Curriculum Research Journal

Volume 6, Issue 3, summer 2025

Table of Contents

Curriculum Research
Volume 6, Issue 3
Sept. 20, 2025

Comparing the Impact of Cooperative and Competitive Flipped Learning on EFL Learners' Speaking

Article Type:

Original Research

Authors:

Hamid Marashi[1]

ORCiD: 0000-0002-7957-671X

Nazanin Mokhlesi[2]

Article History:

Received: 2025.06.01

Accepted: 2025.08.23

Published: 2025.09.20

Abstract

Ever since its introduction, the flipped learning approach has demonstrated promising results in English language teaching. To this end, the goal of the present study was to compare the effectiveness of flipped learning on speaking in two different cooperative and competitive EFL learning settings. To conduct the study, the researchers selected 60 learners through nonrandom convenience sampling and subsequently randomly assigned them into two experimental groups of 30 learners, namely the flipped learning cooperative and the flipped learning competitive groups. Both groups underwent the same flipped learning treatment; the point of departure, however, was the fact that in one group the learners exercised cooperative learning while the other group experienced competitive learning. Prior to the treatment, all 60 learners took a speaking pretest (a sample Preliminary English Test or PET) and after the treatment, a posttest (another sample PET) was administered to them. An analysis of covariance (ANCOVA) was run on the two groups' scores on the pre- and posttest. The result of the study indicated that the learners sitting in the flipped cooperative group outperformed those in the flipped competitive group. The implications of the findings of this study are elaborated in the paper.

Key Words: Competitive Learning, Cooperative Learning, Flipped Learning, EFL Speaking

1. Department of English, CT.C., Islamic Azad University, Tehran, Iran (Corresponding Author). Email: hamid.marashi@iauctb.ac.ir
2. Department of English, CT.C., Islamic Azad University, Tehran, Iran. Email: na71mokhlesi@gmail.com

1. Introduction

Developing the speaking skill is a fundamental premise of second/foreign language learning. The importance of speaking which allows people to express their minds through a language has indeed risen in significance in modern times with the advent of information and communication technology (Ratnasari, 2020). To this end, English enjoys an unrivaled stance as *the* international language of the world; this drives people from different demographic backgrounds to learn this language and thus be able to communicate with one another regardless of their L1. Accordingly, there seems to be a general consensus that speaking is indeed at the heart of second language learning (Dewi et al., 2016).

The process of developing the speaking skill can be really challenging for many language learners, as it is difficult for them to express their opinions through speaking (Leong & Ahmadi, 2017). Perhaps, one issue which contributes to this hardship, as Indrianty (2016) states, is that while the other skills can be practiced alone, learners cannot really speak on their own. The ELT literature is of course overtly replete with studies aimed at investigating how the learning and teaching of speaking could be enhanced through different methods (e.g., Alonso, 2018; Basturkmen, 2002; Goh & Burns, 2012; Marashi & Khosh-Harf, 2019; Salehi et al., 2015; Souzandehfar, 2024; Tuan & Mai, 2015).

One such method of teaching that has been very much in practice is cooperative learning (Marashi & Gholami, 2020). Cooperative learning "is an approach to group work that minimizes the occurrence of those unpleasant situations and maximizes the learning and satisfaction that result from working on a high-performance team" (Felder & Brent, 2007, p. 36). In other words, as asserted by Slavin and Cooper (2002), cooperative learning entails a relatively small group of learners with different levels of academic achievement levels adopting an array of learning activities than can lead to major academic success for each individual learner within the group.

Accordingly, in a cooperative classroom, students work with their peers in order to do a motivational and instructional task in the class and thus feel more incentivized which can enhance their positive attitudes in school (Stevens, 2008). In the words of Shaaban and Ghaith (2005), within a cooperative learning environment, "Learners may learn together in a classroom climate of academic and personal support in order to read and comprehend a certain text, write an essay, and/or prepare a group project or presentation about certain aspects of the target culture" (p. 17).

Contrary to cooperative learning in which learners are members of a group and work together to achieve the goals, competitive learning is another learning method where students

work alone (Johnson et al., 2014). Competitive learning is defined by students "working against each other to achieve an academic goal such as a grade of A that only one or a few students can attain" (Johnson & Johnson, 2010, p. 204). Both cooperative and competitive learning have been and continue to be studied in the ELT literature (e.g., Cecchini et al., 2021; Farzaneh & Nejadansari, 2014; Gillies, 2016; Johnson & Johnson, 2015; Marashi & Hosseini, 2019; Mendo-Lázaro et al., 2022; Nguyen et al., 2021).

In addition to the traditional choices of language teaching methods, e.g., cooperative/competitive learning, modern information and communication technology has changed the way learners learn language skills. There is of course a myriad of techniques and procedures in this regard with one such technology-based method which has emerged in the last decade being flipped learning. This approach was first popularized by two American chemistry teachers Bergman and Sams (2014). "Flipped learning is a relatively new instructional method which emphasizes effective use of class time by changing the traditional tasks of teachers and students inside and outside the classroom" (Ekmekci, 2017, p. 152).

As noted by Hsieh et al. (2016), in conventional teaching, learners gain new knowledge inside the classroom through lecture and then practice it at home through homework, whereas in flipped learning, students learn new content or lessons at home via watching video clips and then practice it inside the class so that the teacher can monitor them and give them feedback. In other words, "This new learner-centered model foregoes unneeded teacher-talk time during class by scaffolding the learning from the pre-class assignment, and expanding or deepening the learning in class" (Mehring, 2016, p. 1). With its significant application in ELT, numerous studies on flipped learning have been reported in the literature (e.g., Abdullah et al., 2019; Alten et al., 2020; Amini et al., 2022; Bauer-Ramazani et al., 2016; Chen & Hwang, 2020; Khosravani et al., 2020; Lee & Wallace, 2018; Marashi & Eghtedar, 2021; Wang, 2024).

2. Review of the Related Literature

2.1. Speaking

The process of speaking can be defined as producing systematic speech to convey one's massage and "expressing or exchanging thoughts through using language" (Mart, 2012, p. 91). As a communication skill, speaking is a must in second language learning and is basically the process of producing words eligibly for the listeners (Bygate, 1998; Richards, 2008).

Speaking is indeed a highly multifactorial human behavior comprising various linguistic

and nonlinguistic parameters (Bailey, 2003; Menggo, 2018). According to Fasold and Connor-Linton (2014), “General linguistics includes the sounds of language, words and their parts, the structure of sentences, meaning, language change, writing, dialect variation and discourse” (p. 12). And regarding nonlinguistic factors, for instance, in some cases the focus is on facial expression (Burleson & Greence, 2003). However, in other cases, not only the face but also the voice and body gestures can be included (Nowicka & Wilczyn´ska, 2011; Richards & Renandaya, 2002). In effect, “Speaking is the process of building and sharing meaning through the use of verbal and non-verbal symbols, in a variety of contexts” (Chaney, 1998, as cited in Bahadorfar & Omidvar, 2014, p. 9).

Speaking ability is arguably the most essential verbal skill since it is highly needed to perform a conversation (Larasati, 2018; Nation & Newton, 2009). Accordingly, Leong and Ahmadi (2017) assert that, “Humans are programmed to speak before they learn to read and write. In any given, human beings spend much more time interacting orally with language rather than using it in its written form” (p. 35). Hedge (2008) maintains that speaking as a two-way process contains expressing ideas and sharing information and feelings where speaking is considered as the collaboration between two or more persons through which they share the time and context.

2.2. Competitive/Cooperative Learning

Perhaps deeply rooted in the implications of social Darwinism emphasizing the pivotal role of competitiveness in the survival of the fittest, competitive learning is where “individuals seek outcomes that are beneficial to themselves but detrimental to all other group members” (Johnson & Johnson, 2010, p. 202). Prevailing as the dominant mode of instruction at least in the developed world throughout the major bulk of the 20th century, competitive learning is thrusted through participation in a rivalry be it between individuals or groups (Tjosvold et al., 2003).

In effect, within an average competitive classroom, “students are concerned with their individual grades and where they fit into the grade curve” (Kolawole, 2008, p. 4) where the emphasis is placed on outperforming everyone else. According to Johnson et al. (2013), competitive learning is all about the students having to work against one another aimed at gaining a high grade; to this end, one student should achieve the goal while another is bound to fail. Generally, there are two modalities of competition: constructive and destructive (Roseth et al., 2008). In the constructive mode, the person who has won tries to help the person who has lost by giving them suggestions. In the destructive model, however, the loser is not helped by the learners who won the activity (Deutsch et al., 2006, as cited in El-Hallim & Abdalla, 2019).

As of the late 1960s when teachers took up training course about the effective use of small-group techniques in classes (Johnson & Johnson, 2009), competitive learning came under challenge by rising social scientists (e.g., Hartup, 1977; Johnson, 1970; Lewis & Rosenblum, 1975). Gradually, cooperative learning or small-group learning which was perhaps common practice in primordial times grew in prominence and the notion of social interdependence became widely spread (Drakeford, 2012; Nisa & Sulisworo, 2019). Cooperative learning offers learners "an opportunity to be grouped not only heterogeneously by academic performance, but also by race, gender, and language proficiency" (Slavin & Cooper, 2002, p. 649). In a sense, the goal of cooperative learning is to enhance learners' academic intake by allowing them to deliberate, learn from one another, and encourage each other to achieve more (Lucena & San Jose, 2016; Ning, 2011; Tran, 2014).

The pioneers of cooperative learning have demonstrated the usefulness of this mode of learning extensively (e.g., Johnson & Johnson, 1983; Johnson et al., 1990; Kagan, 1995; Slavin, 1992). Following this trend, cooperative learning has been investigated extensively in different contexts. For instance, Er and Aksu Ataç (2014) found that most learners held a positive perspective towards cooperation in the ELT class while Altamimi and Attamimi (2014) and Altun and Sabah (2020) reported that cooperative learning improves learners' speaking skill. According to Darmuki et al. (2017), learners were less anxious and more willing to speak in class through cooperative learning. Johnson and Johnson (2010) stated that through providing a democratic and nonthreatening atmosphere, cooperative learning fosters learners' communication.

In the context of Iran too, cooperative learning has attracted major attention. An early study by Marashi and Baygzadeh (2010) proved the advantageousness of cooperative learning for the overall achievement of EFL learners. Memari Hanjani and Li (2017) reported how cooperative learning boosts reading comprehension. Azizinezhadet al. (2013) and Tabatabaei et al. (2015) also presented such positive results for learners' achievement and motivation while Marashi and Khatami (2017) showed that this mode of learning improves creativity as well. Ahangari and Samadian (2014) demonstrated the effectiveness of cooperative learning on learners' writing and more specifically on comparing cooperative and competitive learning, Marashi and Sanatipoor (2015) concluded that content-based instruction is more beneficial in a cooperative setting for learners' reading and writing.

2.3. Flipped Learning

The term flipped learning – introduced by two Chemistry teachers, Bergmann and Sams, 2014 – is derived from the acronym ***FLIP***: a ***F***lexible environment; a ***L***earning culture where

teachers pay attention to learners' input and abilities and classroom amenities during the pedagogy; ***I***ntentional content through which the teacher intentionally selects the teaching content; and ***P***rofessional educator who knows the teaching material and their students prior to developing a plan (Bauer-Ramazani et al., 2016).

Pudin (2017) noted that the main aim of flipped learning is to change a passive classroom to an active environment in which students collaborate with each other. To this end, Mehring (2016) further added that in an EFL classroom in which flipped learning is used to teach language, learners will have more chances to employ the target language in the context of learning that develops and increases use of the language with quick and effective feedback which is provided by the teacher. In addition, Rahman et al. (2020) stated that by freeing up the class time from instruction, learners have more chances to engage, solve problems, and receive quick feedback.

A significant volume of studies has proven the usefulness of flipped learning. To begin with, Shyr and Chen (2017) asserted that flipped learning methodology can be helpful for students who have low vocabulary acquisition. Chen and Hwang (2020) concluded that flipped methodology can improve learners' learning outcomes. Moreover, Yeşilçınar (2019) noted that "The integration of FCM (Flipped Classroom Model) enhanced adult learners' oral proficiency and changed their opinions towards speaking in a positive manner" (p. 227) while Abdullah et al. (2019) discussed that flipped learning can positively affect learners' speaking skill. Regarding students' engagement, Lee and Wallace (2018) noted that learners who attended flipped classes were more engaged and showed more willingness to ask questions.

In Iran, Mohammadi et al. (2019) showed the effectiveness of flipped learning on EFL learners' English achievement and their willingness to communicate while Khosravani et al. (2020) concluded that flipped learning can significantly affect learners' achievement and autonomy. Yousefzadeh and Salimi (2015) noted that the flipped classroom assists learners to correct misunderstandings and use this new knowledge while they have the opportunity to get feedback from the instructor and classmates. Haghighi et al. (2018) also demonstrated that flipped learning can enhance pragmatic competence and provide plenty of opportunities for learners to communicate and cooperate with their classmates while Marashi and Eghtedar (2021) concluded that learners' motivation and willingness to communicate are boosted through flipped learning.

3. Purpose of the Study

In the currently globalized world, communication plays an important role in each person's

life; in order to have effective communication, speaking is arguably the most significant skill among other language skills (Rao, 2018). Accordingly, seeking more optimal ways to boost learners' speaking is a pivotal and never-ending endeavor in the ELT arena.

As discussed earlier, flipped learning has commonly culminated in positive results while there have been some concerns raised regarding the *one-size-fits-all* approach to teaching which is promoted in this model (Bergmann & Sams, 2016) and also the issue of digital equity where flipped learning may not be available for all socioeconomic cohorts (Sargent & Casey, 2020). Likewise, cooperative learning has shown considerably positive results while there have been certain mixed findings as well. For instance, Marashi and Dibah (2013) concluded that extrovert learners perform better in speaking in cooperative settings while introverts outperform extroverts in competitive settings. Furthermore, Marashi and Gholami (2020) asserted that impulsive learners benefit more from cooperative offline planning contrary to reflective ones who gained more in competitive offline planning settings. Hence, with the occasionally different results in favor of competitive learning, there seems to exist adequate grounds for investigating the impact of flipped learning in a cooperative learning and competitive learning setting to see whether there would be a difference between the effect of this modality of learning in the aforesaid two contexts. In line with the purpose of this study stated above, the following research question was formulated:

- *Is there any significant difference between the effect of using flipped learning in cooperative and competitive learning settings on EFL learners' speaking skill?*

4. Method

4.1. Participants

In order to undertake this research, 60 female intermediate EFL students studying in a language school in Tehran were selected through nonrandom convenience sampling in intact classes. The level of these participants' language proficiency was intermediate as attested by their scores on the placement / achievement tests of the language school they were studying at. Subsequently, these participants were randomly assigned into two experimental groups of 30 learners, namely the flipped learning cooperative and competitive groups. As there was an average of six students in each class, the researchers needed five classes in each group (thus a total of 10 classes) to conduct the research.

4.2. Instrumentation and Materials

The tests and materials used in this study are described below.

4.2.1. Speaking Pretest and Posttest

A sample Preliminary English Test (PET) speaking paper consisting of four parts was used as the pretest of the study at the outset. The scoring was done through four analytical criteria: grammar and vocabulary, discourse management, pronunciation, and interactive communication. Once the treatment was over and for the posttest, another sample PET speaking paper was administered for both groups. Both the pretest and posttest were scored by two raters (one of the researchers in this study and a colleague of hers); the inter-rater reliability is presented in the results section below.

4.2.2. Teaching Materials

The main course book used for this study was *Touchstone 3* by McCarthy et al. (2013) which is a four-level series for adults and young adults with the main objective of integrating speaking, grammar, vocabulary, pronunciation, listening, reading, and writing. *Touchstone 3* is designed to develop conversation strategies, present natural language in authentic contexts, and expose learners to the English which is used in the real world.

As materials can vary in flipped learning, the researchers used different kinds of materials such as pictures, recorded voices, and videos. These materials were prepared based on the activities and parts of the course book that was used inside the classroom. The videos were extracted from online networking sites. These materials were shared with both experimental groups in the WhatsApp groups created for this study.

4.3. Procedure

Once the two experimental groups were set up, the teacher (one of the researchers) started the course by informing the students about the procedure: she explained about the flipped classroom and what they were going to do during the instruction. The teacher taught both groups herself using the same materials during 16 sessions of 90 minutes – two of which were allocated to the pretest and posttest – and prepared some materials such as videos or pictures based on the content of the course and she subsequently sent the materials to the WhatsApp groups that had been set up for each class within each of the two groups. She further asked the students to watch, listen to, or read the materials before attending the class. Each unit of the course book

contained four lessons (A, B, C, and D) and a total of two units were taught in each of the two groups throughout the treatment meaning that about one and a half pages were covered each session.

In both experimental groups, the students were familiar with the content of the lesson as they would have access through the WhatsApp group in line with the flipped learning approach. The basic procedure of a flipped classroom is represented in Figure 1 below:

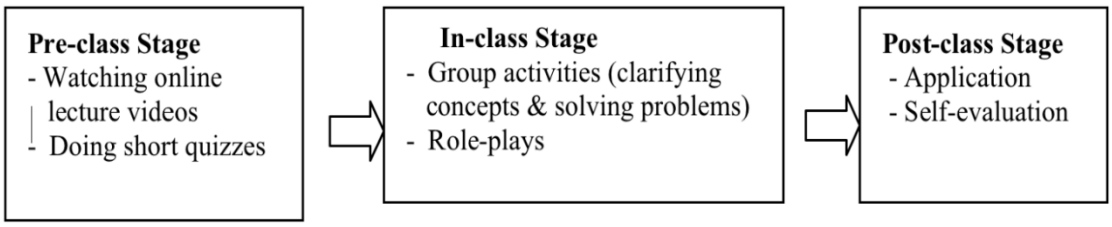


Figure 1.

*Procedure of a flipped classroom (**Estes** et al.**,** 2014)*

In both experimental groups, the teacher/researcher observed the above flipped classroom procedure. In doing so, she reviewed the lesson or, in some cases, taught some parts and asked the students to share their ideas and discuss the issues. Finally, she would ask the students to produce the language they had learned. The point of departure in this study was the duality of the learning settings (i.e., cooperative and competitive learning) and the procedure in each group is discussed in more details in the following parts.

4.3.1. Cooperative Learning Group

On the very first session and seeking to establish a vibrant atmosphere, the teacher and the learners introduced themselves. In this learning setting and depending on the number of the students, the learners were clustered in groups of three. The grouping would change throughout the treatment period so that the learners could experience working with all classmates within small groups. Competitiveness was deemphasized in each group while group work was encouraged.

At the beginning of each session, the teacher engaged the students to build a context and, secondly, in the study stage, they opened their books to do the activities in groups and in the activation part (activating the language), the students did the extra tasks designed by the teacher. Each session, she gave the students linguistic feedback and used various types of error correction methods such as on-the-spot, recasts, and delayed error correction.

As stated earlier, each unit of the book contained four lessons. Lesson A included a

grammar lesson. In the engagement part based on the subject of the lesson, the students discussed a topic in groups (with the students being members of different groups during the term). Since the students were initially introduced to new topics outside the classroom (through the WhatsApp group), they talked about the topic in greater depth. Following the engagement stage, they all opened their books, listened to a conversation, and then practiced it. Subsequently, they listened to a conversation, role played it, did the course book exercises in groups, and checked them.

Next, the teacher taught the grammar box without being the only person to disseminate the information while the students cooperated in doing the exercises and discussing their problems and questions in groups. If there were any problems and questions that had not been solved in the groups, the teacher helped them with their problems. For the activation part, the teacher provided them some extra practices which were supposed to be done in groups and monitored them. The first group to complete the task successfully was the winner.

Lesson B included a grammar section and vocabulary lesson. The class started with the engagement part and the students again discussed a topic in their groups. In this part, the teacher did not apply any error correction methods in order to understand what the main problems were. Next, they did the grammar and vocabulary activities of the course book and checked them in groups. Accordingly, they read their answers first through turn-taking. There was no force on them from the teacher's side.

The teacher revised the vocabularies one more time and then gave them time to look at the definitions. She changed the group and chose two students by chance to come to the board and the other group members had to define the vocabulary and they had to guess. The teacher had written the vocabularies of the lesson on some cards and put them in a box; each group picked up some cards and the teacher asked the students to look the words up in their dictionaries and find some correct examples. The group with the most correct examples in a given time was the winner of that game. The lesson ended with the students doing the designed task in their groups in this part and the teacher monitored them and wrote the mistakes on the board and asked the groups to correct them.

For Lesson C, the students did the exercises of the course book, checked their answers in their groups, and shared any unsolved problems to be solved. The lesson ended with writing a conversation based on the content and the teacher changed the groups. She did not restrict the students to a particular subject in writing the conversation but used the learned structure in their writings. The groups exchanged their writings and they were supposed to correct any mistakes

(the teacher helped them). At the end, when the mistakes were corrected, each group role played the conversation.

Lesson D contained a reading, listening, and speaking section. For the reading part, the teacher used a jigsaw activity where she divided the students into groups of three and chose them randomly. They read a part and summarized it into a short paragraph. Then the teacher asked each group to talk about their paragraph while the other groups listened to retell it. Next, the teacher applied a close reading activity in the groups.

Ultimately, this part normally ended with producing a piece of writing about the content and the group with the least mistakes was the winner. For the listening and speaking part, the teacher changed the groups and asked the students' opinions about the listening topic. The teacher used this opportunity to scaffold and pre-teach the new words. The listening was played twice. During the first time, they took some notes to answer a general question. The second time, they listened to answer the comprehension questions and then checked their answers in the groups; consequently, they talked about the listening content in groups. For the speaking skill, they answered the questions in their books which were related to the topic of the listening. Finally, the teacher asked the students to do the speaking tasks she had designed in advance.

4.3.2. Competitive Learning Group

During the first session, the teacher introduced herself and then asked the learners to introduce themselves one by one. Unlike the cooperative classes, an individualistic learning method was applied in this setting where the whole activities were done individually and not in groups. The teacher applied different error correction techniques in this context as well.

For lesson A, the teacher started the class with the engagement part. She gave them time to think while the students were not allowed to share and talk about the questions with each other. In the study stage, they opened their course books and did the exercises alone and in the activation part, the teacher provided some extra practices as the students were supposed to do the tasks individually and cooperation was nonexistent.

To start Lesson B, the teacher asked the students to discuss a topic in order to engage them. Since they were familiar with the content of the lesson to a certain degree, they were able to use the new content. Each individual had a chance to speak up and express her thoughts. In the study part (after discussing the new content), the students did the course book exercises alone. Then, the teacher designed a game to revise the vocabulary; she called two students to come to the board in this learning setting and they had to compete against each other (whereas

in cooperative learning, the students worked together to find examples for each vocabulary from the box). In this group, each student was supposed to do it alone and the student who could find a correct example for words was the winner. At the end of this lesson, the students did the solo tasks in the form of a worksheet, doing a research, or writing designed by teacher. After doing the tasks, the teacher wrote the mistakes on the board to correct them.

For lesson C, contrary to the cooperative class in which the students were supposed to work in groups, each student did the course book activities alone. They were supposed to write a conversation with some missing parts, they did the task alone, and while they were doing the activity, the teacher monitored and helped them with their questions. Next, the teacher changed their papers and asked another student to fill in the blanks based on the knowledge they had learned throughout the lesson.

Next, they memorized the conversation (the teacher defined the roles and gave them time to role play it and they did not cooperate or practice together before performing in the class). To act out the roles, they stood up and recited the part. The student who got more positive comments from others and managed to perform better than others was the winner.

Lesson D included a reading, listening, and speaking part. For the reading section, the teacher asked each student to read the text and summarize it. The teacher chose two students randomly and asked them to come to the board. The students could write some keywords on the board from their notes to help them with the whole summary and then one by one told the gist of the story. Ultimately a student who could retell the story in a more accurate way was the winner.

For the listening part, the students first listened to take some notes and told the gist of the text while the second time, they listened to answer the comprehension questions (again, there was no cooperation). The speaking part task was done individually and it was mainly in the form of a class discussion in which each student had an equal chance to speak and be active in the class participation. While the students were talking, the teacher was listening to each individual and writing down the mistakes to correct them once the students finished talking.

It is worth noting that in comparison to the speaking questions in the course book, the teacher’s tasks were more expanded and required more knowledge to be accomplished; indeed, the students had to use not only the knowledge they had just learned but also their background knowledge. Immediately after the treatment was over in both competitive and cooperative groups, the 60 learners sat for the same posttest.

5. Results

5.1. Pretest

Once the two experimental groups were randomly assigned, a sample PET speaking was administered as the pretest. Table 1 below shows the descriptive statistics for the pretest. The mean and the standard deviation of the flipped cooperative group were 23.78 and 2.39, respectively, while those of the flipped competitive group stood at 23.95 and 1.81, respectively. In addition, both groups' skewness ratios were inside the acceptable range of ±1.96 (0.107 / 0.427 = 0.25 and 0.389 / 0.427 = 0.91).

Table 1.

Descriptive Statistics of the Scores Obtained by the Two Groups on the Speaking Pretest

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
Coop group	30	20.0	28.0	23.783	2.3914	.107	.427
Comp group	30	21.0	28.0	23.950	1.8116	.389	.427
Valid N (listwise)	30						

Since two raters scored the speaking pretest, the inter-rater reliability of the two was assessed. Table 2 shows the descriptive statistics of the scores they gave the speaking papers of 19 learners selected randomly from among the 60 participants. The mean and the standard deviation of the scores given by rater 1 were 23.16 and 2.14, respectively, while those of rater 2 were 22.79 and 2.14, respectively. As Table 2 reveals, the skewness ratio of both sets of scores (0.14 and 1.22) fell within the acceptable range.

Table 2.

Descriptive Statistics of the Scores Given by the Two Raters to the Speaking Papers

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
Rater 1	19	20	26	23.16	2.141	.075	.524
Rater 2	19	20	28	22.79	2.149	.642	.524
Valid N (listwise)	19						

Consequently, the Pearson Product Moment was run. According to Table 3, the two sets of scores manifested a significant correlation (r = 0.563, p = 0.004 < 0.05).

Table 3.

Inter-Rater Reliability between the Two Raters Scoring the Speaking Papers

		Rater 1	Rater 2
Rater 1			
	Pearson Correlation	1.000	.563**
	Sig. (2-tailed)	.	.004
	N	19	19
Rater 2			
	Pearson Correlation	.563**	1.000
	Sig. (2-tailed)	.004	.
	N	19	19

**Correlation is significant at the 0.01 level

5.2. Posttest

Following the termination of the treatment, the speaking posttest was administered to both groups. Table 4 below shows the descriptive statistics for the posttest. The mean and the standard deviation of the flipped cooperative group were 24.67 and 2.07, respectively, while those of the flipped competitive group stood at 23.12 and 1.96, respectively. Also, both groups' skewness ratios were acceptable (0.49 and 0.07).

Table 4

Descriptive Statistics of the Scores Obtained by the Two Groups on the Posttest

	N	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**	**Skewness**	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
Coop group	30	21.0	28.0	24.667	2.0734	.209	.427
Comp group	30	20.0	27.0	23.117	1.9594	.030	.427
Valid N (listwise)	30						

5.3. Testing the Null Hypothesis

In order to test the null hypothesis, i.e., *there is no significant difference between the effect of using flipped learning in cooperative and competitive learning settings on EFL learners'*

speaking, an analysis of covariance (ANCOVA) was run on both groups' scores on the pre- and posttest. First, the prerequisites for running this parametric test are discussed. To begin with, all sets of scores of course enjoyed normality as demonstrated earlier (Tables 1 and 4). Next, the Levene's test showed no significant difference in the variances ($F_{(1,58)} = 1.509$, $p = 0.224 > 0.05$). Since one covariate was investigated (speaking pretest), the assumption of correlation among covariates was irrelevant. The fourth assumption is linearity; as can be seen in Figure 2 below, the general distribution of the scores is almost linear.

