Wiley Blackwell.

Suzuki, Y., Jeong, H., Cui, H., Okamoto, K., Kawashima, R., & Sugiura, M. (2023). An fMRI validation study of the word-monitoring task as a measure of implicit knowledge: Exploring the role of explicit and implicit aptitudes in behavioral and neural processing. *Studies in Second Language Acquisition*, *45*(1), 109–136. https://doi.org/10.1017/S0272263122000043

Transparent Language (2013). *Declarative acceleration in the classroom | Transparent Language* [White paper]. https://blogs.transparent.com/language-news/2013/06/03/declarative-acceleration-in-the-classroom/?_ga=2.130015247.2084682707.1661743909-608582353.1660999251

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, *4*(2), 105–122. https://doi.org/10.1017/s1366728901000220

Vadivel, B., Khalil, N. R., Asif, M., & Ajanil, B. (2022). Computer-assisted language learning and English learning in Cihan University: A mixed-methods study. *Education Research International*, *2022*, 8700068. https://doi.org/10.1155/2022/8700068

van Deusen-Scholl, N. (2015). Assessing outcomes in online foreign language education: What are key measures for success? *The Modern Language Journal*, *99*(2), 398-400.

Vandergrift, L. (2004). Learning to listen or listening to learn. *Annual Review of Applied Linguistics*, *24*, 3–25. https://doi.org/10.1017/s0267190504000017

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, *40*(3), 191–210. https://doi.org/10.1017/S0261444807004338

Vesselinov, R., & Grego, J. (2012). The Duolingo efficacy study [White paper]. Duolingo. http://static.duolingo.com/s3/DuolingoReport_Final.pdf

Vesselinov, R., & Grego, J. (2016a). The Babbel efficacy study [White paper]. Babbel.

http://comparelanguageapps.com/documentation/Babbel2016study.pdf
Vesselinov, R., & Grego, J. (2016b). The Busuu efficacy study [White paper]. Busuu. http://comparelanguageapps.com/documentation/The_busuu_Study2016.pdf
Vesselinov, Roumen, Grego, J., Sacco, S. J., & Tasseva-Kurktchieva, M. (2019). *The 2019 Rosetta Stone efficacy study*. http://comparelanguageapps.com/documentation/The2019_RS_FinalReport.pdf