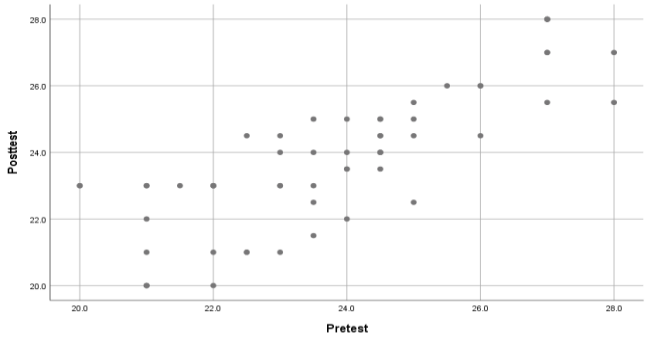


Figure 2.

Histogram of the Linearity of the Scores of the Two Groups on the Pretest and Posttest

As for the fifth assumption, i.e., homogeneity of regression slopes, Table 5 indicates an interaction (i.e. Group * Pretest) of 1.76 which is greater than 0.05; the assumption of homogeneity of regression slopes is thus met.

Table 5.

Tests of Between-Subjects Effects (1)

Source	Type III Sum of Squares	df	Mean Square	*F*	Sig.	Partial Eta Squared
Corrected Model	227.224[a]	3	75.741	94.631	.000	227.224[a]
Intercept	3.863	1	3.863	4.827	.032	3.863
Group	3.138	1	3.138	3.921	.053	3.138
Pretest	184.697	1	184.697	230.759	.000	184.697
Group * Pretest	1.505	1	1.505	1.881	.176	1.505
Error	44.822	56	.800			44.822
Total	34520.750	60				34520.750
Corrected Total	272.046	59				272.046

[a] R Squared = 0.835 (Adjusted R Squared = 0.826)

Table 6 demonstrates that the speaking pretest scores (the covariate in the model) were significant ($F = 233.380$, $p = 0.0001 < 0.05$). Hence, there was a significant difference prior to the

treatment between the two groups' speaking. A significant relationship also existed between the covariate (the pretest) and the dependent variable (the posttest) while the independent variable was controlled ($F = 52.757$, $p = 0.0001 < 0.05$).

Table 6.

Tests of Between-Subjects Effects (2)

Source	Type III Sum of Squares	df	Mean Square	*F*	Sig.	Partial Eta Squared
Corrected Model	225.719[a]	2	112.859	138.860	.000	.830
Intercept	5.719	1	5.719	7.036	.010	.110
Pretest	189.681	1	189.681	233.380	.000	.804
Group	42.878	1	42.878	52.757	.000	.481
Error	46.327	57	.813			
Total	34520.750	60				
Corrected Total	272.046	59				

[a]. R Squared = .830 (Adjusted R Squared = .824)

Hence, the null hypothesis, i.e., there is no significant difference between the impact of flipped cooperative and flipped competitive tasks on learners' speaking was rejected. Those in the flipped cooperative group who achieved a higher mean (Table 4) bore a significantly higher degree of improvement in their speaking than those in the flipped competitive group. Also, the effect size was 0.48; this is a strong effect size (Larson-Hall, 2010).

6. Discussion

The result of this research demonstrating the effectiveness of cooperative learning as compared to competitive learning is in line with the finding of a multitude of studies including *inter alia* those reported by Ghaith (2002), Gillies (2019), Howe (2014), Jacobs and Renandya (2019), and Tamimy et al. (2023), all of which report the more significantly positive impact of cooperative learning on EFL learners' language skills compared to competitive learning. Accordingly, others such as Altamimi and Attamimi (2014) concluded that in a learner-centered atmosphere, learners show more positive attitudes towards speaking. Accordingly, as found by Larasati (2018), learner-centered classes boost learners' speaking skill as this approach provides an action-oriented context in which learners understand the task, take actions towards doing it, and pay attention to their needs and characteristics.

The outcome of this study was also corroborated by Tran (2014) who found that learners having been taught by learning together were able to retain information better. Farzaneh and

Nejadansari (2014) showed that both learners and teachers generally support the application of cooperative learning in reading comprehension classes. Mendo-Lázaro et al. (2022) proved the effectiveness of cooperative learning as a means to encourage students to engage in classroom tasks and develop academic goals. In their study in a university setting, Gul and Shehzad (2015) established the impact of cooperative learning on students' improvement while Nguyen et al. (2021) conducted their study among teachers and found that they too had positive perceptions of implementing cooperative learning. Specifically, the teachers noted that using the cooperative learning procedure in class helped them in "clearly assigning roles for students, setting the stage for learning, and closely monitoring the groups" (p. 246).

In line with the result of this study showing that flipped learning in a cooperative learning setting bore a significantly positive impact on learners' speaking, Abdullah et al. (2019) proved that because of the well-designed tasks in and out of class in flipped learning and the fact that learners have this opportunity to practice, they would be more willing to collaborate and participate in speaking activities. Furthermore, a study by Mohammadi et al. (2019) concluded that employing the flipped learning method provides the opportunity for learners to interact and communicate more with each other. Ekmekci (2017) too found that a flipped classroom "supports and encourages independent and collaborative learning which provides a more flexible learning environment, anytime or anywhere learning, for learners' needs" (p. 163). In addition, Mehring (2016) showed that problem solving, having more time to discuss the problems, and getting feedback from classmates and the teacher are some of the advantages of a flipped model through which a learner-centered procedure is provided where learners are able to use the target language in a more authentic way.

Accordingly, the finding of this study is perhaps of no surprise as it is the very nature of the cooperative learning method that learners have this opportunity to communicate and share their ideas where "since matters are assessed and discussed with reasons and each person defends his/her opinion, a positive and synergistic atmosphere exists and people defend each other and complement each other's thoughts" (Mohammadjani & Tonkaboni, 2015, p. 111). Moreover, since anxiety can negatively affect learners' oral performance, group work can be a solution as, "Cooperative learning is an approach to group work that minimizes the occurrence of those unpleasant situations and maximizes the learning and satisfaction that result from working on a high-performance team" (Felder & Brent, 2007, p. 37).

Another possible justification for the findings of the present study in this respect, as stated by Baker and Clark (2010), is that in a cooperative learning setting, learners feel more encouraged

to do the speaking tasks since they are not afraid of making mistakes in front of other students. Furthermore, as rightly discussed by Namaziandost et al. (2019), implementing cooperative learning can positively affect learners' speaking skill through enhancing their motivation. Contrary to competitive learning settings that are mainly focused on teacher-fronted teaching and learners compete against each other and try to outperform one another individually, cooperative learning encourages learners to be an active participant in the class and have this opportunity to communicate and share their ideas.

The results of the abovementioned studies alongside those of the present study perhaps delineate the point that in terms of the speaking skill, collaboration and learners' togetherness play essential roles and are catalysts for the improvement be it in flipped classrooms or not. The above manifestation is probably indicative of the very nature of cooperative learning in which learners work together, share their ideas, and solve the emerging problems together and there is thus a pretext to gain more through doing the assigned tasks.

7. Conclusion

As highlighted in the previous sections, the results of this study yet again provide further evidence for the significant advantageousness of cooperative learning in the realm of ELT. In simple terms, an increasing number of studies not just in ELT but also in a diversity of educational and vocational fields such as physics (Akinbobola, 2009), management (Anderson, 2006), physical education (Goudas & Magotsiou, 2009), literacy instruction (Stevens, 2003), mathematics (Kolawole, 2008), and chemistry (Nisa & Sulisworo, 2019), only to name a few domains, are factualizing the necessity of incorporating the cooperative learning approach in pedagogy.

It is perhaps a remarkable irony that while there is phenomenal emphasis on cooperation and teamwork in various leading doctrines of management around the globe from total quality management or TQM (Prajogo & Brown, 2004) to the 'no-blame culture' in enterprise leadership and management (Gorini et al., 2012), cooperative learning has yet to become a widespread practice in educational establishments (Slavin, 2015). The perhaps robustly institutionalized theme of competitive learning in the aftermath of the Industrial Revolution and the 20th century, in particular, seems not be easily challengeable despite the plethora of research findings.

The status quo thus requires an active advocacy for a paradigm shift from competitive to cooperative learning in various educational entities with ELT being no exception in this regard.

While teachers stand at the forefront of education inside classrooms and in direct interaction with learners, a multi-stakeholder perspective is called for to facilitate the aforesaid transition smoothly. To this end, policymakers, managers, supervisors, and teachers need to undergo training on cooperative learning in order to be aligned with one another in the entire ecology of an educational institution as it is all too evident that the success of cooperative learning in a classroom necessitates a culture of cooperative work among all those involved (Marashi & Gholami, 2020). Amidst this advocacy for a paradigm shift, one must of course not overlook the pivotal role of the learners themselves and, in the case of young learners, their parents and/or caretakers as they need to be adequately informed on why the well-established practice of individualistic and competitive learning is being substituted with cooperative learning, i.e., a practice which they are most probably not familiar with and hence appears unacceptably unorthodox at first sight and would not receive the support and participation required on their side.

Furthermore, another group of stakeholders in this context is the community of syllabus designers and textbook writers who bear a prominent role. They may decide to add more authentic tasks to course books since in a flipped cooperative classroom, learners have more time to engage in freer practice both before the classroom and within the small learning groups inside the class. Such a procedure would facilitate learning in comparison with conventional course books as they mainly contain controlled or semi-controlled exercises. These tasks and activities appearing in a textbook could be designed and arranged in a way that enables asking learners to cooperate, share, and solve real-life problems while they are completing the task. Needless to say, teachers need to be allowed and encouraged to come on board alongside syllabus designers in the process, thus providing their practical experience and feedback throughout the designing stage.

In conclusion, the researchers must note that they faced a number of limitations in the process of conducting this study which are elaborated here. As noted in the participants section, the learners in this study were all females. It is thus suggested to interested researchers to replicate this research among male learners to identify if gender would bring about differing results. Secondly, the same study could also be conducted in coed classes to see whether the sitting together of male and female learners in cooperative settings would culminate different results or not. Yet another variable that could be involved as an intervening variable thus impacting the finding is age; the same study could be conducted within different age cohorts to check the uniformity of the outcome. Lastly, this study was conducted on the effect of flipped learning in cooperative and competitive learning setting on the speaking skill as a dependent

variable; similar researches could focus on other language skills such as writing to see whether different results would be produced or not.

References

Abdullah, M., Hussin, S., & Ismail, K. (2019). Implementation of flipped classroom model and its effectiveness on English speaking performance. *International Journal of Emerging Technologies in Learning, 14*(9), 130-147. https://dx.doi.org/10.3991/ijet.v14i09.10348.

Ahangari, S., & Samadian, Z. (2014). The effect of cooperative learning activities on writing skills of Iranian EFL learners. *Linguistics and Literature Studies, 2*(4), 121-130. https://dx.doi.org/10.13189/lls.2014.020403.

Anderson, J. R. (2006). On cooperative and competitive learning in the management classroom. *Mountain Plains Journal of Business and Technology. 7*(1), 34-47.

Akinbobola, A. O. (2009). Enhancing students' attitudes towards Nigerian senior secondary school physics through the use of cooperative, competitive, and individualistic learning strategies. *Australian Journal of Teacher Education, 34*(1), 1-9. https://dx.doi.org/10.14221/ajte.2009v34n1.1.

Alonso, R. (2018). *Speaking in a second language*. John Benjamins.

Altamimi, N. O. M., & Attamimi, R. A. (2014). Effectiveness of cooperative learning in enhancing speaking skills and attitudes towards learning English. *International Journal of Linguistics, 6*(4), 27-45. https://dx.doi.org/10.5296/ijl.v6i4.6114.

Alten, D., Phielix, C., Janssen, J., & Kester, L. (2020). Self-regulated learning support in flipped learning videos enhances learning outcomes. *Computers and Education, 158,* 1-16. https://dx.doi.org/10.1016/j.compedu.2020.104000.

Altun, M., & Sabah, R. M. (2020). The effect of cooperative learning strategies in the enhancement of EFL learners' speaking skills. *Asian EFL Journal, 27*(2), 144-171.

Amini, M., Roohani, A., & Jafarpour, A. (2022). Effect of flipped learning on Iranian high school students' L2 grammar achievement and their foreign language anxiety. *Teaching English Language*, *16*(2), 169-201. https://dx.doi.org/10.22132/tel.2022.158977.

Azizinezhad, M., Hashemi, M., & Darvishi. S. (2013). Application of cooperative learning in EFL classes to enhance the students' language learning. *Journal of Procedia Social and*

Behavioral Sciences, 93, 138-141. https://dx.doi.org/10.1016/j.sbspro.2013.09.166.

Bahadorfar, M., & Omidvar, R. (2014). Technology in teaching speaking skill. *International Journal of Multidisciplinary Research, 2*(4), 9-13.

Bailey, K. (2003). *Practical English language teaching: Speaking.* McGraw-Hill.

Baker, T., & Clark, J. (2010) Cooperative learning – a double-edged sword: A cooperative learning model for use with diverse student groups. *Intercultural Education, 21*(3), 257-268. https://dx.doi.org/10.1080/14675981003760440.

Basturkmen, H. (2002). Learner observation of and reflection on spoken discourse: An approach for teaching academic speaking. *TESOL Journal, 11*(2), 26-30. https://dx.doi.org/10.1002/j.1949-3533.2002.tb00080.x.

Bauer-Ramazani, C., Graney, J., Marshall, H., & Sabieh, C. (2016). Flipped learning in TESOL: Definitions, approaches, and implementation. *TESOL Journal, 7*(2), 429-437. https://dx.doi.org/10.1002/tesj.250.

Bergmann, J., & Sams, A. (2014). Flipping for mastery. *Educational Leadership, 71*(4), 24-29.

Bergmann, J., & Sams, A. (2016). *Flipped learning for English language instruction.* ISTE.

Burleson, B. R., & Greence, J. O. (2003). *Handbook of communication and social interaction skills.* Lawrence Erlbaum Associates.

Bygate, M. (1998). Theoretical perspectives on speaking. *Annual Review of Applied Linguistics, 18*(1), 20-42. https://dx.doi.org/10.1017/S0267190500003469.

Cecchini, J. A., Fernandez-Rio, J., Mendez-Gimenez, A., Gonzalez, C., Sanchez-Martínez, B., & Carriedo, A. (2021). High versus low-structured cooperative learning: Effects on prospective teachers' regulation dominance, motivation, content knowledge, and responsibility. *European Journal of Teacher Education, 44,* 486-501. https://dx.doi.org/10.1080/02619768.2020.1774548.

Chen, M. R. A., & Hwang, G. J. (2020), Effects of a concept mapping-based flipped learning approach on EFL students' English speaking performance, critical thinking awareness and speaking anxiety. *British Journal of Educational Technology, 51*(21), 817-834. https://dx.doi.org/10.1111/bjet.12887.

Darmuki, A., Andayani, A., Nurkamto, J., & Saddhono, K. (2017). Cooperative, synectics, and CTL learning models toward speaking ability viewed from students' motivation. *Advances in*

Social Science, Education and Humanities Research, 125, 75-79. https://dx.doi.org/10.2991/icigr-17.2018.18.

Dewi, R., Kultsum, U., & Armadi, A. (2016). Using communicative games in improving students' speaking skills. *English Language Teaching, 10*(1), 63-71. https://dx.doi.org/10.5539/elt.v10n1p63.

Drakeford, W. (2012). The effects of cooperative learning on the classroom participation of students placed at risk for societal failure. *Journal of Psychology Research, 2*(4), 239-246. https://dx.doi.org/10.17265/2159-5542/2012.04.003.

El-Hallim, N. L. A., & Abdalla, M. S. (2019*).* The impact of a competitive learning strategy on developing EFL students' translation skills at the Faculty of Specific Education. *Journal of Research in Curriculum, Instruction, and Educational Technology, 5*(1), 11-44. https://dx.doi.org/10.21608/jrciet.2019.31975.

Ekmekci, E. (2017). The flipped writing classroom in Turkish EFL context: A comparative study on a new model. *Turkish Online Journal of Distance Education, 18*(2), 151-167. https://dx.doi.org/10.17718/tojde.306566.

Er, S., & Aksu Ataç, B. (2014). Cooperative learning in ELT classes: The attitudes of students towards cooperative learning in ELT classes. *International Online Journal of Education and Teaching, 2*(1), 109-122.

Estes, M. D., Ingram, R., & Liu, J. C. (2014). A review of flipped classroom research, practice, and technologies. *The Higher Education Teaching and Learning Portal, 4*. Retrieved on August 1, 2024, from https://www.hetl.org/feature-articles/a-review-of-flipped-classroom-research-practice-and-technologies.

Farzaneh, N., & Nejadansari, D. (2014). Students' attitude towards using cooperative learning for teaching reading comprehension. *Theory and Practice in Language Studies*, *4*(2), 287-292. https://dx.doi.org/10.4304/tpls.4.2.287-292.

Fasold, R. W., & Connor-Linton, J. (2014). *An introduction to language and linguistics*. Cambridge University Press.

Felder, R. M., & Brent, R. (2007). Cooperative learning. *Journal of ACS Symposium Series, 970,* 34-53. https://dx.doi.org/10.1021/bk-2007-0970.ch004.

Ghaith, G. M. (2002). The relationship between cooperative learning, perception of social support, and academic achievement. *System, 30*(3), 263-273. https://dx.doi.org/10.1016/S0346-

251X(02)00014-3.

Gillies, R. (2016). Cooperative learning: Review of research and practice. *Australian Journal of Teacher Education, 41*(3), 39-54. https://dx.doi.org/10.14221/ajte.2016v41n3.3.

Gillies, R. (2019). Promoting academically productive student dialogue during collaborative learning, *International Journal of Educational Research, 97,* 200-209, https://dx.doi.org/10.1016/j.ijer.2017.07.014.

Goh, C. C. M., & Burns, A. (2012). *Teaching speaking: A holistic approach.* Cambridge University Press.

Gorini, A., Miglioretti, M., & Pravettoni, G. (2012). A new perspective on blame culture: An experimental study. Journal of Evaluation in Clinical Practice, 18(3), 671-675. https://dx.doi.org/10.1111/j.1365-2753.2012.01831.x*.*

Goudas, M., & Magotsiou, E. (2009). The effects of a cooperative physical education program on students' social skills. *Journal of Applied Sport Psychology, 21*(3), 356-364. https://dx.doi.org/10.1080/10413200903026058.

Gul, F., & Shehzad, S. (2015). Effects of cooperative learning on students' academic achievement. *Journal of Education and Learning, 9*(3), 246-255. https://dx.doi.org/10.11591/edulearn.v9i3.2071.

Haghighi, H., Jafarigohar, M., Khoshsima, H., & Vahdany, F. (2018). Impact of flipped classroom on EFL learners' appropriate use of refusal: achievement, participation, perception. *Journal of Computer Assisted Language Learning, 32*(3), 261-293. https://dx.doi.org/10.1080/09588221.2018.1504083.

Hartup, W. W. (1977). Peer relations: Developmental implications and interaction in same- and mixed-age situations. *Young Children, 32*(3), 4-13.

Hedge, T. (2008). *Teaching and learning in the language classroom.* Oxford University Press.

Howe, C. (2014). Optimizing small group discourse in classrooms: Effective practices and theoretical constraints. *International Journal of Educational Research, 63,* 107-115. https://dx.doi.org/10.1016/j.ijer.2013.03.011.

Hsieh, J. S. C., Wu, C. V. W., & Marek, M. W. (2016). Using the flipped classroom to enhance EFL learning. *Journal of Computer Assisted Language Learning, 30,* 1-25*.* https://dx.doi.org/0.1080/09588221.2015.1111910.

Indrianty, S. (2016). Students' anxiety in speaking English (a case study in one hotel and tourism college in Bandung). *Journal of English Language Teaching in Indonesia, 4*(1), 28-39.

Jacobs, G. M., & Renandya, W. A. (2019). *Cooperative learning in language education*. Teflin.

Johnson, D. W. (1970). *Social psychology of education*. Holt, Rinehart, & Winston.

Johnson, D. W., & Johnson, R. (1983). Social interdependence and perceived academic and personal support in the classroom. *Journal of Social Psychology, 120*, 77-82. https://dx.doi.org/10.1080/00224545.1983.9712012.

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher, 38*(5), 365-379. https://dx.doi.org/10.3102/0013189X09339057.

Johnson, D. W., & Johnson, R. T. (2010). Teaching students how to live in a democracy. In F. Salili & R. Hoosain (Eds.), *Democracy and multicultural education* (pp. 201-234). Information Age Publishing.

Johnson, D. W., & Johnson, R. T. (2015). Cooperative learning: Improving university instruction by basing practice on validated theory. *Journal on Excellence in College Teaching, 25,* 85-118.

Johnson, D., Johnson, R., & Smith, K. (2014). Cooperative learning: Improving university instruction by basing practice on validated theory. *Journal on Excellence in College Teaching, 25*(3&4), 85-118.

Johnson, D. W., Johnson, R., & Holubec, E. (2013). *Cooperation in the classroom*. Interaction Book Company.

Johnson, D. W., Johnson, R. T., Stanne, M., & Garibaldi, A. (1990). Impact of group processing on achievement in cooperative groups. *Journal of Social Psychology, 130,* 507-516. https://dx.doi.org/10.1080/00224545.1990.9924613.

Kagan, S. (1990). The structural approach to cooperative learning. *Educational Leadership, 47*(4), 12-15.

Khosravani, M., Khoshsima, H., & Mohamadian, A. (2020). On the effect of flipped classroom on learners' achievement, autonomy, motivation, and WTC: Investigating learning and learner variables. *Journal of English Language Teaching and Learning, 12*(25), 175-189. https://dx.doi.org/10.22034/elt.2020.10680.

Kolawole, E. B. (2008). Effects of competitive and cooperative learning strategies on academic performance of Nigerian students in mathematics. *Journal of Educational Research and Review, 3*(1), 3-37.

Larasati, F. (2018). Student centered learning: An approach to develop speaking skill in EFL classroom. *English Community Journal, 2*(1), 153-157. https://dx.doi.org/10.32502/ecj.v2i1.1004.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. Routledge.

Lee, G., & Wallace, A. (2018). Flipped learning in the English as a foreign language classroom: Outcomes and perceptions. *TESOL Quarterly, 52*(1), 62-84. https://dx.doi.org/10.1002/tesq.372.

Leong, L. M., & Ahmadi, M. (2017). An analysis of factors influencing learners' English speaking skill. *International Journal of Research in English Education, 2*(1), 34-41. https://dx.doi.org/10.18869/acadpub.ijree.2.1.34.

Lewis, M., & Rosenblum, L. A. (1975). *Friendship and peer relations*. Wiley.

Lucena, R. E. J., & San Jose, A. (2016). Cooperative learning in enhancing the speaking skills of students: A phenomenological approach. *International Journal of Advanced Multidisciplinary Research, 3*(2), 67-71.

Marashi, H., & Baygzadeh, L. (2010). Using cooperative learning to enhance EFL learners' overall achievement. *Iranian Journal of Applied Linguistics, 13*(1), 73-98. Retrieved on January 23, 2025 from: https://ijal.khu.ac.ir/article-1-46-en.html.

Marashi, H., & Dibah, P. (2013). The comparative effect of using competitive and cooperative learning on the oral proficiency of Iranian introvert and extrovert EFL learners. *Journal of Language Teaching and Research*, *4*(3), 545-555. https://dx.doi.org/10.4304/jltr.4.3.545-556.

Marashi, H., & Eghtedar, D. (2021). Applying the flipped classroom model to foster motivation and willingness to communicate. *Iranian Journal of English for Academic Purposes, 10*(2), 70-89. Retrived on January 31, 2025 from: https://journalscmu.sinaweb.net/article_132909_2b81b204a745f66c66098edd081661d5.pdf.

Marashi, H., & Gholami, M. (2020). Applying cooperative and individual offline planning in speaking classes: A comparison of impulsive and reflective EFL learners. *Journal of Teaching Language Skills, 39*(3.2), 1-34. https://dx.doi.org/10.22099/jtls.2021.38593.2894.

Marashi, H., & Hosseini, S. (2019). Using cooperative group feedback and cooperative group writing in writing classes. *Journal of English Language Pedagogy and Practice, 12*(25),79-103. https://dx.doi.org/10.30495/JAL.2020.675227.

Marashi, H., & Khatami, H. (2017). Using cooperative learning to boost creativity and motivation in language learning. *Journal of Language and Translation, 7*(1), 43-58. Retrieved on February 14, 2025 from: https://sanad.iau.ir/Journal/ttlt/Article/1048102.

Marashi, H., & Khosh-Harf, M. (2019). Comparing critical thinking and autonomy instruction on EFL learners' speaking. *Journal of Language Research, 12*(35), 241-265. https://dx.doi.org/10.22051/jlr.2019.22419.1601.

Marashi, H., & Sanatipoor, S. (2015). Comparing competitive and cooperative content-based instruction in reading and writing classes. *Journal of Studies in Learning and Teaching English, 4*(2), 1-20. https://sanad.iau.ir/Journal/jslte/Article/911522.

Mart, C. T. (2012). Developing speaking skills through reading. *International Journal of English Linguistics, 2*(6), 91-96. https://dx.doi.org/10.5539/ijel.v2n6p91.

McCarthy, M., McCarten, J., & Sandiford, H. (2013). *Touchstone*. Cambridge University Press.

Mehring, J. (2016). Present research on the flipped classroom and potential tools for the EFL classroom. *Computers in the schools, 33*(1), 1-10. https://dx.doi.org/10.1080/07380569.2016.1139912.

Memari Hanjani, A., & Li, L. (2017). Cooperative learning pedagogy: A response to an urgent need in the Iranian EFL reading comprehension context. *Teaching English as a Second Language Quarterly*, *36*(3), 33-58. https://dx.doi.org/10.22099/jtls.2017.27008.2370.

Mendo-Lázaro, S., León-del-Barco, B., Polo-del-Río, M. I., & López-Ramos, V. M. (2022). The impact of cooperative learning on university students' academic goals. *Frontiers in Psychology, 12*. https://dx.doi.org/10.3389/fpsyg.2021.787210.

Menggo, S. (2018). English learning motivation and speaking ability. *Journal of Psychology and Instruction, 2*(2), 64-69. https://dx.doi.org/10.23887/jpai.v2i2.15979.

Mohammadi, J., Barati, H., & Youhanaee, M. (2019). The effectiveness of using flipped classroom model on Iranian EFL learners' English achievements and their willingness to communicate. *Journal of Canadian Center of Science and Education, 12*(5), 101-115. https://dx.doi.org/10.5539/elt.v12n5p101.

Mohammadjani, F., & Tonkaboni, F. (2015). A comparison between the effect of cooperative

learning teaching method and lecture teaching method on students' learning and satisfaction level. *International Education Studies, 8*(9), 107-112. https://dx.doi.org/10.5539/ies.v8n9p107.

Namaziandost, E., Shatalebi, V., & Nasri, M. (2019). The impact of cooperative learning on developing speaking ability and motivation toward learning English. *Journal of Language and Education, 5*(3), 83-101. https://dx.doi.org/10.17323/jle.2019.9809.

Nguyen, C., Trinh, T., Le, D., & Nguyen, K. (2021). Cooperative learning in English language classrooms: Teachers' perceptions and actions. *Anatolian Journal of Education, 6*(2), 89-108. https://dx.doi.org/10.29333/aje.2021.628a.

Ning, H. (2011). Adapting cooperative learning in tertiary ELT. *ELT Journal, 65*(1), 60-70. https://dx.doi.org/10.1093/elt/ccq021.

Nisa, K., & Sulisworo, D. (2019). The effect of cooperative learning and learning style on problem-solving ability in chemistry learning. *Advances in Social Sciences, Education and Humanities Research, 370,* 20-25. https://dx.doi.org/10.2991/adics-elssh-19.2019.5.

Nowicka, A., & Wilczyn´ska, W. (2011). Authenticity in oral communication of instructed L2 learners. In M. Pawlak, E. Waniek-Klimczak, & J. Majer (Eds.), *Speaking and instructed foreign language acquisition* (pp. 24-42). Multilingual Matters.

Nation, I. S. P., & Newton, J. (2009). *Teaching ESL/EFL listening and speaking.* Routledge.

Prajogo, D. I., & Brown, A. (2004). The relationship between TQM practices and quality performance and the role of formal TQM programs: An Australian empirical study. *Quality Management Journal, 11*(4), 31-42. https://dx.doi.org/10.1080/10686967.2004.11919131.

Pudin, C. S. J. (2017). Exploring a flipped learning approach in teaching grammar for ESL students. *Indonesian Journal of English Language Teaching and Applied Linguistics, 2*(1), 51-64. https://dx.doi.org/10.21093/ijeltal.v2i1.47.

Rahman, S. F. A., Yunus, M. M., & Hashim, H. (2020). The uniqueness of flipped learning approach. *International Journal of Education and Practice, 8*(3), 394-404. https://dx.doi.org/10.18488/journal.61.2020.83.394.404.

Rao, S. P. (2018). Developing speaking skills in ESL or EFL settings. *International Journal of English Language, Literature and Translation Studies, 5*(2), 286-293. https://dx.doi.org/10.33329/ijelr.52.286.

Ratnasari, A. G. (2020). EFL students' challenges in learning speaking skills: A case study in

mechanical engineering department. *Journal of Foreign Language Teaching and Learning, 5*(1), 20-38. https://dx.doi.org/10.18196/ftl.5145.

Richards, J. C. (2008). *Teaching listening and speaking: From theory to practice*. Cambridge University Press.

Richards, J. C., & Renandya, W. A. (2002). *Methodology in language teaching.* Cambridge University Press.

Roseth, C. J., Johnson, D. W., & Johnson, R. T. (2008). The relationship between interpersonal relationships and achievement within cooperative, competitive, and individualistic conditions: A meta-analysis. *Psychological Bulletin, 134*(2), 223-246. https://dx.doi.org/10.1037/0033-2909.134.2.223.

Salehi, H., Ebrahimi, M., Sattar, S., & Shojaee, M. (2015). Relationship between EFL learners' autonomy and speaking strategies they use in conversation classes. *Advances in Language and Literacy Studies, 6*(2), 37-43.

Sargent, J., & Casey, A. (2020). Flipped learning, pedagogy and digital technology: Establishing consistent practice to optimise lesson time. *European Physical Education Review*, *26*(1), 70-84. https://dx.doi.org/10.1177/1356336X19826603.

Shaaban, K., & Ghaith, G. M. (2005). The theoretical relevance and efficacy of using cooperative learning in the ESL/EFL classroom. *TESL Reporter, 38,* 14-28.

Shyr, W. J., & Chen, C. H. (2017). Designing a technology-enhanced flipped learning system to facilitate students' self-regulation and performance. *Journal of Computer Assisted Learning, 34*, 53-62. https://dx.doi.org/10.1111/jcal.12213.

Slavin, R. E. (1992). Cooperative learning. In M. D. Gall, P. Joyce, & R. Walter (Eds.), *Applying educational research: A practical guide* (pp. 114-118). Longman.

Slavin, R. E. (2015). Cooperative learning in elementary schools. *Education 3-13, 43*(1), 5-14. https://dx.doi.org/10.1080/03004279.2015.963370.

Slavin, R. E., & Cooper, R. (2002). Improving intergroup relations: Lessons learned from cooperative learning programs. *Journal of Social Issues, 55*(4), 647-663. https://dx.doi.org/10.1111/0022-4537.00140.

Stevens, J. R. (2003). Student team reading and writing: A cooperative learning approach to middle school literacy instruction. *Journal of Educational Research and Evaluation, 9*(2), 137-

160. https://dx.doi.org/10.1076/edre.9.2.137.14212.

Stevens, R. J. (2008). Cooperative learning approach to middle school literacy instruction. In R. M. Gillies, A. F. Ashman, & J. Terwel (Eds.), *The teacher's role in implementing cooperative learning in the classroom* (pp. 92-109). Springer. https://dx.doi.org/10.1007/978-0-387-70892-8_5.

Souzandehfar, M. (2024). Trajectories of multilingual Iranian students' speaking skill development from complex dynamic systems theory perspective. *Applied Research on English Language*, *13*(2), 31-56. https://dx.doi.org/10.22108/are.2024.140705.2240.

Tabatabaei, O., Afzali, M, & Mehrabi, M. (2015). The effect of collaborative work on improving speaking ability and decreasing stress of Iranian EFL learners. *Mediterranean Journal of Social Sciences, 6*(4), 274-280. https://dx.doi.org/10.5901/mjss.2015.v6n4s1p274.

Tamimy, M., Rashidi, N., & Ling Koh, J. H. (2023). The use of cooperative learning in English as foreign language classes: The prevalence, fidelity, and challenges. *Teaching and Teacher Education, 121*. https://dx.doi.org/10.1016/j.tate.2022.103915.

Tjosvold, D., Johnson, D. W., Johnson, R., & Sun, H. (2003). Can interpersonal competition be constructive within organizations? *Journal of Psychology, 137*(1), 63-84. https://dx.doi.org/10.1080/00223980309600600.

Tran, V. D. (2014). The effects of cooperative learning on the academic achievement and knowledge retention. *International Journal of Higher Education, 3*(2), 131-140. https://dx.doi.org/10.5430/ijhe.v3n2p131.

Tuan, N. H., & Mai, T. N. (2015). Factors affecting students' speaking performance. *Asian Journal of Educational Research, 3*(2), 8-23.

Yeşilçınar, S. (2019). Using the flipped classroom to enhance adult EFL learners' speaking skills. *Journal of Language Teaching and Learning, 58*, 206-234.

Yousefzadeh, M., & Salimi, A. (2015). The effects of flipped learning (revised learning) on Iranian Students' learning outcomes. *Advances in Language and Literary Studies, 6,* 209-213. https://dx.doi.org/10.7575/aiac.alls.v.6n.5p.209.

Wang, C. (2024). Six-minute flipped classroom: The effect of using podcasts and collaborative learning on Taiwanese freshmen's English communicative competence. *Teaching English Language*, *18*(1), 1-23. https://dx.doi.org/10.22132/tel.2024.444021.1571

Curriculum Research

Volume 6, Issue 3
Sept. 20, 2025

The Role of Materials in Teaching Listening Strategies in Computer-based and Text-based Classes: EFL Learners' Listening Comprehension and Cognitive Load in Focus

Article Type:

Original Research

Authors:

Maral Katal[1]

ORCiD: 0000-0002-0638-5226

Mojgan Rashtchi[2]

ORCiD:0000-0001-7713-9316

Arshya Keyvanfar[3]

ORCiD: 0000-0003-2803-5910

Article History:

Received: 2025.06.12

Accepted: 2025.09.16

Published: 2025.09.20

Abstract

Listening effectively in a foreign language presents many challenges for students. The aims of this convergent mixed methods study were two-fold. First, to determine whether teaching listening strategies through materials tailored to apply principles related to decreasing Cognitive Load (CL) could help learners improve their L2 listening comprehension. Second, whether using such materials could reduce learners' cognitive load. In doing so, two modes of material presentation (computer-based and text-based) were selected. Eighty-eight TEFL bachelor students in three intact classes were randomly assigned to three treatment conditions: Strategy-Based Cognitive-Load-Decreasing Computer-Based Material (SB-CLD-CBM), Strategy-Based Cognitive-Load-Decreasing Text-Based Material (SB-CLD-TBM), and Strategy-Based Non-Cognitive-Load Decreasing Conventional Learning Material (SB-NCLD-CLM), and received relevant instructions. A general proficiency test was used to examine participants' language proficiency. A listening pretest and posttest, a listening achievement test, and a cognitive load questionnaire were used to collect data in the quantitative phase. A reflection paper, including two questions about students' perceptions administered during the treatment, was used to gather qualitative data. The One-way Analysis of Variance (ANOVA) and Multivariate Analysis of Variance (MANOVA) showed that CLD-CBM and CLD-TBM outperformed the NCLD-CLM in listening comprehension and experienced lower cognitive load. However, no significant difference was found between the experimental groups. The findings are valuable for teachers and materials developers, suggesting they should include Cognitive Load Theory principles and strategy instruction when designing learning materials and instructional approaches.

Key Words: Cognitive Functioning, Computer-based and Text-based Classes, Learning Materials, Listening Comprehension, Listening Strategies

1. Department of Teaching English as a Foreign Language, NT.C., Islamic Azad University, Tehran, Iran. Email: M.katal@iau.ac.ir
2. Department of Teaching English as a Foreign Language, NT.C., Islamic Azad University, Tehran, Iran (Corresponding Author). Email: m_rashtchi@iau.ac.ir
3. Department of Teaching English as a Foreign Language, NT.C., Islamic Azad University, Tehran, Iran. Email: a_keyvanfar@iau.ac.ir

1. Introduction

L2 learners' challenges in completing listening tasks partly stem from the high mental effort required by the skill (Goh & Vandergrift, 2021). The strategic approach, based on the socio-cognitive paradigm, provides a viable solution for enhancing listening comprehension (He & Jiang, 2020). According to Vandergrift and Cross (2018), learning strategies act as practical tools that help students better manage listening tasks by addressing their weaknesses. They give learners cognitive and metacognitive awareness to perform listening activities. Although many studies have explored listening strategies, students and teachers still face difficulties caused by various factors that impede listening success (Goh & Vandergrift, 2021). EFL learners, especially in Iran, face several obstacles in completing listening tasks (Haghighi et al., 2019), emphasizing the need for further research into the variables involved in the comprehension process.

Aside from listening-related elements, one factor influencing listening comprehension is Working Memory (WM), which is affected by learning materials, learning tools, instructional steps, and procedures (Masrai, 2020; Sweller, 2016). Some learning materials place a significant burden on students' mental state and consume their memory resources, resulting in poor listening comprehension (Jiang, 2024). One issue that helps learners handle listening tasks more effectively is to free up their cognitive resources and WM capacity (Paas & van Merriënboer, 2020). Guidance on how to listen effectively can reduce their Cognitive Load (CL) and improve listening skills (Sayyadi, 2019). While CL involves human cognitive structures, designing appropriate educational content and teaching strategies can help learners overcome challenges they face while listening and understanding (Sweller et al., 2011). Cognitive Load Theory (CLT) emphasizes that minimizing the load on WM is essential (Sweller, 2016) and provides guidelines for presenting new information to optimize learners' mental performance (Retnowati et al., 2018).

Similarly, the use of appropriate design procedures in learning materials (Ghalandar-Zehi et al., 2024) is suggested to enhance students' Working Memory Capacity (WMC) (Lwande et al., 2021). Learning materials play a significant role in shaping students' cognitive dispositions and WM mechanisms across various domains of knowledge (Leahy & Sweller, 2011), including listening comprehension (McNamara & Scott, 2001). Accordingly, considering the prominence of listening strategies and CL in fostering learning conditions, and observing students' problems with listening comprehension, the researchers in the current study investigated whether teaching listening strategies through materials designed to apply principles related to decreasing CL could help learners improve their L2 listening comprehension. Besides, whether using such materials could generally reduce their CL. Therefore, by teaching listening strategies and designing materials, the researchers aimed

to alleviate the cognitive burden on students and enhance their listening comprehension. In designing the materials, the researchers followed the twelve instructional design principles identified by Chen et al. (2017) within the CLT framework, which was grounded in empirical research and aimed to optimize learning by managing the cognitive demands placed on learners.

The present study could be significant as it posited that learners' success in listening comprehension was partially contingent upon the CL they experienced during listening tasks. This investigation highlighted the importance of strategies for enhancing listening outcomes. Also, it presumed that reducing the cognitive burden imposed by educational materials could contribute to students' mastery of the skill. The educational materials in this study were tailored through two modes of learning: computer-based and text-based, as explained in the Procedure section.

2. Review of the Related Literature

Listening demands significant mental effort, as listeners must simultaneously discriminate between sounds, interpret intonation and stress, and comprehend vocabulary and grammatical structures (Field, 2008). Vandergrift and Cross (2018) emphasized that listening comprehension and the application of listening strategies help students process language input and perform at the highest level in language learning. The prominence of listening strategies occurs when the listener focuses not only on the listening content but also on how to listen. According to Vandergrift (2004), effective listening requires learners to engage in metacognitive strategies, focusing not only on the content but also on how they process auditory input.

Similarly, adopting a strategic approach is essential for improving receptive auditory processing and has shown promising results in listening comprehension thus far (Nilforoushan et al., 2024); nonetheless, empirical work outside research settings has revealed that students continue to encounter difficulties with listening comprehension (Haghighi et al., 2019). Goh and Vandergrift (2021) and Vandergrift and Cross (2018) proposed effective strategies that enable learners to approach listening tasks with increased efficacy. The instructional phases guide students in organizing cognitive and metacognitive strategies to accomplish a listening task within the framework of pre- listening, while-listening, and post-listening stages.

Teachers should explicitly explain or show each listening strategy in addition to discussing its role in helping learners manage and regulate their listening within a given stage. Teachers should provide students with multiple practice opportunities using strategies, in

addition to encouragement and feedback, to consolidate their understanding. Moreover, exposing learners to a range of different types of listening texts is warranted to promote the transfer of metacognitive strategies across various listening contexts. As learners become more adept at employing metacognitive strategies, teachers reduce the amount of input, feedback, and review to promote independent use and self-evaluation of strategies. As Vandergrift and Cross (2018) argue, separating teaching cognitive from metacognitive strategies is challenging, since while metacognitive strategies direct learning, cognitive strategies involve interaction with learning material; thus, the directive power of metacognitive strategies cannot be fully realized without the application of appropriate cognitive strategies.

An issue that significantly influences the successful performance of listening tasks is encompassed within the CLT proposed by Sweller (1988), which posited that learners have a limited capacity for processing information within WM. When cognitive demands exceed this capacity, learning outcomes decline. In the context of second language (L2) listening comprehension, the concurrent need to decode auditory input, interpret meaning, and retain information can impose a considerable cognitive strain (Field, 2008). Should the task design neglect to consider this load—such as by introducing unfamiliar vocabulary, rapid speech, or complex syntactic structures—learners may encounter difficulties in processing the input (Vandergrift & Goh, 2012). Consequently, instructional strategies aimed at reducing extraneous CL and fostering germane load—such as pre-listening activities, visual aids, or guided strategy instruction—can substantially improve listening performance. Thus, understanding and applying CLT within listening pedagogy offers a valuable framework for optimizing task design and supporting learners' cognitive processing capabilities.

Satori (2021) emphasized that students predominantly face challenges related to memory overload and WM issues, which are critical factors for comprehension. This challenge aligns with CLT, which posits that CL depends on the capacity of WM (Sweller, 2016). Recent research has investigated the impact of CL on various facets of L2 acquisition (Liu et al., 2024). Specifically, some studies have identified the potential effects of CL (Satori, 2021), while others have acknowledged its detrimental impact on listening comprehension (Diao et al., 2007).

An extensively debated topic within the discipline concerns the influence of delivery modality on listening comprehension outcomes (Lehmann & Seufert, 2020) and the enhancement of listening skills through the integration of other competencies (Moussa-Inaty et al., 2012). Nonetheless, the debate about the most effective medium for L2 listening comprehension and how extraneous information affects CL remains unresolved (Chang et al., 2014). Sweller et al. (2011) suggested that the effects of CL vary depending on specific

instructional strategies, arguing that recognizing these effects can enhance learning. Chen et al. (2017) identified twelve key factors that help reduce CL (see Table 6).

A series of studies and meta-analyses on CLT and its effects on CL (Chen et al., 2017) have been proposed in various forms of computer-assisted learning, providing substantial empirical evidence for the instructional guidelines of the CLT framework (Sweller et al., 2011). However, the role of factors that can reduce CL during the process of mastering listening skills—and whether these factors can effectively facilitate listening comprehension—remains under-researched. While CLT (Sweller, 1988) has been widely applied in instructional design across various domains, its specific implications for L2 listening pedagogy are still being developed. Most existing studies focus on identifying challenges learners face during listening tasks, such as speech rate, lexical density, and syntactic complexity, rather than systematically exploring interventions that mitigate cognitive strain. For example, the potential benefits of scaffolding techniques, multimodal input, or strategic pre-listening activities are often acknowledged but not rigorously tested through empirical research.

As a result, there is a gap in understanding how instructional modifications—designed to reduce extraneous load and increase germane load—can be optimized to support learners' cognitive processing during listening. Addressing this gap is crucial for developing evidence-based approaches that not only improve comprehension outcomes but also promote long-term listening proficiency. Therefore, this convergent mixed methods study examined the effect of listening strategy instruction and cognitive load-reducing learning materials on students' listening comprehension and CL. CL principles are applied to redesign learning materials. Such materials are primarily used in computer-based classes; however, this study aimed to investigate whether presenting paper-and-pencil materials versus computer-based materials was more effective in enhancing listening comprehension and reducing CL. The research questions addressed in this study were as follows:

RQ1. Do listening instructions through Strategy-Based Cognitive-Load-Decreasing Computer-Based Material (SB-CLD-CBM), Strategy-Based Cognitive-Load-Decreasing Text-Based Material (SB-CLD-TBM), and Strategy-Based Non-Cognitive-Load Decreasing Conventional Learning Material (SB-NCLD-CLM) differently impact Iranian EFL learners' listening comprehension?

RQ2. Do listening instructions through Strategy-Based Cognitive-Load-Decreasing Computer-Based Material (SB-CLD-CBM), Strategy-Based Cognitive-Load-Decreasing Text-Based Material (SB-CLD-TBM), and Strategy-Based Non-Cognitive-Load Decreasing Conventional Learning Material (SB-NCLD-CLM)

differently impact Iranian EFL learners' cognitive load?

RQ3. How do the participants perceive their cognitive load when using SB-CLD-CBM and SB-CLD-TBM in listening classes?

3. Method

3.1. Design

The researchers employed a convergent mixed methods design to address the research questions. Both quantitative and qualitative data were collected concurrently and analyzed subsequently (Katz-Buonincontro, 2024). The quantitative phase was a quasi-experimental study utilizing a non-equivalent pretest-posttest control group design, implemented across three intact classes (Best & Kahn, 2006). The classes were randomly assigned to three different instructional methods. The study included two experimental groups (SB-CLD-CBM and SB-CLD-TBM) and one control group (SB-NCLD-CLM). The purpose of the control group was to determine whether the observed changes in the experimental groups were attributable to the intervention or chance (Campbell & Stanley, 1963; Mackey & Gass, 2021). Shadish et al. (2002) underscored that the validity of quasi-experimental studies without a comparison or control group is fundamentally compromised. In this investigation, the learning materials served as the independent variable, with three levels: strategy-based, cognitive-load-decreasing computer-based learning materials; strategy-based, cognitive-load-decreasing text-based learning materials; and strategy-based, non-cognitive-load-decreasing conventional learning materials. The dependent variables encompassed listening comprehension and CL.

3.2. Participants

Eighty-eight TEFL bachelor students selected from a pool of 120 learners who had enrolled in a university's regular listening and speaking courses (in three intact classes) participated in this study. The classes were randomly assigned to three treatment groups. The participants included both males and females, with ages ranging from 19 to 25 years old. The selection method was convenience sampling.

3.3 Instruments

A) General Proficiency Test (GPT): A sample of the Preliminary English Test (downloaded from https://www.cambridgeenglish.org) was administered to assess students' homogeneity in English proficiency. The test covered all four skills. The administration took 130 minutes. The test was initially piloted on a group similar to the study participants, consisting of 30 students. Two assistant professors scored the speaking and writing sections. Cronbach's alpha index (r = 0.88) confirmed its reliability. Since the test aimed to ensure homogeneity, 32 out of 120 participants were excluded and instructed by a different instructor in another class, under the supervision of the Head of the TEFL Department.

General Listening Test (GLT): The listening section of the GPT test was considered a General Listening Test (GLT) and served as both a pretest and a posttest. The 25-item test consisted of four listening tasks. The time spent on test administration was 30 minutes. The reliability of the test, as measured by Cronbach's alpha, was acceptable (α = 0.89).

B) Listening Achievement Test (LAT): To control the threat of practice effect, which could occur with the second administration of GLT, the researchers developed the LAT as a post-achievement test. The test consisted of 20 items and involved four sections, each with a different natural setting (e.g., a woman discussing her job). The B estimate was calculated (B: 6.84). The dependability index of the test was calculated using the approach proposed by Subkoviak (2005). The agreement coefficient and Kappa coefficient were 0.83 and 0.35, respectively.

C) Cognitive Load (CL) Questionnaire: NASA Task Load Index (Hart & Staveland, 1988) was used to investigate students' CL. The instrument provides an overall workload score calculated as the sum of six 20-point subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration. The questionnaire was piloted on a sample similar to the study participants. The instrument's reliability index, calculated using Cronbach's alpha before application, indicated its viability (α = 0.91).

D) Reflection Paper: The researchers developed a reflection paper following the suggestions of several researchers for reflection checklists (Chen et al., 2017; Vandergrift & Cross, 2018). Goh (2014) suggests that students' reflection entries should be completed individually by answering questions about what, when, how, why, and who in relation to a specific listening event. Vandergrift and Goh (2012) believed that teachers should use prompts to maintain learners' focus on aspects of learning that require deep thinking and analysis. The reflection papers administered immediately after the listening activities and contained two questions, which helped students evaluate their performance in relation to the CL they experienced. The SB-CLD-TBM and SB-CLD-CBM groups answered the reflection

paper every other session after completing a listening task, answering the following questions:

- Did you feel (mental, physical, temporal) pressure while doing the listening task? How?
- Did you feel frustrated or ineffective? Why?

3.4. Materials

Generally, the materials for the SB-CLD-CBM and SB-CLD-TBM groups consisted of a set of listening strategies (Vandergrift, 1997) and 12 lessons adapted from the listening sections of *Passages One* (Richards & Sandy, 2014). This set of listening strategies included Persian translations and English explanations. Three bilingual experts translated, back-translated, and reviewed the listening strategy set to evaluate its accuracy and content validity. The materials were adapted based on the CL effects demonstrated by Chen et al. (2017) within the CLT framework. They were reorganized and edited according to the validated metacognitive pedagogical sequence for teaching listening strategies (Vandergrift & Cross, 2018). The metacognitive pedagogical sequence served as the fundamental framework for teaching listening strategies. The SB- NCLD-CLM group also covered 12 lessons from the listening sections of *PassagesOne*, but without any manipulation regarding CL effects. The group did not have access to the complete set of listening strategies. Teaching listening strategies was done implicitly, following the metacognitive pedagogical sequence proposed by Vandergrift and Tafaghodtari (2010), which mainly focused on embedded instruction. The addressed models are explained in detail herein.

3.4.1 Sequence of Pedagogical Steps (Vandergrift & Cross, 2018). The researchers redesigned the book's content, following the sequence proposed by Vandergrift and Cross (2018). Corresponding to each stage, the required tasks and listening texts were supplemented with specified listening strategies, and unnecessary parts were removed. The order and content of the presented material, as outlined in these instructional steps, were consistent across all treatment sessions (Tables 1-5).

Table 1.

Pre-listening Stage Activities

Pre-listening Stage	**Strategies**		
Pedagogical Sequence	**Cognitive**	**Metacognitive**	**Affective**
Students were prepared for what they were going to hear and what they were expected to do. Initially, students were informed about the topic (e.g., personality traits) and the type of text they would hear (e.g., a conversation between two friends) by providing students with texts, pictures, discussions, and	Elaboration Note-taking Resourcing Summarizing Inferencing Grouping	Planning Directed-attention	Cooperation Questioning-for Clarification

exercises that activated their prior knowledge, along with any relevant cultural information or supplementary material needed, presented in the form of relevant exercises. Second, students were informed about the purpose of listening (the goal-setting stage). Students would know the specific information they needed to listen for and the degree of detail required. After students got enough information about what they were going to hear, they made predictions to anticipate what they would hear. These predictions formed the backdrop against which listeners could use contextualization to guide their comprehension.			

Table 2.

While-Listening Stage Activities, First Stage

First Listen: First Verification Stage	**Strategies**		
Pedagogical Sequence	**Cognitive**	**Metacognitive**	**Affective**
Students verified predictions and checked comprehension of the desired information. Students were not required to understand every word. After completing their predictions, they listened to the text for the first time. As they listened, they highlighted the predicted words, phrases, and information mentioned in the text and added any other information they understood from the listening task. Students discussed and compared their predictions, adding information in pairs. They also identified their problem in concentrating during the second listen.	Grouping Transfer Elaboration Inferencing Imagery Summarization	Monitoring Planning Selective-attention	Self-encouragement Cooperation Questioning-for clarification

Table 3.

While-Listening Stage Activities, Second Stage

Second Listen: First Verification Stage	**Strategies**		
Pedagogical Sequence	**Cognitive**	**Metacognitive**	**Affective**
As listeners monitored their comprehension, they might face problems that impeded task completion. They might need to revise predictions, using other strategies, or both. They attempted to build upon what they had understood so far to fill in the missing information. Students listened to the text for the second time. They focused on details and what they did not understand during the first listening phase. They wrote down more detailed information and answered the questions presented in the book. Then, they discussed their answers and the text in more detail.	Inferencing Elaboration Grouping Transfer Imagery Repetition Resourcing Note- taking	Monitoring Problem-solving Evaluation Problem-identification	Self-encouragement Cooperation Questioning-for clarification

Table 4.

While-Listening Stage Activities, Third Stage

Third Listen: Third Verification Stage	Strategies		
Pedagogical Sequence	**Cognitive**	**Metacognitive**	**Affective**
Students listened for information that they had not been able to decipher earlier in the class discussion. Students listened to the text for the third time to verify their understanding and get the information they might have missed. After listening, students worked on the focused listening exercises. They attempted to personalize what they had learned by creating stories related to the topic using vocabulary, phrases, and structures presented in the passage.	Summarization Translation Transfer Repetition Elaboration Inferencing	Selective-attention Monitoring Problem-identification	Self-encouragement Cooperation Questioning-for clarification

Table 5.

Activities in Post-Listening Stage

Reflection Stage	Strategies		
Pedagogical Sequence	**Cognitive**	**Metacognitive**	**Affective**
Students reflected on the results of the decision-making process during the listening task, identifying what went well, what went wrong, and what could be done differently. Students reflected on their experience in the listening activity and evaluated the effectiveness of cognitive strategies used and the results of decisions made during the task. Students shared their ideas regarding the approach they have taken. They also discuss what other strategies they will use for the next listening task.	Imagery Transfer Elaboration	Evaluation Problem-identification	Cooperation Taking-emotional-temperature Questioning-for-clarification

3.4.2 Cognitive Load Effects (Chen et al., 2017). In addition to the model mentioned above, the study material was presented in light of CL effects, as outlined by Chen et al. (2017). Using the cognitive-load decreasing material, students could choose whether they needed to access the strategy battery in general, the strategies that might be used for each question, and the examples or definitions of those strategies for each question (Table 6).

Table 6.

Cognitive Load Decreasing Effects on Study Materials (Adopted from Chen et al., 2017)

Goal-free effect: Students used the given information at all or not, or used any of the number of given strategies, or chose the ones they found suitable independently, so they were not following a fixed sequence. They were presented with open-ended tasks rather than goal-specific activities. **Worked example effect:** Students had access to examples of strategies if needed. Examples for doing exercises guide them. **Completion problems:** Students could use previously specified strategies to complete the exercise. They were given completion exercises.

Split-attention effect: Students had access to the complete battery of strategies, including those applicable to each question, along with their corresponding example sentences and definitions.

Redundancy effect: Students could choose to use the available strategy files or not. So the existence of redundant terminologies, definitions, and examples might not exceed students' load.
Modality and the transient information effect: Students had access to the written format of the complete battery of strategies, including the strategies applicable to each question, as well as example sentences and definitions, as the teacher read and explained them. So the visual and auditory channels were integrated.
Imagination effects: the inclusion of an example guided students to imagine and visualize the application of strategies.
Isolated element and element interactivity effect: Students accessed an element of strategies in isolation, including terminology, definitions, and examples for each question, and, if needed, accessed the whole.
Respecting the expertise reversal and the guidance fading effect: The presence of strategies in their material was not in a fixed linear manner, and disappeared little by little as students gained enough expertise. Initially, definitions were provided, followed by examples. Finally, the terminology was eliminated to allow students to feel autonomous and responsible for selecting the appropriate strategy for each question. From the very beginning, each student had the authority to use the available information in the text or not.

3.5. Procedure

The teacher, who was one of the researchers, administered the GPT, GLT, and CL questionnaires to the groups before the treatment. The participants attended 16 sessions, meeting once a week for 180 minutes each. In the experimental groups, the teacher prepared and distributed learning materials based on CL effects and a metacognitive pedagogical sequence to students at every session. The teacher revised the materials to match the learners' skill levels each session. Reflection papers and oral strategy assessments provided feedback to prepare the materials for the next session. Based on this feedback, the teacher adjusted the amount of explicit oral instruction before the listening tasks and the level of scaffolding in the materials for each session. Following Chamot (2004), the teacher then switched from using L1 to L2 for instruction and decreased the explicit teaching of strategies. She focused on providing multiple practice opportunities to ensure that students internalized the strategies and could transfer them to other contexts.

The SB-CLD-CBM group had access to the electronic version of the study materials (Appendix A). Through these materials, implementing CL effects was more promising. Students had access to the strategies (definitions and examples) needed for each listening activity. They could make necessary adjustments based on their needs and use or remove any part of the information they desired. The teacher followed Cross's (2009) model of strategy instruction, which is designed explicitly for technology-integrated instruction. This model combines the use of specific listening strategies with audiovisual technology to encourage ameliorating listening comprehension for students.

Table 7.

Model of Strategy (Cross, 2009)

Instruction Model Using Technology Materials
Identify and analyze factors that may influence the extent of comprehension.
Expose learners to the material and ascertain whether or not they already apply any listening strategies.
Determine suitable metacognitive, cognitive, and social-affective strategies for instruction and consider appropriate activities through which to teach them.
Prepare pre-listening, while listening, and post-listening materials and exercises.
Conduct integrated and informed strategy instruction, provide substantial practice and feedback, and consistently review.
Evaluate the learner instruction regularly and revise where necessary.
Encourage self-evaluation and autonomous use of listening strategies.

Accordingly, during the first sessions, when students were not yet adept at using strategies, all the terminology, definitions, and examples specific to each task were available to them before each task, although the paperclip was attached. By clicking on the paper clip, they could access the required information quickly. As they gained more expertise, less information was provided. Finally, they reached a level at which no information on strategies was available, and they could transfer their learned knowledge to the new listening task.

Explicit strategy instruction continued until the ninth session, during which students could identify which strategy was best suited for a specific listening task. The teacher focused on automating listening strategies in students' minds and removed all strategy-related terminology before each listening task. However, learners still had access to the listening strategies battery at the beginning of each lesson. They practiced strategies for one additional session, session ten, while completing a listening task. In the subsequent sessions, the learners practiced listening comprehension and strategies without access to the battery.

SB-CLD-TBM followed Rubin et al.'s (2007) model of strategy instruction. The model integrates a focus on metacognitive awareness, instruction on how strategies can be applied (through teacher presentation and modeling), the application and practice of strategies, and the evaluation and transfer of strategies (Appendix B). These elements together are at the heart of successful strategy instruction.

Table 8.

Model of Strategy Instruction (Rubin et al., 2007)

Increasing students' awareness of the strategies they were using.
Presenting and modeling strategies to help learners become aware of their own thinking and learning processes.

Providing several practice opportunities to help students use the strategies independently through the gradual elimination of scaffolding Self-evaluation of the usefulness of the strategies Transfer of strategies to new tasks

This group did not work through digital devices. Therefore, to incorporate the requirements of CL effects, the teacher wrote down the names of strategies that could be used for each listening task. She included the definitions and examples of the strategies used for each task on the page. Accordingly, the teacher prepared the materials for each session and handed them out to the students. As the students' mastery of listening strategies improved, the teacher provided less information in the text.

From the beginning to the eleventh session, explicit strategy instruction continued through the provision of definitions, examples, and applications. Students were presented with all the information about the strategies and clues for each listening task, which helped them choose the appropriate strategy. In the seventh session, students understood the definition of strategies but were unable to determine which strategy was most suitable for a specific task. Therefore, in the materials for the next session, only the titles of strategies for each specific task were provided, and instructions were solely in English. In the ninth and tenth sessions, the teacher removed strategies from the materials, requiring students to identify the appropriate strategy for each task. However, they still had access to and could use the complete set of strategies. In the following sessions, students drew on their knowledge and expertise to determine the most suitable strategy for a specific listening task.

The SB-NCLD-CLM group, like the two other groups, was taught the listening sections of *Passages One* (Richards & Sandy, 2014) as their study material and covered 12 chapters. They received the standard instructional sequence of the book (pre-listening, while-listening, and post-listening). However, no explicit mention of listening strategies, scaffolding, or support in this vein, or any other areas, such as extra visual cues, glossaries, or supplementary material, was provided. Strategy instruction was carried out implicitly through the textbook-provided tasks, following Vandergrift and Tafaghodtari (2010) approach, known as the Metacognitive Pedagogical Sequence (Table 9). The sequence employs a holistic approach, providing an implicit, embedded platform for teaching listening strategies through task performance. In this group, the teacher led the students in learning the strategies without explicitly mentioning their definitions and examples. However, cognitive load principles were not included in their study material. This design enabled the SB-NCLD-CLM group to serve as a baseline for evaluating the impact of cognitive load-decreasing strategy-based lessons implemented in the experimental conditions.

Table 9.

Metacognitive Pedagogical Sequence

Pedagogical Stages	Metacognitive Processes
Pre-listening: Planning/ Prediction Stage After students are informed about the topic and text type, they predict the types of information and possible words they may hear.	Planning and directing attention
First Listen: First Verification Stage Students verified their initial hypotheses, making corrections as required, and noted any additional information they understood. Students compared what they understood/wrote with their peers, modified as required, identified what needed resolution, and decided on the important details that required special attention.	Elective attention, monitoring, and evaluation
Second Listen: Second Verification Stage Students' points of earlier disagreement, make corrections, and write additional details that are understood. Class discussion in which all class members contribute to the reconstruction of the text's main point and most pertinent details, interspersed with reflections on how students arrived at the meaning of certain words or parts of the text.	Selective attention, monitoring, evaluation, and problem solving Monitoring, evaluation, and problem-solving
Third Listen: Final Verification Stage Students listen specifically for the information revealed in the class discussion, which they were not able to decipher earlier.	Selective attention, monitoring, and problem solving
Reflection Stage Based on earlier discussion of strategies used to compensate for what was not understood, students write goals for the next listening activity.	Evaluation and planning

The participants retook the GLT during the fourteenth session. Two weeks later, they took the LAT (the unseen listening test) to allow the researchers to examine their achievement and control the threat of practice effect, which could occur with the second administration of GLT. The researchers also assessed the participants' CL.

Students in SB-CLD-CBM and SB-CLD-TBM classes completed a reflection paper every other session based on their experiences with the listening task. The reflection papers focused on students' perception of CL and were used only in the experimental groups who were instructed based on cognitive load-decreasing materials.

4. Results

4.1. Research Question One

The skewness indices (obtained by dividing the statistic by the standard error) ensured that the distribution of scores was normal, as they fell within the range of ±1.96 (Table 10).

Table 10.

Descriptive Statistics for Pretest (GPT, GLT) and Posttest (GLT and LAT)

Test	Group	N	Mean	SD.	Skewness Statistic	Std. Error
GPT	SB-CLD-CBM	29	118.28	5.713	-.180	.434
	SB-CLD-TBM SB-NCLD-CLM	31	116.77	5.993	.243	.434
		28	116.96	6.432	.424	.441
GLT	SB-CLD-CBM	29	14.59	.780	-.550	.434
	SB-CLD-TBM	31	14.87	.991	-.165	.421
	SB-NCLD-CLM	28	14.50	.923	.304	.441
GLT (posttest)	SB-CLD-CBM	29	20.21	1.840	.154	.434
	SB-CLD-TBM	31	19.32	1.851	-.306	-.421
	SB-NCLD-CLM	28	16.68	1.786	-.189	.441
LAT	SB-CLD-CBM	29	20.45	2.339	-.281	.434
	SB-CLD-TBM	31	19.77	1.783	-.088	.421
	SB-NCLD-CLM	28	16.82	1.679	-.355	.441

Two one-way ANOVAs were conducted on English language proficiency (GPT) and listening skills (GLT). The homogeneity of variances for GPT was confirmed ($F_{(2, 85)} = 0.33$, $p = 0.71$). Results showed that the groups did not significantly differ in GPT scores ($F_{(2, 85)} = 0.53$, $p = 0.58$). The homogeneity of variances for GLT was also verified ($F_{(2,85)} = 0.53$, $p = 0.58$). The ANOVA results also indicated no significant difference among the three groups before the treatment ($F_{(2, 85)} = 1.37$, $p = 0.25$) (Table 11).

Table 11.

One-Way ANOVA between GPT and GLT

		Sum of Squares	df	Mean Square	F	Sig.
GPT	Between Groups	39.266	2	19.633	.537	.587
	Within Groups	3108.177	85	36.567		
	Total	3147.443	87			
GLT	Between Groups	2.254	2	1.127	1.37	.258
	Within Groups	69.518	85	.818		
	Total	71.773	87			

After the treatment, a MANOVA was conducted to examine differences among the

groups on two posttests (GLT and LAT). Table 12 presents the descriptive statistics.

Table 12.

Descriptive Statistics on the Two Posttests

Dependent Variable	Group	Mean	Std. Error	97.5% Confidence Interval Lower Bound	Upper Bound
GLT	SB-CLD-CBM	20.207	.339	19.433	20.981
	SB-CLD-TBM	19.323	.328	18.574	20.071
	SB-NCLD-CLM	16.679	.345	15.891	17.466
LAT	SB-CLD-CBM	20.448	.363	19.620	21.276
	SB-CLD-TBM	19.774	.351	18.973	20.575
	SB-NCLD-CLM	16.821	.369	15.979	17.664

The similarity of the post-GLT (*M*=18.77) and post-LAT (*M*=19.06) values showed that the assumption of linearity was satisfied. An examination of the assumption of multivariate outliers revealed no substantial outliers. The Mahalanobis value of 11.076 was less than the critical value. The homogeneity of variance-covariance matrices ($p < 0.001$) indicated a violation of the assumption; therefore, Pillai's Trace was used to check the results (Tabachnick & Fidell, 2013).

Regarding the assumption of multicollinearity, the significant value ($p < 0.001$) indicated a positive correlation between GLT and LAT. The assumption of normality of variances was satisfied for GLT: ($F_{(2, 85)} = .023$, $p = 0.97$); LAT: ($F_{(2, 85)} = 1.38$, $p = 0.25$).

The MANOVA results (Table 13) revealed a statistically significant difference between groups on GLT and LAT, $F_{(4,170)} = 11.60$, $p < .001$; Pillai's Trace=.42; ηp2=.21.

Table 13.

Multivariate Tests for GLT and LAT Posttest

Effect		Value	F	Hypothesisdf	Error df	Sig.	Partial Eta Squared ηp^2
Intercept	Pillai's Trace	.991	4591.410b	2.000	84	.000	.991
	Wilks' Lambda	.009	4591.410b	2	84	.000	.991
	Hotelling's Trace	109.319	4591.410b	2	84	.000	.991
	Roy's Largest Root	109.319	4591.410b	2	84	.000	.991
Group	Pillai's Trace	.429	11.599	4	170.	.000	.214
Wilks' Lambda		.579	13.188b	4	168	.000	.239
Hotelling's Trace		.713	14.792	4	166	.000	.263
Roy's Largest Root		.693	29.452c	2	85	.000	.409

When the results for GLT and LAT were examined separately (Table 14), using a Bonferroni adjustment with an alpha level of 0.25, a statistically significant difference was identified in post-GLT ($F_{(2, 85)}$ = 28.74, p < 0.001; ηp2 =.40) and post-LAT ($F_{(2,85)}$ = 27.75, p < 0.001; ηp2 = .39).

Table 14.

Tests of Between-Subject Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared η^2_p
Corrected Model	LAT	191.815a	2	95.907	28.741	.000	.403
	GLT	212.017b	2	106.008	27.751	.000	.395
Intercept	LAT	30836.381	1	30836.381	9240.914	.000	.991
	GLT	31760.313	1	31760.313	8314.246	.000	.990
Group	LAT	191.815	2	95.907	28.741	.000	.403
	GLT	212.017	2	106.008	27.751	.000	.395
Error	LAT	283.640	85	3.337			
	GLT	324.699	85	3.820			
Total	LAT	31488.000	88				
	GLT	32495.000	88				
Corrected Total	LAT	475.455	87				
	GLT	536.716	87				

a. R Squared = .403 (Adjusted R Squared = .389)
b. R Squared = .395 (Adjusted R Squared = .381)

The Bonferroni test (Table 15) showed a statistically significant difference between SB-CLD-CBM and SB-NCLD-CLM (p < 0.001, 95%, CI = 2.22 to 4.84) and between SB- CLD-TBM and SB-NCLD-CLM (p < 0.001, 95%, CI = 1.36 to 3.93) in GLT. However, no statistically significant differences were observed between the two groups. Regarding LAT, a statistically significant difference was also detected between SB-CLD-CBM and SB-NCLD-CLM (p < 0.001, 95%, CI = 2.23 to 5.03) and between SB-CLD-TBM and SB- NCLD-CLM (p < 0.001, 95%, CI = 1.58 to 4.33).

Table 15.

Bonferroni Post Hoc Multiple Comparisons

Dependent Variable	(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	97.5% Confidence Interval Lower Bound	Upper Bound
GLT	SB-CLD-CBM	SB-CLD-TBM	.88	.472	.193	-.39	2.16
		SB-NCLD-CLM	3.53*	.484	.000	2.22	4.84
	SB-CLD-TBM	SB-CLD-CBM	-.88	.472	.193	-2.16	.39

		SB-NCLD-CLM	2.64*	.476	.000	1.36	3.93
	SB-NCLD-CLM	SB-CLD-CBM	-3.53*	.484	.000	-4.84	-2.22
		SB-CLD-TBM	-2.64*	.476	.000	-3.93	-1.36
LAT	SB-CLD-CBM	SB-CLD-TBM	.67	.505	.556	-.69	2.04
		SB-NCLD-CLM	3.63*	.518	.000	2.23	5.03
	SB-CLD-TBM	SB-CLD-CBM	-.67	.505	.556	-2.04	.69
		SB-NCLD-CLM	2.95*	.510	.000	1.58	4.33
	SB-NCLD-CLM	SB-CLD-CBM	-3.63*	.518	.000	-5.03	-2.23
		SB-CLD-TBM	-2.95*	.510	.000	-4.33	-1.58

4.2. Research Question Two

The researchers compared the groups using data from the CLT questionnaire. The skewness indices (Table 16) ranged between ±1.96, verifying the normality assumption.

Table 16.

Descriptive Statistics of the Scores Distribution

Test	Group	N	Mean	SD.	Skewness Statistic	Std. Error
CL (pretest)	SB-CLD-CBM	29	333.4	96.3	-.543	.434
	SB-CLD-TBM	31	318.0	130.3	-.555	.421
	SB-NCLD-CLM	28	303.9	85.4	-.448	.441
CL (posttest)	SB-CLD-CBM	29	235.5	81.4	.428	-.434
	SB-CLD-TBM	31	225.8	111.9	-.031	.421
	SB-NCLD-CLM	28	307.5	88.1	-.436	.441

The test for homogeneity of variances on cognitive load revealed that the assumption of homogeneity of variances was violated ($F_{(2, 85)} = 3.26$, $p = 0.43$). The ANOVA (Table 17) revealed no statistically significant difference in CL scores among the groups ($F_{(2, 85)} = 0.54$, $p = 0.58$). Welch and Brown-Forsythe tests corroborated the finding by showing no statistically significant difference between the groups (Welch: $p = 0.48$; Brown-Forsythe: $p = 0.57$).

Table 17.

One-way ANOVA for CL

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	12429.463	2	6214.732	.546	.581
Within Groups	966806.901	85	11374.199		
Total	979236.364	87			

The homogeneity of variances was violated for CL ($F_{(2, 85)} = 4.35$, $p = 0.01$). One- way

ANOVA (Table 18) indicated a statistically significant difference in the groups' CL scores ($F_{(2, 85)} = 6.31$, $p < 0.003$).

Table 18.

One-way ANOVA for CL

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	114602.920	2	57301.460	6.316	.003
Within Groups	771197.080	85	9072.907		
Total	885800.000	87			

The results of the Welch and Brown-Forsythe tests supported the finding, showing a statistically significant difference between the groups (Welch: $p < 0.002$ and Brown-Forsythe: $p < 0.003$). Table 19 shows a statistically significant difference between SB-NCLD-CLM and SB-CLD-CBM ($p = 0.02$, 95%, CI = 9.10 to 134.8) and SB-NCLD-CLM and SB-CLD-TBM ($p = 0.006$, 95%, CI = 19.82 to 143.5). However, no statistically significant differences were shown between the SB-CLD-CBM and SB-CLD-TBM groups.

Table 19.

Scheffé Post Hoc Comparisons

		Mean Difference			95% Confidence Interval	
(I) Group	(J) Group	(I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
SB-CLD-CBM	SB-CLD-TBM	9.711	24.608	.925	-51.60	71.02
	SB-NCLD-CLM	-71.983*	25.237	.021	-134.86	-9.10
SB-CLD-TBM	SB-CLD-CBM	-9.711	24.608	.925	-71.02	51.60
	SB-NCLD-CLM	-81.694*	24.834	.006	-143.57	-19.82
SB-NCLD-CLM	SB-CLD-CBM	71.983*	25.237	.021	9.10	134.86
	SB-CLD-TBM	81.694*	24.834	.006	19.82	143.57

4.3. Research Question Three

Students submitted their reflection papers almost simultaneously after receiving the treatment. The analysis revealed that the responses followed a consistent pattern in both the SB-CLD-CBM and SB-CLD-TBM groups, with indications of CL among students from the first to the final session. Nearly half of the students experienced CL in all sessions for three reasons: lack of knowledge and understanding, lack of concentration, and sequential classes.

The first theme, *lack of knowledge and understanding*, emerged as the most common response in this category. Students primarily felt pressure due to their inability to grasp the material and keep pace with the course. Some students expressed that the listening was too fast, while others felt their vocabulary was not extensive enough for complete comprehension. Some excerpts are: "I have a problem hearing the correct

pronunciation of words." Or, "I couldn't understand some parts, and I had to put pressure on my mind to understand." Or, "When I cannot understand a word in a sentence, I feel pressure."

The second theme, "*lack of concentration,*" indicated that listening was a mentally demanding activity that required great concentration and attention. The participants believed that one momentary lapse in their engagement could lead to a total loss of listening, as the comments illustrate: "This class needs much concentration, and if you miss a part, you cannot understand the rest." Or, "I couldn't concentrate."

The third extracted theme, "*sequential classes*," highlighted the role of extraneous factors in the learning procedure. For some students, it was not the listening process or the task itself that required more cognitive involvement; instead, they identified the long hours of classes and successive schedules as the factors that placed them under pressure: "I had many classes before this class."

5. Discussion

Findings related to the first research question revealed that the SB-CLD-CBM and SB-CLD-TBM groups performed better in listening comprehension than the SB-NCLD-CLM group. However, no differences were found between the SB-CLD-CBM and SB-CLD-TBM groups, which can be explained by considering the critical role of WM capacity in processing listening input. WM capacity is limited and cannot function effectively when new information consists of multiple elements (Sweller, 2023) with varying durations (Xie et al., 2020). By adhering to the principles of CLT (Chen et al., 2017), the study materials for these groups enabled students to allocate cognitive resources effectively and maintain adequate processing capacity. The materials helped learners remain mentally alert to relevant information and avoid irrelevant, cognition-consuming details. Consequently, they could utilize their WM capacity more effectively. This finding aligns with Chen et al. (2017), who suggested that less-skilled learners struggle to grasp relevant information while listening; thus, they feel overwhelmed by the significant amount of information, a phenomenon referred to as the narrow limits of change principle in CLT. Listening inefficiency results from the limitations and failures of L2 learners' memory and cognitive systems (Goh, 2023; Satori, 2021). Thus, using appropriate materials can reduce the load on WM and lead to successful listening performance.

The finding aligns with studies that emphasize the importance of listening strategies in developing a skilled listener who can effectively coordinate and synchronize various strategies (Goh & Vandergrift, 2021; Vandergrift & Cross, 2018). In this vein, the researchers

postulate that through listening strategies, students are empowered to control their comprehension process in various listening tasks and manage multiple sources of information and interacting elements simultaneously, resulting in a lower cognitive load. Goh and Vandergrift (2021) argued that through cognitive strategies, students form mental connections with the input by creating mental images or applying their previously learned skills and concepts.

The findings also receive support from cognitive theories of listening, such as Baddeley's (2000) model of WM and the connectionist model (Bechtel & Abrahamsen, 1991). Cognitive theories share fundamental principles regarding WM and cognitive processing capacity. Firstly, learners should attend to the input to process information and decode the signals. They should recognize words and break them into meaningful units simultaneously. This process happens automatically for skilled learners, while less-skilled learners must engage in controlled information processing. These factors underscore the importance of strategy instruction in helping learners quickly recognize and parse input, which was the primary goal of this study. Strategy-based, cognitive-load-decreasing materials helped learners actively listen, constantly manage and regulate their thinking processes, and allocate their cognitive resources for upcoming information.

The lack of differences between the SB-CLD-CBM and SB-CLD-TBM groups verifies earlier findings in this respect, which hold that the human cognitive system processes information regardless of whether the material is presented digitally or in hard copy form. More specifically, in line with Chen et al. (2017), this study shows that WM and long-term memory function similarly in both modes of presentation. Similar studies on receptive skills contradict this finding, as Macedo-Rouet et al. (2003) found that hypertext reading leads to greater comprehension than hard copy. They showed that material provided through computers enhanced learners' perceived CL. Unlike this study, Genç and Gulozer (2013) also found that presenting reading material through hypertext enhances comprehension and promotes students' success.

The answer to the second research question highlighted two issues: first, the SB-CLD-CBM and SB-CLD-TBM groups experienced lower CL than the SB-NCLD-CLM group. Second, no difference existed in CL between the SB-CLD-CBM and SB- CLD-TBM groups. The first issue, the lower CL scores of SB-CLD-CBM and SB-CLD- TBM compared to the SB-NCLD-CLM, can be explained by CLT (Paas & Merrienboer, 2020), the strategy mediation hypothesis (Dunning & Holmes, 2014; Peng & Fuchs, 2017), and the cognitive-affective motivation learning model (McGrew, 2022). Referring to CLT, Sweller (2020, 2023) specified three types of CLs: intrinsic, extrinsic, and germane. The underlying assumption was that for optimal learning conditions, instructional material must be presented in a way

that (1) controls the intrinsic load (e.g., through managing element interactivity and providing earlier information to learners), (2) lessens extraneous cognitive load (e.g., by eliminating learning-irrelevant data), and (3) ensures that cognitive resources to manage both loads are accessible (e.g., by encouraging learners to engage in the activity). Thus, studies in this field focus more on redesigning instructional materials and methodologies to meet the criteria for reducing cognitive load (Sweller, 2020, 2023). Accordingly, the researchers of the current study assume that the study materials helped students maintain enough mental capacity to stay on track, focus on relevant information, address discrepancies and knowledge gaps, process their previous schemas quickly, and manage input without being overwhelmed by large amounts of information. In other words, in line with Chen et al. (2017), the tailored materials controlled intrinsic CL and provided students with shortcuts to navigate the barriers they faced while listening. By using appropriate materials, teachers may decrease learners' cognitive load and enhance their engagement in learning (Dong et al., 2020).

The findings of this study can also be explained by the strategy mediation hypothesis (Dunning & Holmes, 2014; Peng & Fuchs, 2017), which proposes that through adequate WM exercise, it is possible to modify the choice strategies for a specific task. This modification yields improved performance on both trained and untrained tasks (Dunning & Holmes, 2014). The hypothesis assumes that practice-generated differences in WM are provided by compensatory strategies that learners develop during practice (Peng & Fuchs, 2017). Accordingly, in this study, students' WM was enhanced through continuous practice of listening strategies, allowing them to employ appropriate strategies for completing listening tasks. They automatically utilized a task-specific strategy that facilitated their performance and reduced cognitive load. Likewise, Goh (2023) believed that through adequate training, listening strategies can be expected to develop into well-structured, automatized listening skills over time. Learners will continually move between skills and strategies while listening to various forms of discourse with different levels of challenge, as competent listeners sometimes must do.

The findings of this study also support McGrew's (2022) cognitive-affective motivation learning model. He believed motivation, affective factors, and cognitive constructs should be integrated into an optimal learning model. Accordingly, in this study, cognitive load theory principles and listening strategies simultaneously provided the SB-CLD-TBM and SB- CLD-CBM students with both affective and cognitive support to accomplish the learning task.

Regarding the second point, which indicated no significant difference between the SB-CLD-CBM and SB-CLD-TBM groups in cognitive load, the researchers, in line with Chen et al. (2017), believe that the cognitive system exhibits an identical pattern in processing

information, regardless of whether the information is presented digitally or in hard copy format. The existing research reveals that no specific study has focused on the variables examined in this study. However, related studies indicate that the findings of the present study align with those of Chevalère et al. (2021), who investigated differences in inquiry- based learning and Computer-Assisted Instruction. They defined inquiry-based learning as a different approach in which students are encouraged to interact with one another in the classroom. They demonstrated that students with higher working memory capacity learn more effectively through computers. Chang et al.'s (2017) study showed that outdoor ubiquitous learning is more effective than indoor computer-assisted group learning.

To investigate the third research question, the researchers analyzed the reflection papers, which confirmed the quantitative findings, indicating that the SB-CLD-CBM and SB-CLD-TBM groups shared similar views on CL. Thus, working with digital instruments or using a traditional paper format made no difference in CL. Furthermore, half of the students reported experiencing CL up to the final session due to several factors, including a lack of knowledge and understanding, insufficient concentration, and the sequential nature of the classes.

Lack of knowledge and understanding was the most problematic factor in creating CL. This finding broadly supports CLT and assumes that more knowledgeable learners experience less intrinsic CL (Sweller et al., 2019). However, contradictory results from the theory, as well as partially from the current study, have been detected in the literature. For example, Endres et al. (2022) argued that the reverse is relatively accurate in some circumstances, including complex systems (ecological systems). They believed that amateur learners underestimate the issue's complexity, resulting in a lower intrinsic load. In the current study, students repeatedly reported that they could not follow the track because the listening appeared too fast to them, or they lost the track when they could not understand a single word, which led to a total loss of comprehension. What becomes evident by examining students' responses to the researchers' question is that, apart from mastering the language, which facilitates understanding and promotes a peaceful state of mind and a lower cognitive load, it is necessary to employ listening strategies. The reported problems of speed and lack of vocabulary knowledge are factors that could be compensated for using relevant listening strategies (Vandergrift & Cross, 2018). This issue highlights the role of language educators in employing strategic teaching in their programs.

The second item affecting cognitive load was the lack of concentration. In line with Purwanto et al.'s study (2021), the participants reported that they often became distracted. If they lost their focus for even a moment, they missed the entire listening excerpt. The participants' comments indicated that approaching an enormous listening task might result

in missing some parts of the content. In such situations, listening strategies can be helpful (Goh & Vandergrift, 2021). For instance, by employing the elaboration strategy, learners can draw on prior knowledge outside the listening text and connect it to the information acquired from the material to infer missing details. Integrating principles from CLT might be one solution to this issue. Teachers should create a setting in which students have sufficient mental capacity to process new information and apply previously learned knowledge, such as listening strategies, to manage listening challenges, including distractions and a lack of concentration.

The second inference from the students' notes is that they became distracted by irrelevant factors while listening, which reflected the underlying assumptions of CLT. Providing optimal learning conditions through appropriate instructional design can eliminate learning-irrelevant information (Sweller, 2020). Therefore, if students encounter instructional materials intentionally designed to align with the human cognitive structure, including WM and CL effects, the issue of distraction by irrelevant factors might be mitigated. However, this issue is controversial. One group of scholars believes that CL utilizes executive resources that are available for attentional management; thus, CL decreases disturbance. Conversely, another group states that CL demands high levels of concentration, which leads to reduced peripheral processing and decreased distraction.

The next influential factor was sequential classes. The students argued that their inefficiency in the listening class was due to attending too many other classes, which left them exhausted and prevented them from fully engaging mentally in the task. The participants identified cognitive-load-increasing factors, as identified in NASA-TLX, showing consistency between the quantitative and qualitative findings. For example, they mentioned mental demands when they stated that a lack of knowledge, understanding, and focus hindered their success in the task.

6. Conclusion

This study demonstrated that CL is a crucial factor in listening comprehension, and neglecting its impact can prevent teachers and students from reaching their full potential in listening activities. Students' mental capacity is a key concern; by applying CL principles in the design of learning materials, educators can help learners filter out irrelevant information from their cognitive resources, enabling them to achieve goals and complete tasks. The current research emphasizes the importance of teachers' understanding of the comprehension process and how it is accomplished. To support students effectively, teachers should identify the factors that influence this process. They need to understand

that, like other aspects of language learning, students' listening comprehension can be enhanced through carefully planned activities and learning resources. The findings provide valuable insights for teachers and materials developers to incorporate principles of CLT into instructional design. CLT, grounded in our understanding of human cognitive architecture, provides an integrated framework for developing practical learning principles in any educational setting. By offering guidelines for creating successful learning experiences, the theory can also help identify some instructional shortcomings.

Although the researchers attempted to minimize the study's flaws, some limitations remain. First, the CL questionnaire was assessed based on the respondents' self-reports. Second, students' subjective responses to the qualitative questions might not be precise. Further research on three kinds of cognitive load (germane, intrinsic, and extraneous) can elucidate the underlying factors that affect the process of listening comprehension. This study can be illuminating in research that focuses on learners' mental processes in language learning. The researchers suggest integrating teaching language skills while decreasing CL.

References

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Bechtel, W., & Abrahamsen, A. (1991). *Connectionism and the mind: An introduction to parallel processing in networks.* Blackwell.

Best, J. W., & Kahn, J. V. (2006). *Research in education* (10th ed.). Pearson.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Houghton Mifflin.

Chamot, A. (2004). Issues in language learning strategy research and teaching. *Electronic Journal of Foreign Language Teaching*, *1*(1), 14–26.

Chang, C. C., Lei, H., & Tseng, J. S. (2014). *Do English listening outcome and cognitive load change for different media delivery modes in u-learning?* International Association for the Development of the Information Society.

Chen, L., Huang, T-W., & Chen, Y-H. (2017). The effects of inquiry-based information literacy instruction on memory and comprehension: A longitudinal study. *Library & Information Science Research, 39*(4), 256–266. https://doi.org/10.1016/j.lisr.2017.11.003

Chen, O., Paas, F., & Sweller, J. (2023). A cognitive load theory approach to defining and measuring task complexity through element interactivity. *Educational Psychology Review, 35*, Article 63. https://doi.org/10.1007/s10648-023-09782-w

Diao, Y., Chandler, P., & Sweller, J. (2007). The effect of written text on comprehension of spoken English as a foreign language. *The American Journal of Psychology, 120*(2), 237–261. https://doi.org/10.2307/20445397

Dong, A., Jong, M., & King, R. (2020). How does prior knowledge influence learning engagement? The mediating roles of cognitive load and help-seeking. *Frontiers in Psychology, 11*, 1–10, Article 591203. https://doi.org/10.3389/fpsyg.2020.591203

Dunning, D. L., & Holmes, J. (2014). Does working memory training promote the use of strategies on untrained working memory tasks? *Memory & Cognition, 42*(6), 854– 862. https://doi.org/10.3758/s13421-014-0410-5

Endres, T., Lovell, O., Morkunas, D., Rieß, W., & Renkl, A. (2022). Can prior knowledge increase task complexity?–cases in which higher prior knowledge leads to higher intrinsic cognitive load. *British Journal of Educational Psychology*, *93*(S2), 305– 317. https://doi.org/10.1111/bjep.12563

Field, J. (2008). *Listening in the language Classroom*. Cambridge University Press.

Ghalandar-Zehi, B., Rahmanpanah, H., & Mohseni, A. (2024). Exploring inner voice in the EFL classroom: From materials development perspective. *Curriculum Research, 5*(4), 1–25. https://doi.org/https://doi.org/10.71703/cure.2024.1130467

Genç, H., & Gulozer, K. (2013). The effect of cognitive load associated with instructional formats and types of presentation on second language reading comprehension performance. *The Turkish Online Journal of Educational Technology, 12*(4)*,* 171– 182.

Goh, C., & Vandergrift, L. (2021). *Teaching and learning second language listening: Metacognition in action.* Routledge. https://doi.org/10.4324/9780429287749

Goh, C. C. (2023). Learners' cognitive processing problems during comprehension as a basis for L2 listening research. *System*, *119*, Article 103164.

Haghighi, M., Rashtchi, M., & Birjandi, P. (2019). Exploring the role of strategic knowledge and strategic regulation in Iranian EFL learners' listening performance: A structural equation modeling approach. *Journal of Language and Translation, 9*(3*),* 121–136. https://dorl.net/dor/20.1001.1.20088590.2019.9.3.9.1

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp.139–183). North- Holland. https://doi.org/10.1016/S0166-4115(08)62386-9

He, L., & Jiang, Z. (2020). Assessing second language listening over the past twenty years: A review within the socio-cognitive framework. *Frontiers in Psychology, 11*, Article 2123.

Jiang, D. (2024). Cognitive load theory and foreign language listening comprehension. Springer.

Katz-Buonincontro, J. (2024). Convergent mixed methods designs. In J. Katz- Buonincontro (Ed.), *How to mix methods: A guide to sequential, convergent, and experimental research designs* (pp.73–82). American Psychological Association. https://doi.org/10.1037/0000404-005

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology, 25,* 943–951. https://doi.org/10.1002/acp.1787

Lehmann, J., & Seufert, T. (2020). The interaction between text modality and the learner's modality preference influences comprehension and cognitive load. *Frontiers in Psychology. 10,* Article 2820. https://doi.org/10.3389/fpsyg.2019.02820

Liu, H., Liu, Z., Yuan, M., & Chen, T. (2024). The effect of cognitive load on code- switching. *International Journal of Bilingualism*, *28*(3), 513–530. https://doi.org/10.1177/13670069231170142

Lwande, Ch., Muchemi, L., & Oboko, R. (2021). Identifying learning styles and cognitive traits in a learning management system. *Heliyon, 7*(1), Article e07701. https://doi.org/10.1016/j.heliyon.2021.e07701

Macedo-Rouet, M., Rouet, J.-F., Epstein, I., & Fayard, P. (2003). Effects of online reading on popular science comprehension. *Science Communication, 25*(2), 99–128. https://doi.org/10.1177/1075547003259209

Macaro, E. (2001). Analysing student teachers? Code switching in foreign language classrooms: Theories and decision making. *The Modern Language Journal, 85*(4), 531–548. https://doi.org/10.1111/0026-7902.00124

Mackey, A., & Gass, S. M. (2021). *Second language research: Methodology and design (3rd ed.)*. Routledge. https://doi.org/10.4324/9781003188414

Masrai, A. (2020). Exploring the impact of individual differences in aural vocabulary knowledge, written vocabulary knowledge and working memory capacity on explaining L2 learners' listening comprehension*. Applied Linguistics Review, 11*(3), 423–447.

McGrew, K. S. (2022). The cognitive-affective-motivation model of learning (CAMML): Standing on the shoulders of giants. *Canadian Journal of School Psychology, 37*(1), 117–134. https://doi.org/10.1177/08295735211054270

McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition, 29*(1), 10–17. https://doi.org/10.3758/BF03195736

Moussa-Inaty, J., Ayres, P., & Sweller, J. (2012). Improving listening skills in English as a foreign language by reading rather than listening: A cognitive load perspective. *Applied Cognitive Psychology, 26*(3), 391–402. https://doi.org/10.1002/acp.1840

Nilforoushan, S., Rashtchi, M., & Abbasian, Gh. (2024). The role of text type on the use of metacognitive reading strategies: Iranian EFL learners' perceptions in focus. *Curriculum Research, 5*(1), 99–120. https://doi.org/https://doi.org/10.71703/cure.2024.1107725

Paas, F., & Van Merrienboer, J.J.G. (2020). Cognitive load theory: Methods to manage working memory load in the learning of complex tasks. *Current Directions in Psychological Science, 29*(4), 394–398. https://doi.org/10.1177/0963721420922183

Peng, P., & Fuchs, D. (2017). A randomized control trial of working memory training with and without strategy instruction: Effects on young children's working memory and comprehension. *Journal of Learning Disabilities, 50*(1), 62–80. https://doi.org/10.1177/0022219415594609

Purwanto, D., Fadhly, F.Z., & Rahmatunisa, W. (2021). Listening comprehension study: Difficulties and strategies used by college students. *Indonesian Journal of Learning and Instruction, 4*(1), 55–62. https://doi.org/10.25134/ijli.v4i1.4345

Retnowati, R., Suharyati, H., Manurung, R., Maknun, D., Armariena, D. N., Hetilaniar, H., Hasanudin, C., Murniviyanti, L., Anwar, S., & Setyorini, R. (2018). The Effect of environmental teaching method and the level of natural intelligence on the environmental view of the students' behavior. *Journal of Physics: Conference Series, 1114*(1), 1–7, Article 012104. https://doi.org/10.1088/1742- 6596/1114/1/012104

Richards, J. C., & Sandy, C. (2014). *Passages level 1 student's book.* Cambridge University Press.

Rubin, J., Chamot, A. U., Harris, V., & Anderson, N. J. (2007). Intervening in the use of strategies. In A. Cohen & E. Macaro (Eds.), *Language learner strategies: Thirty years of research and practice* (pp. 141–160). Oxford University Press.

Satori, M. (2021). Effects of working memory on L2 linguistic knowledge and L2 listening comprehension. *Applied Psycholinguistics, 42*(5), 1313–1340. https://doi.org/10.1017/S0142716421000345

Sayyadi, M. (2019). The cognitive load of listening activities of a cognitive-based listening instruction. *Indonesian Journal of Applied Linguistics, 9*(2), 202–216.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002*). Experimental and quasi- experimental designs for generalized causal inference*. Houghton Mifflin.

Subkoviak, M. (2005). Empirical investigation of procedures for estimating reliability of mastery tests. *Journal of Educational Measurement. 15*, 111–116. https://doi.org/10.1111/j.1745-3984.1978.tb00062.x

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285.

Sweller, J. (2016). Cognitive load theory, evolutionary educational psychology, and instructional design. In D. C. Geary & D. B. Berch (Eds.), *Evolutionary perspectives on child development and education* (pp. 291–306). Springer. https://doi.org/10.1007/978-3-319-29986-0_12

Sweller, J. (2020). Cognitive load theory and educational technology. *Education TechnologyResearch & Development, 68*(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Sweller, J. (2023). The development of cognitive load theory: Replication crises and incorporation of other theories can lead to theory expansion. *Educational Psychology Review, 35*(4), Article 95. https://doi.org/10.1007/s10648-023-09817-2

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4_7

Sweller, J., Van Merrienboer, J.J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261– 292. https://doi.org/10.1007/s10648-019-09465-5

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Pearson.

Vandergrift, L. (1997). The Cinderella of communication strategies: Reception strategies in interactive listening. *The Modern Language Journal*, *81*(4), 494–505. https://doi.org/10.2307/328892

Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics*, *24*, 3–25.

Vandergrift, L., & Cross, J. (2018). Metacognitive listening strategies. In J. I. Liontas (Eds.), *The TESOL Encyclopedia of English language teaching* (pp. 1–6). John Wiley & Sons. https://doi.org/10.1002/9781118784235.eelt0567

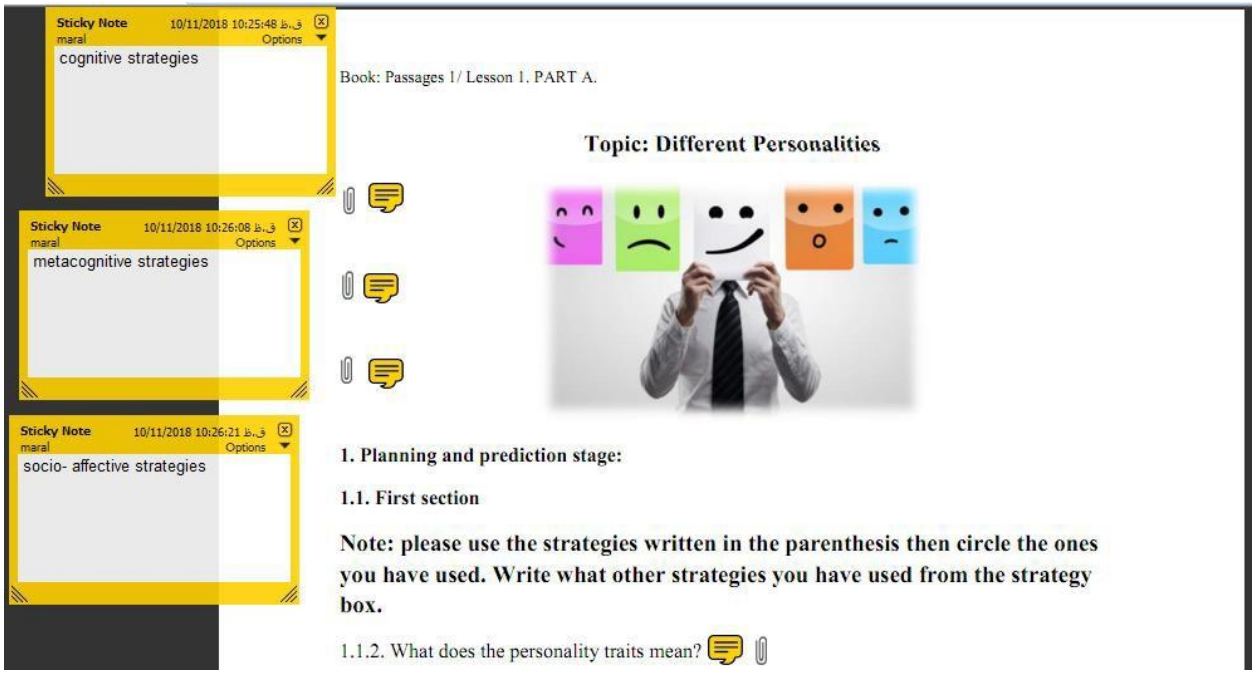
Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action.* Routledge.

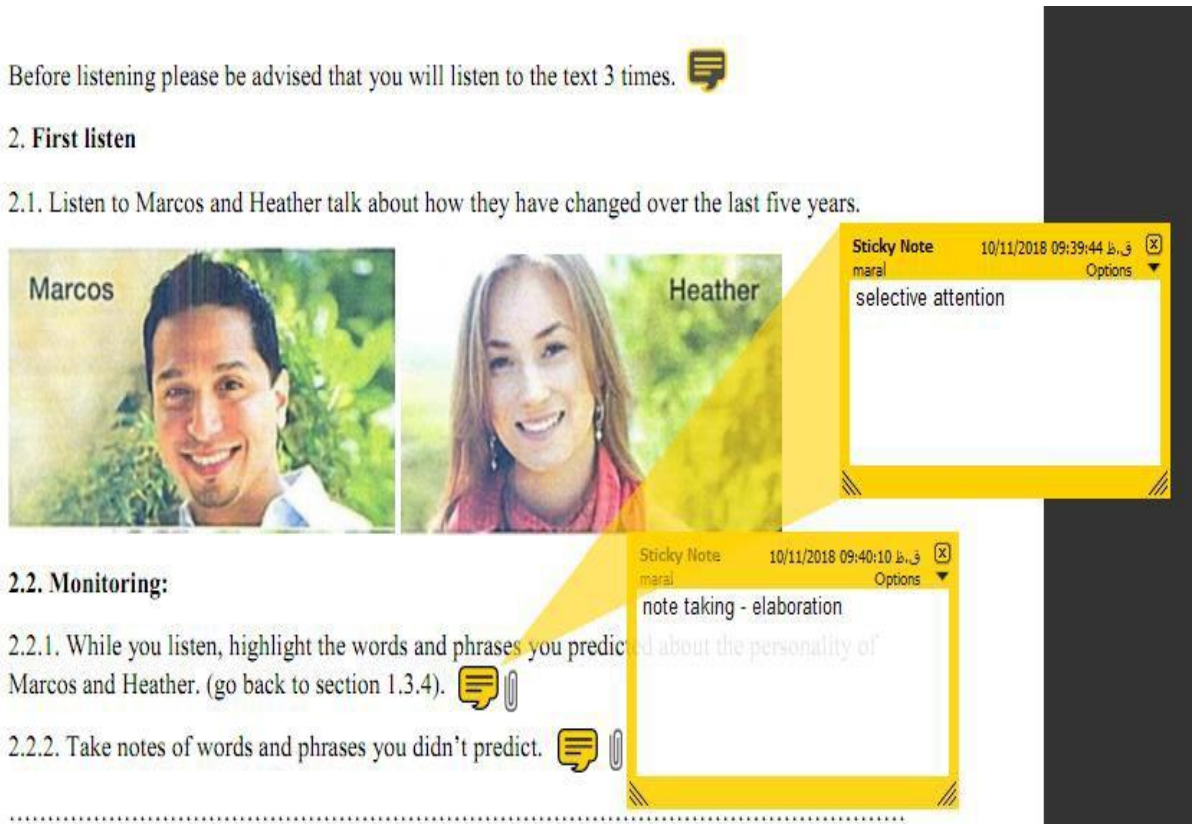
Vandergrift, L., & Tafaghodtari, M. (2010). Teaching L2 learners how to listen does make a difference: An empirical study. *Language Learning, 60*(2), 470–497. https://doi.org/10.1111/j.1467-9922.2009.00559.x

Xie, W., Campbell, S., & Zhang, W. (2020). Working memory capacity predicts individual differences in social distancing compliance during the COVID-19 pandemic in the U.S. *Proceedings of the National Academy of Sciences of the USA, 117*(30),17667–17674. https://doi.org/ 10.31234/osf.io/3j69f

Appendix A

Sample Lesson for Computer-based Instruction





Appendix B

Sample Lesson for Text-based Instruction

Topic: Different Personalities

1. Planning and prediction stage:

1.1. First section

Note: please use the strategies written in the parenthesis then circle the ones you have used. Write what other strategies you have used from the strategy box.

1.1.2. What does the personality traits mean? *(resourcing + elaboration)*

.....
.....
.....

1.1.3. Can you provide me with different personality traits you know? *(elaboration + grouping)*

.....
.....
.....

1.1.4. Search and write on a piece of paper what you have found regarding personality traits in the dictionary or internet. *(resourcing + grouping + note taking)*

Elaboration: using prior knowledge outside the text or conversational context and relating it to knowledge gained from the text or conversation in order to predict outcomes or fill in missing information

Resourcing: using available reference ources of information abut the target language, including dictionaries, textbooks

Grouping: Recalling information based on grouping according to common attributes

e.g., I try to relate the words that sound the same

Note-taking: writing down key words and concepts in abbreviated verbal, graphic, or numerical form to assist performance

Curriculum Research
Volume 6, Issue 3
Sept. 20, 2025

The Impact of Explicit Writing Strategy Instruction on EFL Learners' Strategy Use across Proficiency Levels

Article Type:

Original Research

Authors:

Maryam Khezrinejad[1]

Nooshin Azin[2]

Leila Amiri Kordestani[3]

ORCiD: 0000-0002-4819-9931

Article History:

Received: 2025.05.17

Accepted: 2025.09.03

Published: 2025.09.20

Abstract

Over the past decades, the use of strategies by language learners has been the subject of considerable research. Yet comparatively few studies have examined the direct effects of strategy instruction on the actual writing strategy use of learners, with specific comparisons involving students at different levels of proficiency. This study examined if focused writing strategy instruction can increase the use of writing strategies by EFL learners. Using a mixed-methods design, 64 undergraduate learners in Iran were divided into experimental and control groups. The experimental group received strategy-based instruction grounded in Oxford's (1990) framework, while the control group received conventional writing instruction. Data were collected through pre- and post-tests, stimulated recall protocols, and a validated strategy-use questionnaire. Findings from a mixed between-within ANOVA revealed significant time-group- proficiency interactions, indicating that strategy instruction benefited learners differently across proficiency levels. Qualitative analysis further highlighted how learners integrated strategies into their writing practices. The results have implications for tailoring writing instruction to proficiency-specific needs.

Key Words: Academic Writing, L2 Writing, Learner Proficiency, Metacognitive Awareness, Writing Strategy Instruction

1. English Language Department, Isl.C., Islamic Azad University, Islamshahr, Iran (Corresponding Author). Email: khezrinejad@iau.ac.ir
2. English Language Department, Isl.C., Islamic Azad University, Islamshahr, Iran. Email: nooshin.azin@iau.ac.ir
3. English Language Department, ShQ.C., Islamic Azad University, Shahr-e Qods, Iran. Email: l.amirikordestani@iau.ac.ir

1. Introduction

Within the developing discipline of English as a Foreign Language (EFL), the significance of learner strategies has grown more salient, particularly with respect to productive skills like writing. Although research on writing strategies has abounded over the past decades, a considerable lacuna still exists with regard to the impact of explicit strategy instruction on the actual writing behavior of learners, especially at various proficiency levels (Macaro, 2006). L2 writing is a multidimensional cognitive and linguistic task, which frequently demands that learners utilize both their linguistic resources and strategic competence to successfully orchestrate the writing process (Anderson, 2003; Kellogg, 1996).

In spite of broad recognition of the advantages of strategy use, it is not clear whether learners at different proficiency spectrums benefit equally from strategy instruction. Some research indicates that high-proficiency learners more easily internalize and utilize strategies (Ikeda & Takeuchi, 2003), yet other research has reported substantial gains for lower-proficiency students where instruction is suitably scaffolded (Vandergrift & Tafaghodtari, 2010). Such mixed outcomes point to a need for more sensitive examination of how strategy instruction works within varied learner groups.

In addition, most of the research that does exist has emphasized test scores over an investigation of how learners actually use strategies in the process of writing. To gain insight into the strategic processes learners employ—not merely their results—methodological windows into internal cognitive activity are necessary. Stimulated recall, specifically, has emerged as a fruitful method for prompting reflective explanation of strategic choices during writing, particularly when verbalization is scaffolded through recorded performance (Gass & Mackey, 2000).

This research adds to the burgeoning literature on L2 writing strategies through its investigation of the effect of an intervention intended to foster strategic writing behaviors in university students participating in an L2 writing course. More specifically, it explored how learners of different proficiency levels used strategies after instruction, with particular emphasis on their planning, monitoring, and revision activities. Through the combination of qualitative insights from stimulated recall interviews and quantitative comparison of strategy frequency, the study sought to present a fuller portrait of the developmental effect of strategy instruction. The results may have implications for writing practitioners, curriculum developers, and educational policymakers interested in improving writing instruction in diverse and multilingual academic environments.

2. Theoretical and Empirical Background

Research into second language (L2) writing strategies has developed considerably over the last decades with an increasing agreement that strategic competence is a key aspect of successful writing (Macaro, 2006; Manchón, 2008). Unlike any other language skill, writing is distinguished by its recursive nature in which constant interaction among cognitive, linguistic, and metacognitive processes is necessary (Flower & Hayes, 1980; Kellogg, 1996). Strategic writing behaviors of planning, monitoring, and revising have been found to play an instrumental role in determining the quality and coherence of learners' written production (Anderson, 2005; Graham, 1997).

2.1. Models of Strategic Writing

Among theoretical models undergirding writing strategy research, Kellogg's (1996) working memory model continues to be central. It positions writing as an executive function relying strongly on the coordination of idea generation, language formulation, and monitoring. Likewise, Flower and Hayes (1980) underscored writing as a problem-solving process, during which writers are continually assessing goals, audience, and rhetorical structures. More recently, Grabe and Zhang (2023) maintained that strategic L2 writing needs to be understood as a dynamic between long-term writing development and moment-by-moment decision-making, requiring both procedural knowledge and strategic flexibility.

The applicability of these models is especially evident in the case of learners with varied proficiency levels. As proficiency level rises, learners tend to demonstrate more capacity to balance various sub-processes in writing, including idea generation and linguistic accuracy, while using metacognitive strategies concurrently (Teng, 2024). In contrast, lower-proficiency learners can be less able to distribute cognitive resources effectively, making explicit strategy instruction necessary.

2.2. Strategy Instruction and L2 Writing Development

There is a strong body of research that attests to the efficacy of explicit strategy instruction in enhancing the writing performance of learners (Plonsky, 2011; Sasaki, 2002). Strategy instruction has demonstrated to improve both the quality and quantity of strategy deployment, resulting in higher metacognitive control and greater writing fluency. Nevertheless, the degree to which learners take up and transfer instructed strategies is moderated by aspects of language proficiency, task type, and instructional design (Macaro, 2010; Manchón et al., 2007).

Empirical research has provided mixed findings on the relative advantage of strategy training by proficiency level. Ikeda and Takeuchi (2003) reported greater benefits for more

proficient learners from reading strategy instruction, while Vandergrift and Tafaghodtari (2010) reported larger gains for less proficient learners in a listening-based intervention. Such inconsistencies indicate that the structure of strategy instruction needs to be attuned to learners' starting strategy repertoires and cognitive readiness (Teng & Zhang, 2021).

Additionally, a recent longitudinal study conducted by Horbach and Giannakaki (2023) illustrated that differentiated instruction, whereby learners choose and consolidate strategies on the basis of reflective assessment, results in more robust writing performance and strategic awareness gains. This corroborates the case presented by Manchón (2008), who advocated for individualized, needs-driven instruction that supports learner independence and problem-solving conduct.

2.3. Capturing Strategic Thinking: Methodological Considerations

Accessing learners' internal strategic processes continues to be a methodological issue for L2 writing research. Though questionnaires like Oxford's (1990) Strategy Inventory for Language Learning (SILL) have been extensively employed, they tend to be subject to self-report biases and miss the dynamic process of strategy use (Wharton, 2000). Think-aloud protocols have been proposed as a more immediate way of accessing strategy use in the course of task performance (Anderson, 2005), though they have the potential to interfere with the normal writing process, especially for lower level learners (Sasaki, 2000).

Stimulated recall, on the other hand, has become an increasingly versatile and learner-responsive technique for eliciting strategic behavior. By having learners comment on recorded writing sessions, it enables researchers to tap into both overt and covert strategies, with minimal interference with the task itself (Gass & Mackey, 2000). Current methodological overviews suggest stimulated recall as a complementary instrument that adds qualitative depth and ecological validity to research on strategy use (Zhang & Teng, 2023).

While the literature warrants the benefit of strategy instruction, there is limited research on its effect in the specific area of L2 academic writing and across different proficiency levels. In addition, the majority of existing studies have focused on performance outcomes instead of strategic development. The current study attempted to fill this research void by exploring how explicit writing strategy instruction affected the strategic behavior of students at different proficiency levels through the use of stimulated recall interviews and frequency analysis of strategy deployment. Accordingly, the following research questions were proposed:

1. Does explicit writing strategy instruction significantly increase the use of writing strategies among EFL learners?

2. Does learners' proficiency level influence the effects of explicit writing strategy instruction on their writing strategy use?

3. What insights do EFL Learners provide regarding the effectiveness of writing strategy instruction?

3. Method

3.1. Research Design

This research followed a quasi-experimental, mixed-methods design in examining the effect of focused writing strategy instruction on students' utilization of academic writing strategies at various proficiency levels. The design involved a pre-test/post-test control group design, augmented by qualitative data gathering using stimulated recall interviews. The aim was to examine both the frequency and quality of using writing strategies before and after the intervention, with emphasis on metacognitive and cognitive strategy engagement in the course of writing tasks.

3.2. Participants

The participants were 64 undergraduate students enrolled in an English course at a large public university in Iran. They were in their second year of study and represented diverse academic disciplines, including science, engineering, and the social sciences. All participants had studied English as a foreign language for at least seven years through the national curriculum. They were selected through non-random convenience sampling, as intact classes were available and accessible for the purposes of this study.

A placement writing task and an institutional language exam were used to categorize students into higher proficiency (CEFR B2–C1) and lower proficiency (A2–B1) groups. The participants were then randomly assigned into two groups: experimental (n = 32) and control (n = 32). Each group contained 16 higher proficiency and 16 lower proficiency learners. Gender distribution was roughly balanced (58% female, 42% male), and all students gave informed consent for their participation. Ethical approval for the study was obtained from the university's research ethics committee.

3.3. Data Collection Instruments

3.3.1. Writing Tasks and Stimulated Recall. Both groups completed two writing tasks: one administered before and one after the instructional period. For each task, students produced a brief argumentative essay (250–300 words) on socially relevant topics designed to provoke genuine cognitive engagement.

Immediately after each writing task, a subsample of 24 students (six per subgroup: high/low proficiency × experimental/control) participated in stimulated recall interviews. The decision to select six students from each subgroup was guided by methodological considerations of qualitative research, where smaller samples allow for in-depth exploration while ensuring representativeness across conditions (cf. Sasaki, 2000). This number also reflected practical constraints of time and resources in conducting and transcribing detailed recall sessions.

All writing sessions were video-recorded to capture observable behaviors (e.g., pausing, rereading, consulting notes). During the recall sessions, participants reviewed their own recordings and were prompted to verbalize their thoughts, decision-making, and strategy use. To minimize cognitive load and elicit more detailed responses, the interviews were conducted in Persian.

The interview protocol included six core open-ended questions. These questions targeted learners' planning, monitoring, and revision behaviors, as well as their perceptions of task difficulty and strategy usefulness. The questions were adapted from established stimulated recall protocols in L2 writing research (e.g., Sasaki, 2000; Wenden, 1991) and reviewed by two experts in applied linguistics to ensure content validity. The sessions were audio-recorded and subsequently translated and transcribed for analysis. Transcripts, along with observation notes, were coded for the occurrence and type of strategy use.

3.3.2. Learner Diaries and Strategy Use Questionnaire. Participants in the experimental group maintained reflective learner diaries throughout the intervention, documenting the strategies they attempted, their perceptions of the strategies' effectiveness, and any challenges they faced. To ensure comparability, students in the control group also kept diaries; however, they were instructed to reflect on their writing experiences in general, without specific reference to strategies. This design minimized the possibility that the reflective component itself would bias the results.

In addition, all participants (both control and experimental groups) completed a task-specific writing strategy questionnaire at the beginning and the end of the semester. The questionnaire was adapted from Oxford's (1990) Strategy Inventory for Language Learning and tailored to the writing context. Content validity was established through expert review by two applied linguistics specialists, and a pilot test confirmed its clarity and reliability. Cronbach's alpha for the adapted version was 0.82, indicating acceptable internal consistency. The inclusion of both groups in pre- and post-testing allowed for a meaningful comparison of changes in perceived strategy use.

3.4. Procedure

The experimental group received direct instruction in writing strategies over an 18-week period (two academic semesters). The instructional module was developed based on current L2 writing strategy taxonomies (e.g., Anderson, 2005; Macaro, 2006) and a pilot needs analysis. Drawing on Hayes' (2012) model, the training targeted three broad strategy categories—planning, self-monitoring, and revision—which were selected due to their foundational role in academic writing and their noted underuse during the pre-test phase. Instruction was delivered through bi-weekly 90-minute workshops that combined teacher modeling, group activities, task-based exercises, and guided reflections. In the initial six sessions, each strategy was introduced and modeled explicitly. In the subsequent sessions, strategy use was embedded into authentic writing tasks, during which students were encouraged to adapt strategies to task requirements and document their experiences in learner diaries.

The control group followed the same syllabus and completed identical writing tasks and assignments. Importantly, they also received regular instruction in academic writing, including essay structure, coherence and cohesion, grammar, and vocabulary. However, unlike the experimental group, they were not provided with explicit strategy instruction. To ensure comparability, both groups maintained reflective journals; the control group used them for general writing reflections, whereas the experimental group was guided to record their application of specific strategies. This design helped control for the possible influence of self-reflection activities on learning outcomes.

Both groups were taught by instructors with equivalent academic qualifications and teaching experience, and the control group's instructor was explicitly instructed to maintain consistency in syllabus coverage, instructional materials, and classroom management. The only difference between the groups was the presence or absence of explicit strategy training, which allowed for a meaningful comparison of the instructional effect.

3.5 Data Analysis

3.5.1. Quantitative Analysis. The quantitative analysis drew on two complementary data sources: (1) responses from the strategy-use questionnaire and (2) frequency counts of strategies identified in the stimulated recall transcripts. Questionnaire data were analyzed to assess pre- and post-test changes in perceived strategy use across groups and proficiency levels. Descriptive statistics and a mixed between-within ANOVA (time × group × proficiency) were employed to evaluate differences in strategy use patterns.

For the stimulated recall data, a coding framework adapted from Roca de Larios et al. (2008) and Sasaki (2000) was applied to classify strategy use into seven categories: Task Analysis, Planning, Formulating, Self-Monitoring, Revision, Evaluation, and Resourcing. While explicit instruction targeted only three categories (Planning, Self-Monitoring, and Revision), the broader framework allowed us to capture both taught and emergent strategies that learners spontaneously employed. This provided a more comprehensive picture of actual strategy use.

Two trained coders independently analyzed a subset of the transcripts. Inter-coder reliability, assessed with Cohen's Kappa ($\kappa = 0.85$), was reported in the qualitative analysis section rather than the quantitative results, in order to clarify its methodological role. This distinction ensured consistency with the description of instruments and avoided conflating questionnaire data with qualitative coding.

3.5.2. Qualitative Analysis. The qualitative data comprised three main sources: Stimulated recall interviews, learner diaries, and classroom video-taped observations. Unlike in the quantitative analysis—where stimulated recall transcripts were coded to obtain frequency counts of strategy use—here they were analyzed thematically to explore the *nature* and *quality* of learners' strategy application. This distinction ensured that the same dataset contributed to both quantitative breadth and qualitative depth without conceptual overlap.

Thematic analysis followed a recursive coding approach, focusing on learners' self-awareness, adaptability, and capacity to coordinate multiple strategies during writing. Particular attention was paid to episodes where students adjusted or reconfigured strategies in response to task-related challenges. Video-taped observations, which had captured learners' on-task behaviors (e.g., pausing, consulting notes, rereading drafts), were integrated into this analysis to provide an external perspective on the cognitive processes reported during stimulated recall.

To enhance the credibility and trustworthiness of the findings, methodological triangulation was employed, combining evidence from questionnaires, recall interviews, diaries, and observational notes. This integration allowed us to identify points of convergence across data sources, thereby strengthening the interpretation of learners' strategic development.

4. Results

This section presents the findings of the study based on the quantitative and qualitative data collected before and after the intervention. Results are reported in relation to changes in

writing strategy use among learners across different proficiency levels and experimental conditions.

4.1. Quantitative Findings

Analysis of strategy usage during the writing tasks indicated a clear increase in both the number and range of strategies employed by the experimental group after the instructional intervention. In contrast, the control group showed no substantial change in their strategy use from the pre-test to the post-test. These findings suggest the intervention had a notable impact on strategic writing behavior.

4.1.1. Strategy Use across Groups. To investigate whether the experimental and control groups differed in their use of writing strategies across time, a mixed between-within subjects ANOVA was conducted with time (pre-test vs. post-test) as the within-subjects factor and group (experimental vs. control) as the between-subjects factor. Results revealed a significant main effect for time, $F_{(1, 62)} = 56.41$, $p < 0.001$, partial $\eta^2 = .48$, indicating that overall strategy use increased from pre- to post-test. A significant main effect was also found for group, $F_{(1, 62)} = 29.17$, $p < 0.001$, partial $\eta^2 = .32$, with the experimental group employing more strategies overall than the control group. Most importantly, the time × group interaction was significant, $F_{(1, 62)} = 42.06$, $p < 0.001$, partial $\eta^2 = .40$, showing that the experimental group experienced a far greater increase in strategy use compared with the control group. Table 1 presents the descriptive statistics for each group across the two time points, while Gain scores were calculated by subtracting the mean pre-test score from the mean post-test score for each group.

Table 1.

Descriptive Statistics for Strategy Use by Group

Group	Pre-Test (M, SD)	Post-Test (M, SD)	*Gain Score*
Experimental	4.1 (1.3)	8.5 (1.9)	+4.4
Control	4.0 (1.2)	4.4 (1.4)	+0.4

4.1.2. Proficiency-Level Comparisons. To explore whether learners' proficiency levels moderated the effects of strategy instruction, a two-way mixed ANOVA was conducted with time (pre-test vs. post-test) as the within-subjects factor, and group (experimental vs. control) and proficiency (high vs. low) as between-subjects factors. Results showed a significant main effect for group, $F_{(1, 60)} = 34.72$, $p < 0.001$, partial $\eta^2 = .37$, confirming that the experimental group outperformed the control group overall. However, the time × group × proficiency interaction was not statistically significant, $F_{(1, 60)} = 2.13$, $p = 0.15$. This suggests that the benefits of instruction were broadly consistent across proficiency bands.

In table 2, Gain scores were calculated by subtracting individual pre-test scores from post-test scores, then averaging across learners within each proficiency subgroup. Moreover, descriptive analyses of gain scores within the experimental group revealed interesting trends: Higher-proficiency learners appeared to demonstrate greater improvement in self-monitoring and evaluation strategies, whereas lower-proficiency learners showed larger gains in planning and task analysis. Although these differences did not reach statistical significance, they provide insight into how learners at different proficiency levels may differentially internalize and apply strategy instruction.

Table 2.

Strategy Gain Scores by Proficiency Level (Experimental Group)

Strategy Type	High Proficiency (Gain)	Low Proficiency (Gain)
Planning	+2.1	+3.3
Self-Monitoring	+2.7	+1.8
Revision	+1.5	+1.4
Evaluation	+2.9	+1.6
Total	+9.2	+8.1

4.2. Qualitative Findings

Data from stimulated recall interviews and learner diaries offered deeper insight into how learners conceptualized and applied strategies during writing.

4.2.1. Increased Strategic Awareness. Prior to the instructional intervention, most participants tended to focus on a limited range of strategies—like translating word-for-word or just running through basic grammar corrections. After the intervention, though, those in the experimental group reported a much greater awareness of the overall structure of the task, the intended audience, and the rhetorical aims behind their writing. In other words, their approach shifted from surface-level edits to a deeper engagement with purpose and context. *For example,* Student 14 (low proficiency, experimental group) noted that: *"In the past, I would simply write and hope for the best; now, I pause to consider my message and how to express it clearly."*

4.2.2. Emergence of Flexible Strategy Use. Several students demonstrated strategic flexibility—the ability to switch between planning and revision strategies depending on the demands of the task. For instance, Student 6 (high-proficiency learner, experimental group) argued that: *"I usually write everything first, then I fix mistakes. Now I plan more and revise in stages, especially when I'm not sure how to express complex ideas."*

4.2.3. Metacognitive Reflections. Entries from learner diaries revealed that many students began reflecting on their own writing process. Some low-proficiency learners indicated they

had never thought of writing as a process before the workshops. As an example, Student 22 (low proficiency, experimental group) expressed that: *"I didn't realize I could take control of my writing; now I feel that I have a method, not just an outcome."*

5. Discussion

This investigation aimed to determine the impact of explicit instruction in writing strategies on learners' academic writing practices across various proficiency levels. The findings corroborated the earlier research indicating that targeted strategy training enhances learners' metacognitive regulation and strategic engagement (see Anderson, 2005; Sasaki, 2002). More specifically, the study demonstrated that learners at different stages of proficiency benefit in distinct yet complementary ways from explicit instruction. Higher-proficiency learners gained more in areas such as self-monitoring and evaluation, while lower-proficiency learners showed greater improvements in planning and task analysis. These outcomes suggest that strategy training can be tailored to proficiency levels in order to maximize effectiveness for diverse groups of learners.

The significant increase in the usage of strategies from the experimental group as compared to the control group indicates the potential of explicit teaching in fostering strategic composition behavior. The finding confirms previous studies which asserted that strategies in composition are not innate but can be taught and practiced if properly scaffolded (Manchón, 2008; Plonsky, 2011). The shift from heavy usage of mechanical or surface strategies towards a more balanced usage of planning, self-monitoring, and revision indicates the growing control over the composition procedure by the students.

Interestingly, both the more proficient and the less proficient experimental group students exhibited significant improvement in the utilization of strategies, though in non-overlapping areas. More proficient students exhibited a bigger growth in strategies involving self-monitoring and evaluation, which indicates the superior potential for real-time self-reflection and correction in the production process. This finding corresponds with the notion by Grabe and Zhang (2023) regarding more proficient students benefiting the most from the teaching of metacognitive strategies owing to the wider cognitive bandwidth and linguistic versatility they possess.

In comparison, low proficiency students showed significant improvement in planning and task analysis. Such results indicate that if the instruction is adequately contextualized, even students with minimal linguistic resource can internalize and transfer strategic behavior in the way of scaffolding their own writing. As Teng (2024) maintained, beginners and

intermediate students can construct strategic awareness with adequate models and repeated practice.

Qualitative data persisted in substantiating the results. Stimulated recall interviews and learner diaries revealed students had become more conscious about the choices they made in their writing and began performing tasks more intentionally. Prior to the intervention, the majority of students linearly composed from introduction through the conclusion with minimal rewriting and reflection. Data collected after the intervention revealed growing awareness on the recursive nature of the act of writing and the need for coming back at earlier levels in response to growing task demands.

This growth in employment of strategies coincides with Kellogg's (1996) description of working memory in composition in which the blending of the processes for planning, formulation, and revision is emphasized. The growth in adaptive employment of strategies most prominently in the group of high proficiency similarly underpins the most recent study by Horbach and Giannakaki (2023), which asserts flexibility as the distinguishing feature of composition expertise.

6. Conclusion

This study investigated the effect of explicit writing strategy instruction on EFL learners' use of cognitive and metacognitive strategies within an academic writing context, with particular attention to proficiency-level differences. The results demonstrated that sustained, targeted instruction significantly enhanced learners' ability to plan, monitor, and revise their writing, thereby fostering greater self-regulation and reflective engagement with the writing process.

Although no statistically significant interaction between group and proficiency level were obtained, descriptive trends suggested that higher-proficiency learners tended to improve more in strategies associated with self-monitoring and evaluation, whereas lower-proficiency learners gained more in planning and task-analysis strategies. These patterns, while not conclusive, indicate that learners may internalize and apply strategy instruction in ways shaped by their initial proficiency.

From a curricular perspective, the findings support the integration of writing strategy instruction as a core component of L2 writing courses rather than as supplementary material. Encouraging strategic awareness through tools such as reflective journals and stimulated recall can strengthen learner autonomy and long-term competence in academic writing.

These findings have several pedagogical implications. First, they may highlight the necessity of explicit instruction in strategy use, particularly in L2 writing contexts where learners are expected to produce cognitively demanding texts. Second, the results may suggest that differentiated instruction—that is, instruction responsive to learners' initial proficiency levels—may yield meaningful gains across a broad spectrum of learners. Lower proficiency students may benefit from explicit modeling and guided practice in planning and organizing ideas, while more advanced learners may require support in monitoring and refining their textual output.

Moreover, the success of stimulated recall and reflective journaling in revealing learners' strategic behavior underscores the value of integrating metacognitive reflection into writing curricula. Such practices not only make strategy use visible to learners themselves but also encourage the development of learner autonomy.

Despite the promising results, several limitations should be acknowledged. The sample size, though sufficient for statistical analysis, was restricted to one institution, limiting the generalizability of the findings. Future studies might expand the scope to include multiple institutions or compare outcomes across different educational systems.

Additionally, the study focused primarily on short-term gains in strategy use. Longitudinal research is needed to determine whether the observed strategic behaviors are sustained over time and transferred to other writing contexts, such as academic assignments or standardized assessments. Future work could also explore the intersection of writing strategy instruction with other learner variables such as motivation, self-efficacy, or anxiety (Zhang & Teng, 2023).

References

Anderson, N. J. (2005). *L2 strategy research*. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 757–771). Lawrence Erlbaum.

Flower, L., & Hayes, J. R. (1980). The cognition of discovery: Defining a rhetorical problem. *College Composition and Communication*, 31(1), 21–32. https://doi.org/10.2307/356630

Gass, S., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Lawrence Erlbaum Associates.

Grabe, W., & Zhang, C. (2023). Revisiting strategic L2 writing: A multidimensional model of development. *Journal of Second Language Writing*, 63, 101023. https://doi.org/10.1016/j.asw.2022.100690

Graham, S. (1997). Effective language learning: Positive strategies for advanced level language learning. *Modern Language Journal, 81*(4), 564–567.

Horbach, J., & Giannakaki, M. (2023). Differentiated instruction and strategic flexibility in academic writing. *System*, *112*, 102976. https://doi.org/10.1016/j.ssaho.2024.101230

Ikeda, M., & Takeuchi, O. (2003). Can strategy instruction help EFL learners? An empirical study. *Language Learning Journal*, *31*(1), 37–46.

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing* (pp. 57–71). Lawrence Erlbaum.

Macaro, E. (2006). Strategies for language learning and for language use: Revising the theoretical framework. *Modern Language Journal*, *90*(3), 320–337. https://doi.org/10.1111/j.1540-4781.2006.00425.x

Macaro, E. (2010). Continuum companion to second language acquisition. Continuum.

Manchón, R. M. (2008). Writing strategies in second language acquisition. In R. M. Manchón (Ed.), *Learning, teaching, and researching writing in foreign language contexts* (pp. 229–256). Multilingual Matters.

Oxford, R. (1990). *Language learning strategies: What every teacher should know*. Newbury House.

Plonsky, L. (2011). The effectiveness of second language strategy instruction. *Language Learning*, 61(4), 993–1038. https://doi.org/10.1177/13621688211072981

Roca de Larios, J., Manchón, R., & Murphy, L. (2008). The foreign language writer's strategic behavior: A qualitative approach. In R. M. Manchón (Ed.), *Learning, teaching and researching writing in foreign language contexts* (pp. 141–164). Multilingual Matters.

Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing*, *9*(3), 259–291. https://doi.org/10.1016/S1060-3743(00)00028-X

Sasaki, M. (2002). Building an empirically-based model of EFL learners' writing processes. *Language Learning*, *52*(1), 1–35. https://doi.org/10.1007/978-94-010-0363-6_3

Teng, L. S. (2024). Teaching writing strategies to EFL learners: The role of explicitness and reflection. *TESOL Quarterly*, *58*(1), 115–140.

Teng, L. S., & Zhang, L. J. (2021). Empowering learners in the EFL classroom: The role of strategy-based instruction. *Language Teaching Research*, *25*(5), 690–712.

Vandergrift, L., & Tafaghodtari, M. H. (2010). Teaching L2 learners how to listen does make a difference. *Language Learning*, *60*(2), 470–497. https://doi.org/10.1111/j.1467-9922.2009.00559.x

Wharton, S. (2000). Language learning strategy use of bilingual foreign language learners in Singapore. *Language Learning*, *50*(2), 203–243. https://doi.org/10.1111/0023-8333.00117

Zhang, L. J., & Teng, L. S. (2023). Researching writing strategies in multilingual contexts: A methodological review. *System*, *113*, 103016. https://doi.org/10.1016/j.jslw.2023.101038

Curriculum Research
Volume 6, Issue 3
Sept. 20, 2025

The Role of Reflective Practice in Iranian EFL Teachers' Professional Development and Classroom Effectiveness

Article Type:

Original Research

Authors:

Mehrdad Rezaee[1]
ORCiD: 0000-0002-5604-3064

Saeed Sayyad Chamani[2]

Mohsen Noroozieh[3]

Article History:

Received: 2025.06.09
Accepted: 2025.09.18
Published: 2025.09.20

Abstract

Reflective practice is widely recognized as a cornerstone of teacher professional growth and effective classroom teaching. This study investigated its impact on Iranian EFL teachers' development and their students' learning outcomes, specifically focusing on engagement and academic achievement. A nine-month longitudinal mixed-methods study was conducted with 30 Iranian EFL teachers (aged 24–50) selected based on snow ball sampling technique and 300 students (mixed gender, aged 12–18, intermediate proficiency) taught by those teachers from Safir Language Academy in Tehran. Teachers engaged in structured reflective activities, including journaling, peer discussions, and guided reflection sessions. Data were collected through reflective journals, semi-structured interviews, and student performance assessments, and were analyzed using thematic analysis and paired samples *t*-tests. The findings revealed that reflective practice enhanced teachers' self-awareness and adaptability in instructional strategies, while students of reflective teachers demonstrated higher engagement and improved academic achievement. Barriers such as time constraints and limited institutional support were also identified. Overall, the study highlights the transformative potential of reflective practice in advancing teacher professional development and promoting more effective EFL classrooms in the Iranian context.

Key Words: Reflective Practice, Teacher Development, Classroom Effectiveness, Iranian EFL Teachers

1. Department of English, CT.C., Islamic Azad University, Tehran, Iran (Corresponding Author). Email: Dr.mehrdadrezaee@iau.ac.ir
2. Safir Language Academy, Tehran, Iran. Email: saeedsayyadchamani@gmail.com
3. Safir Language Academy, Tehran, Iran. Email: mohsennoroozieh@gmail.com

1. Introduction

Reflective practice is deemed as a transformative tool in teacher professional development and classroom effectiveness. It involves the systematic analysis of teaching experiences with the aim of improving pedagogical strategies and fostering professional growth (Farrell, 2018; Loughran, 2020). Unlike routine self-evaluation, reflective practice requires critical engagement with one's instructional decisions, classroom interactions, and learning outcomes (Larrivee, 2000). This continuous process enhances teachers' self-awareness, adaptability, and evidence-based teaching strategies, making it particularly valuable in the dynamic contexts of modern classrooms (Farrell, 2015a).

In the field of Teaching English as a Foreign Language (TEFL), reflective practice has gained special importance because teachers should constantly adapt their methods to address linguistic diversity and cultural differences (Chaika, 2023). Research has shown that teachers who engage in structured reflection develop stronger classroom management skills, demonstrate higher adaptability, and adopt more student-centered approaches (Boud et al., 2020). However, the application of reflective practice in different educational contexts is uneven, with Iranian EFL classrooms remaining underexplored (Farrell, 2022; Kurosh et al., 2020).

Iranian teachers face distinctive institutional and cultural constraints, including hierarchical school structures, limited access to professional development opportunities, and time pressures, all of which hinder systematic reflection (Ashraf et al., 2016; Moradkhani et al., 2017). While Western studies have consistently highlighted the benefits of reflection for teacher growth and student engagement (Darling-Hammond et al., 2019; Zembylas, 2018), fewer studies have empirically tested its effectiveness in Iranian contexts. As a result, reflective practice is often discussed theoretically rather than evaluated through measurable outcomes such as student engagement and academic achievement.

The present study sought to address this gap by empirically examining the role of structured reflective activities in Iranian EFL classrooms. Specifically, it investigated how reflection influenced teachers' self-awareness, instructional adaptability, and classroom practices, as well as how it impacted students' engagement and achievement. The study also explored barriers that might prevent Iranian EFL teachers from sustaining reflective practices, including institutional and cultural challenges, and considered potential strategies for overcoming them.

By combining qualitative and quantitative evidence, this study aimed to provide a clearer understanding of the transformative potential of reflective practice in Iranian EFL contexts. It

contributes to both global debates on teacher development and to the local improvement of classroom effectiveness by aligning reflective teaching with measurable outcomes.

Accordingly, this research pursued the following objectives:

1. To evaluate the impact of reflective practice on Iranian EFL teachers' self-awareness, instructional adaptability, and classroom strategies.
2. To investigate the relationship between teacher reflection and student engagement and achievement.
3. To identify barriers that hinder Iranian EFL teachers from adopting reflective practices and propose practical solutions.

In line with these objectives, the study addressed the following three research questions:

1. How does engaging in reflective practice influence Iranian EFL teachers' self-awareness, adaptability, and instructional strategies?
2. What is the impact of reflective practice on students' engagement and academic achievement in Iranian EFL classrooms?
3. What challenges do Iranian EFL teachers face in sustaining reflective practice, and what strategies can support its effective implementation?

By examining these factors, this study attempted to provide evidence-based recommendations for integrating reflective practice into teacher education and professional development programs in Iranian EFL contexts.

2. Review of the Related Literature

Reflective practice has been widely deemed as the key element of teacher development, particularly within the field of English language teaching (ELT). Since Schön's (2017) seminal work on the "reflective practitioner", scholars have emphasized the value of teachers critically examining their own practices to improve pedagogical decision-making. Wallace (1991) further formalized this perspective, framing reflection as an essential stage in the "reflective model" of teacher education. Building on these foundations, Farrell (2018) argued that reflection is not merely an individual activity but a dialogic and collaborative process that supports professional identity construction.

Research has consistently highlighted the role of reflective practice in enhancing teacher professional development. Akbari (2007) identified reflection as a key dimension of effective teacher education programs, noting that reflective teachers are more likely to engage in continuous self-assessment and adapt their methodologies to meet learner needs. Similarly,

Richards and Lockhart (1995) demonstrated that structured reflection enables teachers to recognize and adjust habitual teaching patterns, ultimately leading to greater classroom responsiveness. Reflection is also linked to broader constructs of teacher cognition, including beliefs, knowledge, and decision-making processes, which shape how teachers interpret and act upon classroom realities (Borg, 2006, 2011).

A related body of literature connects reflective practice with classroom effectiveness. Farrell (2015b) contends that reflective teachers develop greater instructional adaptability, which translates into more effective classroom management and improved learner engagement. Other studies have shown that reflection helps teachers incorporate student feedback, experiment with varied techniques, and integrate contextually appropriate strategies (Farrell, 2018; Farrell & Macapinlac, 2021; Richards & Lockhart, 1995). These findings suggest that reflection directly contributes to teachers' ability to create more supportive and interactive learning environments. Furthermore, reflective practice has been found to nurture teachers' emotional resilience and professional confidence, qualities that are increasingly recognized as central to sustaining long-term effectiveness (Day & Gu, 2010).

More recently, attention has turned to the impact of reflection on student outcomes. Research indicates that when teachers engage in systematic reflection, students benefit not only from improved instructional clarity but also from heightened motivation and academic achievement (Akbari, 2007; Farrell, 2018). Student outcomes in reflective practice research are commonly measured through standardized proficiency tests, classroom-based formative assessments, and end-of-term course grades (Darling-Hammond et al., 2019; Farrell, 2022). Formative assessments capture students' ongoing engagement and learning progress by evaluating quizzes, writing assignments, and classroom activities, while course grades provide a cumulative indicator of achievement at the end of a term. These measures offer a comprehensive understanding of both the process and the product of student learning (Guskey, 2003).

Taken together, the literature suggests that reflective practice not only strengthens teachers' professional growth but also enhances classroom effectiveness and student performance (Timperley et al., 2007). However, challenges such as institutional barriers, time constraints, and cultural norms may limit teachers' ability to engage in sustained reflection (Akbari, 2007; Farrell, 2018). These tensions underscore the importance of investigating reflective practice within specific contexts, such as Iranian EFL classrooms, where unique cultural and institutional dynamics may shape both teacher development and student outcomes. Despite the recognized benefits of reflection, relatively few empirical studies in the Iranian setting have systematically examined its dual impact on teachers and learners, leaving a gap that the present study seeks to address (Borg, 2011).

This research holds both theoretical and practical significance. Theoretically, it builds on sociocultural perspectives such as Vygotsky's (1978) Zone of Proximal Development, framing reflective practice as a scaffolding mechanism that helps teachers critically evaluate and refine their instructional strategies within their social and cultural contexts (Lantolf & Poehner, 2018). Practically, it offers evidence-based recommendations for integrating structured reflection into teacher training programs, including tools such as journaling, peer discussions, and guided reflection sessions.

These approaches can improve teachers' self-awareness and adaptability, while also fostering collaborative professional development through peer coaching and mentoring models. For policymakers and administrators, the study highlights the importance of institutional support—such as dedicated time, resources, and collaborative platforms—for embedding reflection into EFL teacher education.

3. Method

3.1. Design

This study employed an explanatory sequential mixed-methods design (Creswell & Plano Clark, 2018) to examine the impact of reflective practice on Iranian EFL teachers' professional development and student outcomes. The design consisted of an initial quantitative phase (pre- and post-tests of student achievement) followed by a qualitative phase (teacher journals and interviews) to provide deeper insights into the quantitative results.

3.2. Participants

The study involved two groups of participants: teachers and students. There were thirty Iranian EFL teachers (aged 24–50) from different branches of Safir Language Academy in Tehran recruited through snowball sampling. They represented three experience levels: novice (less than 5 years, 40%), experienced (5–10 years, 35%), and veteran (over 10 years, 25%). Eligibility criteria included at least two years of teaching experience and willingness to engage in structured reflective activities. In addition, 300 students (aged 12–18; 55% female, 45% male) taught by the participating teachers also took part. They were placed at intermediate proficiency level according to the Oxford Quick Placement Test (OQPT). Inclusion criteria required regular classroom attendance and participation in pre- and post-testing.

All participants provided informed consent prior to data collection. Teachers' and students' identities were anonymized. Participation was voluntary, and data were stored securely.

3.3. Instruments

Three complementary instruments were used to collect data. Reflective journals were maintained weekly by teachers for nine months, documenting teaching challenges, instructional decisions, and classroom interactions, and serving as the primary source of qualitative data on teachers' evolving self-awareness and adaptability. Semi-structured interviews were conducted with all 30 teachers in three rounds (early, mid, and late stages of the study), each lasting 30–45 minutes, conducted in Persian, and audio-recorded with participants' consent. Interview prompts focused on teachers' perceptions of reflective practice, instructional strategies, and barriers to sustained reflection, and transcripts were anonymized and returned to participants for member checking. Finally, student outcomes were evaluated using three complementary measures to capture both language proficiency and classroom achievement:

1. **Oxford Quick Placement Test (OQPT):**

   This standardized proficiency test was administered to all students at the beginning and end of the study. It provided a reliable benchmark of overall language proficiency gains.

2. **Formative Assessments:**

   These included classroom-based tasks such as short quizzes, written assignments, and in-class activities administered throughout the nine-month study. Formative assessments reflected students' ongoing engagement and progress in the learning process. Scores were systematically recorded by teachers as part of their routine evaluation practices.

3. **End-of-term Grades:**

   Final grades assigned by the institution at the end of each course cycle served as a cumulative measure of student achievement. These grades aggregated performance across examinations, assignments, and classroom participation, providing a summative indicator of overall learning outcomes.

Together, these measures offered a multi-dimensional picture of student learning by combining standardized testing, ongoing engagement indicators, and institutional achievement records.

3.4. Data Collection Procedures

Data were collected in three phases over nine months:

1. **Pre-Intervention (Month 1)**: Teachers completed baseline reflective journals and interviews; students took the OQPT and initial classroom assessments.
2. **Intervention (Months 2–7)**: Teachers engaged in structured reflective activities: (a) weekly journals, (b) biweekly peer reflection groups, and (c) monthly guided reflection sessions facilitated by the researchers. Classroom observations were conducted twice per teacher (n = 60 observations) to document reflective practices in action.
3. **Post-Intervention (Months 8–9)**: Teachers completed final journals and interviews; students repeated OQPT and classroom assessments.

3.5. Data Analysis

- **Qualitative Data (Journals & Interviews)**: Analyzed thematically using Braun & Clarke's (2006) framework with NVivo software. Two coders independently coded the data; inter-rater reliability ($\kappa = 0.81$) indicated strong agreement. Themes included self-awareness, instructional adaptability, and barriers to reflection.
- **Quantitative Data (Student Assessments)**: Pre- and post-test scores were compared using paired-samples t-tests to measure gains in student performance. Cohen's d was calculated for effect sizes. Correlation analysis examined the relationship between frequency/depth of teacher reflection and student outcomes.
- **Integration**: Results from the two strands were merged in the interpretation stage, with qualitative insights used to explain quantitative patterns.

4. Findings

This section presents the findings according to the three research questions. Quantitative and qualitative data were combined to demonstrate the impact of reflective practice on teachers' professional development and student outcomes.

4.1 Research Question One

4.1.1 Quantitative Findings. Analysis of journal entries revealed consistent growth in teachers' reflective practice over the nine-month period. As shown in Figure 1, the number of entries increased steadily across the study period. Coding of journals identified three recurring themes—self-awareness, instructional adaptability, and collaboration—whose frequency increased substantially from pre- to post-intervention (Table 1).

Table 1.

Frequency of Reflective Themes across Phases

Theme	Pre-Intervention	Mid-Intervention	Post-Intervention
Self-awareness	40	65	90
Instructional Adaptability	35	60	85
Collaboration	20	35	60

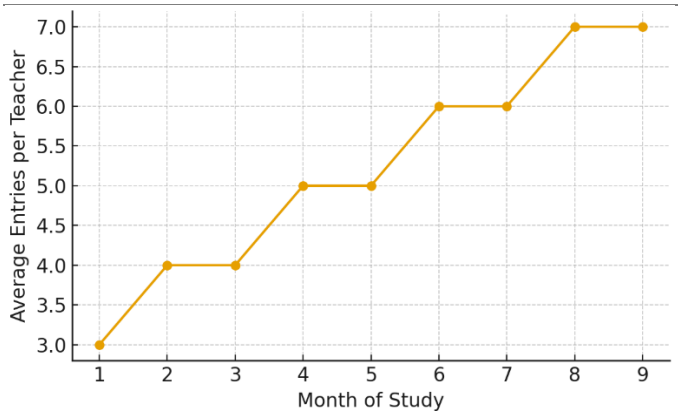


Figure 1.

Growth in reflective journal entries over nine months

4.1.2 Qualitative Findings. The thematic analysis highlighted three dominant areas of teacher development:

1. **Improved Self-Awareness** – Teachers became more conscious of their teaching behaviors. One of the teachers stated:

   "Writing journals forced me to reflect on my feedback style. I realized I was too general and started giving specific comments." (Teacher 12)

2. **Instructional Adaptability** – Teachers increasingly modified strategies in response to student needs. One of the teachers explained that:

   "After reflecting on disengagement, I introduced pair work. It was remarkable to see the students' interaction improved." (Teacher 7)

3. **Collaboration** – Peer reflection groups encouraged sharing and experimentation with new methods. One of the teachers noted that:

   "Listening to colleagues helped me rethink classroom discipline. I tried their approach and it worked better." (Teacher 19)

In conclusion, it can be said that engaging in structured reflective practice produced demonstrable growth in teachers' professional capacities, significantly enhancing self-awareness, prompting adaptive instructional changes, and increasing collaborative problem-

solving as evidenced by rising frequencies of reflective themes and supporting interview data.

4.2. Research Question Two

4.2.1. Quantitative Findings. Students demonstrated significant improvements across all three measures: OQPT scores, formative assessments, and end-of-term grades. As displayed in Table 2, paired-samples t-tests confirmed statistically significant gains ($p < .001$), with moderate to large effect sizes. Figure 2 visually illustrates the pre to post improvements across all measures.

Table 2.

Student Performance Pre- and Post-Intervention

Measure	Pre-Intervention (M, SD)	Post-Intervention (M, SD)	t-value	p-value	Cohen's d
OQPT Scores	75.4 (10.2)	82.6 (9.8)	4.35	< .001	0.70
Formative Assessments	68.3 (12.5)	77.1 (11.4)	4.10	< .001	0.72
End-of-term Grades	70.5 (11.3)	79.8 (10.7)	4.25	< .001	0.76

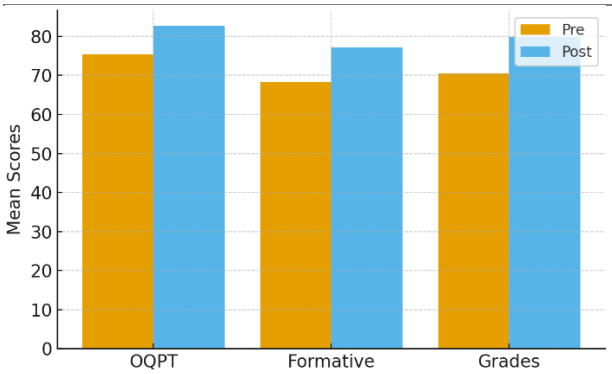


Figure 2.

Pre- and post-intervention student performance

Correlation Analysis. Further analysis revealed strong positive correlations between the frequency of teacher reflection (journal entries and peer group participation) and student performance gains ($r = .62$, $p < .01$), as well as between the depth of reflection (coding complexity) and student performance ($r = .58$, $p < .01$). These findings are presented in Table 3. A scatter plot of reflection frequency and student gains (Figure 3) illustrates this relationship.

Table 3.

Correlation between Reflection and Student Performance

Variable	Correlation Coefficient (r)	p-value
Reflection frequency	0.62	< .01
Reflection depth	0.58	< .01

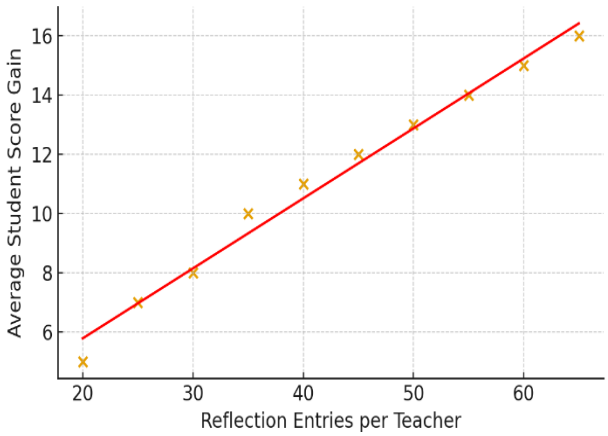


Figure 3.

Correlation between reflection frequency and student gains

4.2.2. Qualitative Findings. Teachers noted higher levels of student engagement and achievement following reflective adjustments to teaching strategies. For instance, one of the teachers remarked that "*Before reflecting, I thought my students were just not interested. After adding interactive tasks, their participation soared."*(Teacher 4). Another teachers also expressed that "*Students became more responsive when I used real-life scenarios after reflecting on their needs."* (Teacher 22)

In summary, the finding of the second research question revealed that reflective practice had a clear, measurable positive impact on students: learners of reflective teachers showed statistically significant gains on the OQPT, formative assessments, and end-of-term grades ($p < .001$; Cohen's $d \approx 0.70–0.76$), with strong positive correlations between teacher reflection and student performance gains.

4.3. Research Question Three

4.3.1. Quantitative Findings. Teachers reported three main barriers: time constraints (73%), lack of institutional support (60%), and cultural barriers (40%). These results are summarized in Table 4 and illustrated in Figure 4.

Table 4.

Teacher-Reported Barriers to Sustained Reflection

Barrier	Frequency	Percentage
Time constraints	22	73%
Lack of institutional support	18	60%
Cultural barriers	12	40%

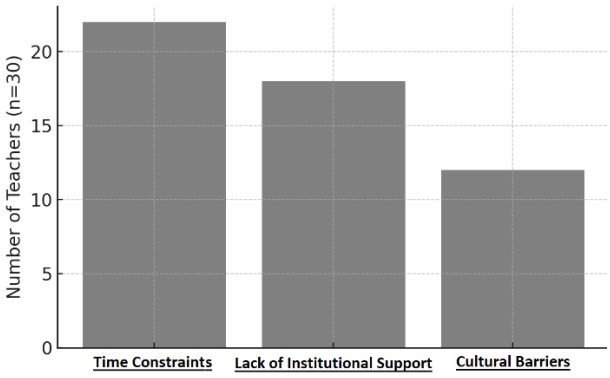


Figure 4.

Reported barriers to sustained reflection

4.3.2. Qualitative Findings. Interview data reinforced these challenges as can be seen from the following interview extracts. One of the teachers explained that *"Between lesson planning and grading, I struggled to keep up with journaling"* (Teacher 6). Or another teacher stated that, *"Some schools do not encourage peer reflection, so it was difficult to sustain discussions"* (Teacher 15). Also one teacher expressed that "*Open reflection is sometimes seen as criticism of authority, so teachers hesitate to share honestly"* (Teacher 27).

Teachers suggested strategies to address these barriers, such as institutional support for reflective activities, formal inclusion of reflection in training curricula, and anonymous platforms for sharing sensitive reflections.

To conclude, the results of the third research question indicated that although reflective practice yields clear benefits, its sustainability is constrained mainly by time pressures, limited institutional support, and cultural barriers. Nevertheless these challenges can be alleviated by formal institutional backing (allocated reflection time, curricular inclusion) and safe/collaborative platforms for teacher reflection, as proposed by the participants.

5. Discussion

The findings of this study confirm the central role of reflective practice in promoting teachers' professional development and improving student outcomes. With regard to the first research question, teachers who engaged in systematic reflection reported gains in self-awareness, instructional adaptability, and collaborative learning. These results are consistent with the findings of Farrell (2018), who emphasized that reflection strengthens teacher identity and enhances decision-making in the classroom. Similarly, Schön's (1992) notion of reflection-in-action is evident in the way teachers adjusted strategies during lessons, while Kolb's (2015) experiential learning cycle is reflected in their iterative process of planning, acting, reflecting,

and revising. The present findings thus extend earlier work by showing that these theoretical models remain relevant in the Iranian EFL context, where reflective practices are often less institutionalized.

The second research question concerned the impact of reflective teaching on students. Quantitative data revealed significant gains in engagement and academic achievement, aligning with Richards and Lockhart (1995), who argued that reflective teachers create more responsive classrooms. Similarly, Darling-Hammond et al. (2019) found that reflective practice enhances classroom management and fosters higher student achievement. The strong correlation observed between teachers' frequency of reflection and students' performance provided empirical support for Bera & Mohalik (2015), who reported that reflective journals contribute to more effective teaching strategies. However, while most international studies report qualitative evidence of improved student outcomes, the present study added quantitative confirmation within a non-Western EFL context. This could address the gap noted by Farrell (2022), who emphasized that Iranian research on reflection rarely includes robust, longitudinal, or mixed-methods data.

At the same time, the results diverge from some earlier findings. For example, Akbari (2007) argued that reflective practice in Iran often remains superficial due to cultural and institutional constraints. While barriers were indeed identified in this study, the significant improvements in both teacher and student outcomes suggested that structured and sustained reflection, when supported by guided activities, can overcome some of these challenges. In this sense, the study demonstrated that reflection can be effectively implemented in hierarchical educational contexts if institutional backing and facilitation are provided.

The third research question highlighted barriers to sustaining reflection, with time constraints, limited institutional support, and cultural norms being the most frequently reported ones. These findings resonate with Moradkhani et al. (2017), who also documented teachers' concerns about workload and lack of support. They also parallel Farrell's (2015b) observation that reflective practice requires a supportive environment to move beyond individual experimentation. Unlike some previous studies, however, teachers in the present research proposed specific solutions, such as embedding reflection into official teacher education curricula and offering institutional time for reflection. These recommendations underscore the practical importance of aligning reflective initiatives with systemic reforms.

Taken together, these findings contribute to ongoing debates about the value and feasibility of reflective practice in teacher education. They suggest that reflective practice should not be regarded solely as an individual cognitive process, but rather as a socially mediated and institutionally supported activity (Vygotsky, 1978). When structured reflection is

integrated into teacher development programs, it not only advances professional growth but also produces measurable gains in student engagement and achievement.

6. Conclusion

This study has demonstrated that reflective practice, when implemented systematically and supported institutionally, can play a pivotal role in enhancing teacher development and improving student outcomes in Iranian EFL classrooms. Teachers who engaged in structured reflection developed stronger self-awareness, greater adaptability in instructional strategies, and more collaborative approaches to teaching. Significantly, students of these teachers also exhibited higher engagement and significant gains in academic achievement, underscoring the direct pedagogical impact of reflection.

The findings contribute to the wider literature by providing quantitative evidence from a non-Western EFL context, where reflective practice has traditionally been underexplored. While much of the existing scholarship (e.g., Darling-Hammond et al., 2019; Farrell, 2018; Richards & Lockhart, 1995) has emphasized the benefits of reflection, this study extended the field by offering longitudinal, mixed-methods data that illustrated its transformative effects in Iran. At the same time, the challenges identified—time limitations, lack of institutional support, and cultural barriers—highlighted the need for systemic solutions that move reflection beyond individual initiative.

Implications for practice include the integration of reflection into teacher education curricula, the provision of institutional time and resources for reflective activities, and the encouragement of collaborative reflective communities among teachers. Policymakers and school administrators should recognize that reflective practice is not merely a personal habit but a professional competency that requires formal support. Future research might investigate the role of digital tools and AI-assisted platforms in sustaining reflective practices, particularly in contexts where time and workload pressures hinder traditional approaches. Longitudinal comparative studies across cultural contexts would further illuminate how reflection can be adapted to diverse educational systems. In conclusion, reflective practice emerges not only as a means of individual professional growth but also as a strategic approach to raising teaching quality and student achievement in EFL education.

References

Akbari, R. (2007). Reflections on reflection: A critical appraisal of reflective practices in L2 teacher education. *System, 35*(2), 192–207. https://doi.org/10.1016/j.system.2006.12.008

Ashraf, H., Samir, A., &, Yazdi, M. (2016). Reflective teaching Practice in an EFL context: A qualitative study. *International Journal of English Linguistics. 6*, 48-48. DOI:10.5539/ijel.v6n7p48

Bera, S., & Mohalik, R. (2015). Reflective practice in professional development of teachers.

Borg, S. (2006). *Teacher cognition and language education: Research and practice.* London: Continuum.

Borg, S. (2011). The impact of in-service teacher education on language teachers' beliefs. *System, 39*(3), 370–380. https://doi.org/10.1016/j.system.2011.07.009

Boud, D., Keogh, R., & Walker, D. (2020). *Reflection: Turning experience into learning.* Routledge.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chaika, O. (2023). Role of reflective practice in foreign language teaching and learning via multicultural education. *International Journal of Social Science and Human Research, 6*(2), 1343–1350. https://doi.org/10.47191/ijsshr/v6-i2-74

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed-methods research* (3rd ed.). Thousand Oaks, CA: SAGE.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2019). *Effective teacher professional development.* Learning Policy Institute.

Day, C., & Gu, Q. (2010). *The new lives of teachers.* London: Routledge.

Farrell, T. S. C. (2015a). *Promoting Teacher Reflection in Second Language Education: A Framework for TESOL Professionals.* Routledge.

Farrell, T. S. C. (2015b). *Reflective language teaching: From research to practice.* London, New York: Bloomsbury.

Farrell, T. S. C. (2018). *Reflective language teaching: Practical applications for TESOL teachers.* Bloomsbury Publishing.

Farrell, T. S. C., & Macapinlac, M. (2021). Professional development through reflective practice: A framework for TESOL teachers. *Canadian Journal of Applied Linguistics, 24*, 1-25. https://doi.org/10.37213/cjal.2021.28999

Farrell, T. S. C. (2022). *Reflective practice in language teaching.* DOI: 10.1017/9781009028783

Guskey, T. R. (2003). How classroom assessments improve learning. *Educational Leadership, 60*(5), 6–11. ERIC Accession No.: EJ660875

Kolb, D. A. (2015). *Experiential learning: Experience as the source of learning and development* (2nd ed.). Pearson Education.

Kurosh, S., Yousefi, M. H., & Kashef, S. (2020). Iranian teachers' reflective teaching practice in relation to self-efficacy perceptions: investigating teachers' discipline. *Reflective Practice. 21*. 1-15. DOI:10.1080/14623943.2020.1748591

Lantolf, J. P., & Poehner, M. E. (2018). *Sociocultural theory and the pedagogical imperative in L2 education.* Routledge.

Larrivee, B. (2000). Transforming teaching practice: Becoming the critically reflective teacher. *Reflective Practice, 1*(3), 293–307. https://doi.org/10.1080/713693162

Loughran, J. (2020). *Developing reflective practice: Learning about teaching and learning.* Routledge.

Moradkhani, S., Raygan, A., & Moein, M. (2017). Iranian EFL teachers' reflective practices and self-efficacy: Exploring possible relationships. System. 65. 1-14. DOI:10.1016/j.system.2016.12.011

Richards, J. C., & Lockhart, C. (1995). *Reflective teaching in second language classrooms.* Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511667169

Schön, D. A. (1992). *The reflective practitioner: How professionals think in action* (1st ed.). Routledge. https://doi.org/10.4324/9781315237473

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional learning and development: Best evidence synthesis iteration.* Wellington, New Zealand: Ministry of Education.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press.

Wallace, M. J. (1991). *Training foreign language teachers: A reflective approach.* Cambridge University Press.

Zembylas, M. (2018). Rethinking the demands for 'preferred' teacher professional identities: Ethical and political implications. *Teaching and Teacher Education, 76*, 78-85. https://doi.org/10.1016/j.tate.2018.08.011

Curriculum Research

Volume 6, Issue 3
Sept. 20, 2025

Fairness in High-stakes Testing: Analyzing Differential Item Functioning (DIF) by Gender, School type, and Ethnicity in Iran's National University Entrance Exam for Foreign Languages

Article Type:

Original Research

Authors:

Sanaz Behboudi Nehzomi[1]
ORCiD: 0009-0002-9591-2602

Masood Siyyari[2]
ORCiD: 0000-0002-6273-2739

Gholam-Reza Abbasian[3]
ORCiD: 0000-0003-1507-1736

Article History:

Received: 2025.05.27

Accepted: 2025.09.08

Published: 2025.09.20

Abstract

Numerous experts have underscored the need of fairness in National Entrance Examination items. This study examines whether examinees' performance on items of the National University Entrance Exam for Foreign Languages (NUEEFL), known as “Konkour,” varies based on background, specifically gender, school type, and ethnicity, rather than language proficiency, as the detection of differential item functioning (DIF) may enhance the fairness of high-stakes tests. The research employed a quantitative non-experimental, cross-sectional design. The participants included 200 male and female students, who were chosen randomly from students studying at Islamic Azad University, Science and Research branch in Tehran, Iran. The instruments consisted of a mock NUEEFL test and a researcher-made questionnaire. Upon taking the participants’ consent, the researcher took the mock version of NUEEFL. Next, the participants were asked to answer the questionnaire about their demographic information, including their gender, school type, and ethnicity. A three-phase DIF analysis was conducted to explore examinees' performance across these demographic variables. The results indicated that school type exhibited the most significant DIF, particularly in grammar and cloze assessments, whereas gender DIF was mostly seen in grammar and language function. Moreover, ethnically differential item functioning was significant in vocabulary and cloze assessments. Furthermore, reading comprehension was mostly impartial, with the exception of school type. The results underscore the need for test developers to consider demographic factors to ensure fairness and validity in high-stakes testing contexts.

Key Words: Differential Item Functioning, Ethnicity, Gender, School Type, Testing Fairness

1. Department of Language Teaching, SR.C., Islamic Azad University, Tehran, Iran. Email: sanaz.behboudi@srbiau.ac.ir
2. Department of Language Teaching, SR.C., Islamic Azad University, Tehran, Iran (Corresponding Author). Email: m.siyyari@srbiau.ac.ir
3. Department of English Language and Literature, Imam Ali University, Tehran, Iran. Email: gabbasian@gmail.com

1. Introduction

Student assessments have transformed from their original purpose of gauging job eligibility into a means of holding schools, instructors, and students accountable, as well as allocating educational resources (Yang et al., 2019). The aim of administering exams is to determine how well learners have learnt a given topic over a specified period of time to fulfill predetermined goals (Yang et al., 2019). For most students, it marks a major turning point in their scholastic path, opening doors to greater opportunities for growth. Testing was crucial in China about 200 BC for deciding who could hold public service jobs. It was common practice in Italy to grade educators by the end of the fourteenth century on the basis of their students' exam scores. These days, test scores are used for a lot of different things by politicians in different countries. First, they help with accountability (for students, teachers, and schools), second, they show where limited funds should go, and third, they inform political and government decisions that try to improve education (Russell et al., 2009).

The history of research on bias in testing has been contentious, particularly since Jensen's seminal work, "Bias in Mental Testing" (1980). Bias, in its statistical and ethical dimensions, remains a critical concern in educational assessments. A student's real knowledge or skills may not be accurately reflected by certain parts of the exam, leading to inaccurate or misleading judgments due to bias in testing. French (2020) points out that the test's psychometric qualities or general design might be the source of bias. Differences in prediction accuracy or the criteria used to pick the test are two examples of external variables that could introduce bias. According to French, bias is an inherent flaw in the measuring procedure that might have varying impacts on different groups' test results. Unfair results in student assessments may also result from assessment bias, which can have a disproportionate effect on pupils according to personal factors such as ethnicity, gender, socioeconomic level, religion, and origin (Kim & Zabelina, 2015).

As scholars, like Chalhoub-Deville (2015) and Kunnan (2018) argue, fairness in testing is paramount for social justice, necessitating valid and equitable measurement tools. Furthermore, a test must be equitable for diverse test takers. Specifically, it should not exhibit bias against the characteristics of the examinees (e.g., gender, race). Addressing this issue necessitates a statistical methodology for test analysis that can initially determine whether test items function differentially among various groups and ultimately identify the sources of this variance (Geranpayeh & Kunnan, 2007). One proposed method for this purpose is Differential Item Functioning (DIF). DIF arises when examinees with equivalent capability levels from distinct groups have differing probabilities of affirming an item (Mazor et al., 1998).

In Iran, the official name for the Iranian National University Exam is the “Konkour” examination. It is perhaps an altered version of the French word "concours," which means "to

source, screen, and select" and describes a variety of human resource management initiatives. Languages other than English are among the three main areas in which Konkour is offered. According to Razmjo (2006), this national English proficiency exam was first given to all citizens in 2002. In order to administer the exams, the National Organization of Educational Testing (NOET) collaborates with the major institutions in every city in the country. People who want to attend public/state universities that do not charge tuition have to pass this very tough test (Hosseini, 2007). The exam evaluates not just reading comprehension and vocabulary but also linguistic functions, the capacity to complete multiple-choice questions, and understanding of English syntax and structure. All of the materials included on the exam are taken from the textbooks that are used as primary resources for education in Iran. This assessment is used to gauge the students' proficiency and is known as a norm-referenced exam. Those who take the exam often are fluent Farsi speakers who visit designated testing locations (Khodi et al., 2021).

The importance assigned to Konkour generates concern in both candidates and their parents, due to its capacity to significantly influence their socio-economic standing (Parviz, 2023). A number of scholars have shown that Konkour remains a significant barrier to attending higher education institutions and is a substantial challenge to attaining educational parity (Ghorbani, 2012; Kamyab, 2007, 2008; Safari, 2016; Safari & Rashidi, 2018).

This study is significant as, in the Iranian context, it is assumed that most language tests in high stakes are not fair because they do not have validity (Safari, 2016). Due to the fact that the Konkoor determines examinees' future in terms of their study and career, as well as their personal life, it must be free from any kind of bias, and treats all examinees fairly (Khodi et al., 2021). Zumbo (1999) asserts that the concept of item bias has significant implications for policy, administration, and teaching settings. Consequently, bias may result in systemic mistakes that skew the conclusions drawn in the categorization and selection of pupils. This indicates that test-takers with comparable understanding of the test content should get equal results on individual questions, irrespective of their affiliation with other groups, including gender, culture, ethnicity, or race (Weijters et al., 2013). The concept of group is central to the definitions of bias (Davis, 2013), and this concept could be studied also in relation to other groups, such as social class, age, religion, or any other sociodemographic characteristic of the learners. As a result, this study was an attempt to investigate whether examinees' performances on test items differ due to their background, including gender, school type, and ethnicity rather than their language proficiency.

2. Review of the Related Literature

Fairness is seen as the core principle of classroom assessment notions and quality assessment techniques (Baniasadi et al., 2022; Rezai et al., 2021), consistently highlighted as a vital quality and a key trait of assessment-literate instructors (DeLuca et al., 2016). Mislevy (2018) posited fairness as a logical foundation for accommodating the interests and past knowledge of test-takers. He contendwd that comprehending job performance requires consideration of the individual, their past, the tasks, the environment, and the contextual factors.

The prospect of learning often prioritizes equity above evaluation and is closely linked to education (Rasooli et al., 2018). It comprises exposure to test material or more broadly refers to the alignment between curriculum and assessment (Tierney, 2016). Furthermore, it encompasses many educational possibilities tailored to learners' distinct learning styles, abilities, and exceptionalities (Lantolf & Poehner, 2013). An opportunity to showcase learning involves offering diverse, equitable, and meaningful chances for learners to display their knowledge (Tierney, 2016). Mauldin (2009) experimentally showed that evaluation is equitable when numerous assessment chances are provided, hence supporting the idea of fairness in assessment.

The non-existence of bias in a test is frequently employed as a component of the test fairness framework when data about the outcomes is gathered (Bachman, 2005). Test biases encompass: a) Offensive content or language, which pertains to material that is derogatory to examinees from diverse backgrounds, including stereotypes of group members and explicit or implicit slurs or insults related to gender, race and ethnicity, religion, age, language, national origin, and sexual orientation. b) Unjust penalization is influenced by the examinee's history, which pertains to material that may lead to inequitable penalties based on a test taker's group affiliation, including but not limited to gender, race, ethnicity, religion, age, language, country origin, and sexual orientation. c) Disparate influence and standard setting, which pertains to varying performances and following consequences among examinees from distinct group affiliations. To ensure fairness and validity in assessment, potential group disparities linked to key test-taker attributes, including gender, race, ethnicity, religion, age, language, national origin, and sexual orientation, should be rigorously investigated through Differential Item Functioning (DIF) and Differential Test Functioning (DTF) analyses at both the item and subtest levels (Zumbo, 1999). A differential validity study should be performed to see whether a test predicts success more effectively for one group compared to another. Test results should be evaluated about the criteria measure and selection judgments in the context of standard-setting. Test creators and consumers must be assured that valid measures and statistically robust, unbiased selection methods are used (Bachman, 2005). These studies

should educate test makers and users that group differences correlate with the tested abilities and that irrelevant constructs should not be included.

Khodi (2020) performed a generalizability research including 5000 Konkur examinees, demonstrating that 86% of the overall variation is due to individual differences, indicating a high degree of test reliability. The interaction between individuals' fields of study and the common questions in the test sections led to an error of around 1.5%. The national entrance examination demonstrates impartiality towards persons from diverse educational backgrounds. Also, Khodi et al. (2021) examined the admission examination for Iranian universities, referred to as "Konkour". Given the significance of this high-stakes examination, which may have social and long-term ramifications for the participants, they assessed the test and its psychometric properties. Their results suggested that the test offers a constrained context for assessing participants' "knowledge of language" rather than their "knowledge about language." Consequently, the dimensionality and validity of the assessment remain contentious.

In their study on graduate students' perceptions of assessment fairness, Darabi et al. (2022) found that "equity and interactional fairness" were the most important variables in deciding whether an assessment was fair or unfair. By interviewing 27 seasoned high school teachers, Rasooli et al. (2022) sought to understand what factors influence teachers' perspectives on equality in the classroom. Individual mechanisms, societal mechanisms, and the dialectical links between the two, were the three main topics that affected how instructors perceived assessment fairness. The research showed that teachers' views and behaviors on assessment fairness are shaped by a complex web of factors, including their own beliefs and experiences as well as the social, institutional, and classroom settings in which they work.

Dadvar and Tabatabaee-Yazdi (2023) examined the correlation between Iranian EFL learners' views of equity in assessment and their cognitive test anxiety. The research used a quantitative methodology and included a total of 201 participants. A convenience and random sample approach was used, using a fair evaluation and a cognitive test anxiety questionnaire. The study's findings, derived from Pearson-moment correlation and regression analysis, indicated no significant association between Iranian EFL learners' judgments of fairness in classroom assessment and their cognitive test anxiety. Besides, no substantial difference between male and female learners' sense of fairness in classroom evaluation and their cognitive test anxiety was shown.

Educators of English as a foreign language (EFL) have their perceptions of fairness in evaluation (Tofighi & Ahmadi Safa, 2023). These two researchers first aimed to develop and assess a Classroom Assessment Fairness Scale (CAFS) as part of a whole methodology. A

total of 120 EFL teachers in Iran were given the validated scale. According to the results, EFL educators had a deep comprehension of what constitutes a fair assessment strategy for the classroom. Equity in classroom assessment techniques was identified as critically important by their recognition of the importance of learning opportunities, the capacity to show knowledge, a positive classroom atmosphere, the no-harm principle, and openness. Score pollution was not, however, guaranteed to have the same level of familiarity or awareness. Further, there were statistically significant variations in how EFL teachers perceived the fairness of classroom assessment based on instructors' gender, educational background, years of experience teaching, and learning environment.

Parviz (2023) performed an exhaustive analysis of the previous developments and current status of Konkour. A systematic-narrative hybrid literature review (SNHLR) was used to meet the study's aims. Seventy-four peer-reviewed research papers, both national and international, were meticulously selected and analyzed to identify significant themes, patterns, gaps, and trends regarding the ramifications of Konkour. The findings revealed that Iranian university applicants had faced many negative consequences, including economic, psychological, and educational impacts owing to the Konkour.

Due to the importance of test fairness, numerous studies have been carried out and various models have been proposed in the international and Iranian contexts (e.g., Dadvar & Tabatabaee-Yazdi, 2023; Haertel & Herman, 2005; McNamara & Roever, 2006; Shohamy & Eldar, 2000; Tofighi & Ahmadi Safa, 2023). However, the studies conducted so far do not yield a compelling account of the fairness associated with the Konkoor examination in the Iranian EFL context. They only propose the general constructs of fairness without going into details of the issue (e.g., Kodi et al., 2021). Thus, the present research aimed to provide a quantified and objective account of fairness in the Konkoor examination to fill the gap in the literature. Therefore, the overarching research question was as follows:

- Do examinees perform differently on any test items as a result of their background (e.g., gender, school type, or ethnicity) rather than their language proficiency?

3. Method

3.1. Design

The design of the present study was a quantitative, non-experimental, cross-sectional research design with a focus on DIF analysis (Ary et al., 2018)**.** Data were collected at a single point in time (after administering the mock NUEEFL test and the demographic questionnaire). It did not involve experimental manipulation but rather examined naturally occurring group

differences in test performance.

3.2. Participants

The participants of the present study included B.A. Konkoor candidates, including 200 students of both genders (100 male and 100 female) who had taken National University Entrance Exam for Foreign Languages (NUEEFL) known as “Konkour”. They were selected randomly from university students from different regions studying at Islamic Azad University, Science and Research branch in Tehran, Iran. The first language of the participants was Persian, and their age ranged between 18-20 years old. Prior to the study, written consent regarding the participation in this study was obtained from all the participants.

3.3. Instruments

3.3.1. Konkour Test. A mock Konkour was administered to the participants. The purpose of this mock Konkour test was to evaluate the test in terms of its differential item functioning. The exam was structured into six components, each reflecting the content of the textbooks provided to high school students. The test had 70 items as follows:

Grammar Test: This section comprised ten questions about English grammar. The questions were formatted as unfinished sentences that students had to complete by selecting from alternatives, which included phrases, nouns, prepositions, or verbs. The succession of questions was not governed by a predetermined rule; instead, it was arranged randomly for each participant, rather than uniformly across all participants. Occasionally, two grammatical rules were conflated inside a single question, making it very complicated for students to discern the underlying concept and identify the proper answer.

Vocabulary Test: This section included 15 questions formatted as incomplete sentences. The students had to choose the best alternative for completing the meaning of the sentences. The right answer for the participants was already communicated in the classroom environment, however the distractors consisted of unfamiliar vocabulary for them. The parts of speech could vary among questions, although efforts were made to maintain consistency across the possibilities of each individual item to reduce the likelihood of random guessing by participants.

Sentence Structure: In this part, there were five questions; each item of the question presented a sentence and the participants had to select the item in which there was no grammatical mistake based on the stem of the question. The sentences were relatively long, often containing complex structures, and errors could occur in any part of the phrase.

Language Functions: This section was consisted of various conversations and comprised ten questions. Participants had to complete the dialogues with the most appropriate responses from the provided alternatives. The accurate response should operate as a complement to the

interaction occurring between both parties in the conversation.

Cloze Test: In this section, participants were required to read a text containing 15 blanks, almost positioned at regular intervals, such as every ten words, and choose the choice that most effectively completed each sentence. The presentation of blanks inside a single text could result in misunderstandings or errors in identifying the right answer for one blank, perhaps leading participants to pick inappropriate options for subsequent blanks.

Reading Comprehension: Each exam encompassed three reading comprehension passages, each ranging from 350 to 500 words. The passages included a diverse array of subjects, such as scientific and social themes. Each text included five multiple-choice questions about the topic, word meanings, and sentence interpretations.

3.3.2. Demographic Questionnaire. A researcher-designed questionnaire was employed in the present study to collect participants' demographic information, including gender, school type, and ethnicity.

3.4. Procedure

Participants were randomly selected from among B.A. Konkour candidates, consisting of 200 students of both genders who had taken the National University Entrance Examination of Foreign Languages (NUEEFL), commonly known as the Konkour, at the Islamic Azad University, Science and Research Branch. The researcher explained the study's objectives to the participants, obtained their informed consent, and then administered a mock version of the NUEEFL. The test was conducted in a single 100-minute session. Following the exam, participants completed a demographic questionnaire providing information on gender, school type, and ethnicity. The participants differed in age and field of study. While all were proficient in Persian as the medium of instruction, their native languages were not systematically collected and therefore are not reported. These data were subsequently analyzed using Differential Item Functioning (DIF) to address the research question, with examinees' performance on the NUEEFL examined in relation to gender, school type, and ethnicity.

3.5 Data Analysis

The collected data were analyzed using **Differential Item Functioning (DIF)** to examine the fairness of the NUEEFL across gender, school type, and ethnicity. After data screening and descriptive statistics, DIF analysis was used to identify whether test items functioned differently for subgroups of examinees who were matched on overall ability. This approach allowed the study to detect potential sources of bias and evaluate the extent to which item performance reflected differences in construct-relevant ability rather than group membership.

4. Results

4.1. Demographic Characteristics of the Participants

Table 1 displays the demographic characteristics of the participants. With respect to ethnicity, Fars (34.2%) and Turk (32.3%) students constituted the largest groups, while Kurd (16.9%) and Gilaki (16.5%) students accounted for smaller but still substantial portions of the sample. The gender distribution reveals a clear imbalance, with females representing nearly three-quarters of the participants (74.8%) and males comprising only one-quarter (25.2%). Regarding school type, a majority of the examinees attended private schools (56.0%), whereas 44.0% were enrolled in public schools.

Table 1.

Frequencies and Percentages of Demographic Information

		Frequency	Percent
Ethnicity	Fars	91	34.2
	Gilaki	44	16.5
	Kurd	45	16.9
	Turk	86	32.3
	Total	266	100.0
Gender	Female	199	74.8
	Male	67	25.2
	Total	266	100.0
School Type	Private	149	56.0
	Public	117	44.0
	Total	266	100.0

4.2. Differential Item Functioning on Gender

Table 2 displays the results of the DIF for the grammar sub-section of NUEEFL. The results indicated that there were significant DIF's on first (χ^2 = 16.95, p = .002), fourth (χ^2 = 10.26, p = .020), and seventh items (χ^2 = 13.35, p = .006). Before discussing the results, it should be noted that two sets of p-values were computed. The last column included the p-values adjusted for multiple comparisons made to reduce the inflated error rate.

Table 2.

Differential Item Functioning Grammar by Gender

	Chi-square	p-value	Adjusted p-value	Significant DIF
Item1	16.952	0.000	0.002	**
Item2	1.401	0.497	0.621	
Item3	0.336	0.845	0.845	
Item4	10.264	0.006	0.020	*
Item5	4.850	0.089	0.221	
Item6	1.866	0.393	0.562	
Item7	13.352	0.001	0.006	**
Item8	2.074	0.355	0.562	

Item9	1.104	0.576	0.640
Item10	3.326	0.190	0.379

Table 3 displays the results of DIF which compare male and female groups' performance on 15 items of the vocabulary test. The results showed no significant DIF's for the two groups on vocabulary test.

Table 3.

Differential Item Functioning Vocabulary by Gender

	Chi-square	p-value	Adjusted p-value	Significant DIF
Item11	4.435	0.109	0.611	
Item12	3.630	0.163	0.611	
Item13	0.954	0.621	0.776	
Item14	1.335	0.513	0.776	
Item15	2.139	0.343	0.735	
Item16	1.720	0.423	0.776	
Item17	2.603	0.272	0.680	
Item18	0.253	0.881	0.881	
Item19	0.960	0.619	0.776	
Item20	0.428	0.807	0.881	
Item21	4.096	0.129	0.611	
Item22	0.299	0.861	0.881	
Item23	3.686	0.158	0.611	
Item24	2.950	0.229	0.680	
Item25	1.177	0.555	0.776	

Table 4 displays the results of DIF which compare male and female groups' performance on five items of the sentence structure test. The results showed no significant DIF's for the two groups on sentence structure test.

Table 4.

Differential Item Functioning Sentence Structure by Gender

	Chi-square	**p-value**	**Adjusted p-value**	**Significant DIF**
Item26	0.330	0.848	0.848	
Item27	1.337	0.513	0.848	
Item28	3.400	0.183	0.848	
Item29	1.142	0.565	0.848	
Item30	0.609	0.738	0.848	

Table 5 displays the results of the DIF for the language function sub-section of NUEEFL. The results indicated that there were significant DIF's on item 34 ($\chi^2 = 15.95$, $p = .002$), and item 39 ($\chi^2 = 12.00$, $p = .012$).

Table 5.

Differential Item Functioning Language Function by Gender

	Chi-square	p-value	Adjusted p-value	Significant DIF
Item31	1.421	0.492	0.614	
Item32	0.984	0.611	0.679	
Item33	7.964	0.019	0.062	
Item34	15.954	0.000	0.003	**
Item35	0.318	0.853	0.853	
Item36	3.560	0.169	0.281	
Item37	4.474	0.107	0.214	
Item38	2.304	0.316	0.451	
Item39	12.004	0.003	0.012	*
Item40	4.597	0.100	0.214	

Table 6 displays the results of DIF which compare male and female groups' performance on 15 items of the cloze test. The results showed no significant DIF's for the two groups on cloze test.

Table 6.

Differential Item Functioning Cloze Test by Gender

	Chi-square	p-value	Adjusted p-value	Significant DIF
Item41	0.549	0.760	0.966	
Item42	0.069	0.966	0.966	
Item43	0.389	0.823	0.966	
Item44	0.422	0.810	0.966	
Item45	1.315	0.518	0.966	
Item46	1.964	0.375	0.966	
Item47	1.792	0.408	0.966	
Item48	0.326	0.850	0.966	
Item49	3.047	0.218	0.966	
Item50	6.168	0.046	0.687	
Item51	0.256	0.880	0.966	
Item52	1.059	0.589	0.966	
Item53	3.459	0.177	0.966	
Item54	2.321	0.313	0.966	
Item55	0.134	0.935	0.966	

Finally, Table 7 displays the results of DIF which compare male and female groups' performance on 15 items of the reading comprehension test. The results showed no significant DIF's for the two groups on reading comprehension test.

Table 7.

Differential Item Functioning Reading Comprehension by Gender

	Chi-square	p-value	Adjusted p-value	Significant DIF
Item56	0.004	0.998	0.998	
Item57	0.643	0.725	0.998	
Item58	0.066	0.968	0.998	
Item59	0.704	0.703	0.998	
Item60	2.937	0.230	0.998	
Item61	0.032	0.984	0.998	
Item62	1.197	0.550	0.998	
Item63	0.209	0.901	0.998	
Item64	2.594	0.273	0.998	
Item65	1.461	0.482	0.998	
Item66	1.179	0.555	0.998	
Item67	0.313	0.855	0.998	
Item68	1.498	0.473	0.998	
Item69	2.711	0.258	0.998	
Item70	5.362	0.069	0.998	

4.2. Differential Item Functioning on School Type

Table 8 displays the results of the DIF for the grammar sub-section of NUEEFL by school type. The results indicated that there were significant DIF's on second (χ^2 = 7.60, p = .045), third (χ^2 = 32.90, p = .000), fourth (χ^2 = 13.78, p = .003), fifth (χ^2 = 531.02, p = .000), and tenth (χ^2 = 652.96, p = .000) items. Before discussing the results, it should be noted that the private schools were labeled as focal.

Table 8.

Differential Item Functioning Grammar by School Type

	Chi-square	p-value	Adjusted p-value	Significant DIF
Item1	0.226	0.893	0.893	
Item2	7.609	0.022	0.045	*
Item3	32.909	0.000	0.000	***
Item4	13.782	0.001	0.003	**
Item5	531.025	0.000	0.000	***
Item6	2.509	0.285	0.357	
Item7	3.297	0.192	0.275	
Item8	1.889	0.389	0.432	
Item9	5.346	0.069	0.115	
Item10	652.961	0.000	0.000	***

Table 9 displays the results of DIF which compare public and private schools' performance on 15 items of the vocabulary test. The results showed there were significant

DIF's on items 14 (χ^2 = 17.65, p = .002), and 23 (χ^2 = 10.04, p = .049).

Table 9.

Differential Item Functioning Vocabulary by School Type

	Chi-square	p-value	Adjusted p-value	Significant DIF
Item11	4.885	0.087	0.145	
Item12	7.208	0.027	0.068	
Item13	2.412	0.299	0.321	
Item14	17.653	0.000	0.002	**
Item15	7.775	0.021	0.068	
Item16	2.439	0.296	0.321	
Item17	0.267	0.875	0.875	
Item18	5.551	0.062	0.117	
Item19	2.498	0.287	0.321	
Item20	3.577	0.167	0.251	
Item21	6.875	0.032	0.069	
Item22	7.434	0.024	0.068	
Item23	10.044	0.007	0.049	*
Item24	9.005	0.011	0.055	
Item25	2.980	0.225	0.307	

Table 10 displays the results of DIF which compare public and private schools' performance on five items of the sentence structure test. The results showed no significant DIF's for the two groups on sentence structure test.

Table 10.

Differential Item Functioning Sentence Structure by School Type

	Chi-square	p-value	Adjusted p-value	Significant DIF
Item26	3.676	0.159	0.746	
Item27	1.608	0.447	0.746	
Item28	0.574	0.751	0.938	
Item29	0.023	0.989	0.989	
Item30	2.259	0.323	0.746	

Table 11 displays the results of the DIF for the language function sub-section of NUEEFL. The results indicated that there were not any significant DIF's for ten items of language function.

Table 11.

Differential Item Functioning Language Function by School Type

	Chi-square	p-value	Adjusted p-value	Significant DIF
Item31	0.123	0.940	0.940	
Item32	1.463	0.481	0.688	

Item33	5.981	0.050	0.181
Item34	5.715	0.057	0.181
Item35	0.602	0.740	0.910
Item36	5.249	0.073	0.181
Item37	1.880	0.391	0.688
Item38	1.536	0.464	0.688
Item39	0.399	0.819	0.910
Item40	10.442	0.005	0.054

Table 12 displays the results of DIF which compare public and private schools' performance on 15 items of the cloze test. The results showed there were significant DIF's on 11 items out of the 15 items of the cloze test; item 42 (χ^2 = 12.28, p = .007), item 45 (χ^2 = 8.40, p = .025), item 47 (χ^2 = 7.96, p = .029), item 48 (χ^2 = 8.40, p = .028), item 49 (χ^2 = 17.92, p = .001), item 50 (χ^2 = 18.18, p = .001), item 51 (χ^2 = 6.88, p = .044), item 52 (χ^2 = 27.17, p = .000), item 53 (χ^2 = 10.44, p = .014), item 54 (χ^2 = 7.86, p = .029), and item 55 (χ^2 = 22.29, p = .000).

Table 12.

Differential Item Functioning Cloze Test by School type

	Chi-square	p-value	Adjusted p-value	Significant DIF
Item41	4.704	0.095	0.102	
Item42	12.281	0.002	0.007	**
Item43	5.941	0.051	0.064	
Item44	1.108	0.575	0.575	
Item45	8.402	0.015	0.028	*
Item46	5.048	0.080	0.093	
Item47	7.967	0.019	0.029	*
Item48	8.408	0.015	0.028	*
Item49	17.923	0.000	0.001	***
Item50	18.189	0.000	0.001	***
Item51	6.880	0.032	0.044	*
Item52	27.176	0.000	0.000	***
Item53	10.441	0.005	0.014	*
Item54	7.868	0.020	0.029	*
Item55	22.299	0.000	0.000	***

Finally, Table 13 displays the results of DIF which compare public and private schools' performance on 15 items of the reading comprehension test. The results showed there were two significant DIF's on items 59 (χ^2 = 24.71, p = .000), and 62 (χ^2 = 10.46, p = .040).

Table 13.

Differential Item Functioning Reading Comprehension by School type

	Chi-square	p-value	Adjusted p-value	Significant DIF
Item56	2.405	0.300	0.349	
Item57	3.591	0.166	0.277	
Item58	8.909	0.012	0.058	
Item59	24.710	0.000	0.000	***
Item60	0.093	0.955	0.974	
Item61	6.690	0.035	0.078	
Item62	10.463	0.005	0.040	*
Item63	2.792	0.248	0.349	
Item64	6.621	0.037	0.078	
Item65	5.196	0.074	0.140	
Item66	2.393	0.302	0.349	
Item67	7.615	0.022	0.078	
Item68	0.053	0.974	0.974	
Item69	7.194	0.027	0.078	
Item70	2.606	0.272	0.349	

4.3. Differential Item Functioning on Ethnicity

Table 14 displays the results of the DIF for the grammar sub-section of NUEEFL by ethnicity. The results indicated that there were significant DIF's on first ($\chi^2 = 10.32$, $p = .029$), and tenth ($\chi^2 = 17.33$, $p = .002$) items.

Table 14.

Differential Item Functioning Grammar by Ethnicity

	Chi-square	p-value	Adjusted p-value	Significant DIF
Item1	10.320	0.006	0.029	*
Item2	0.825	0.662	0.769	
Item3	0.022	0.989	0.989	
Item4	0.795	0.672	0.769	
Item5	0.736	0.692	0.769	
Item6	2.351	0.309	0.617	
Item7	5.181	0.075	0.250	
Item8	1.684	0.431	0.718	
Item9	3.279	0.194	0.485	
Item10	17.335	0.000	0.002	**

Table 15 displays the results of DIF which compare ethnicity groups' performance on 15 items of the vocabulary test. The results showed there were significant DIF's on items 14 (χ^2

= 9.27, p = .036), 15 (χ^2 = 13.36, p = .009), 19 (χ^2 = 31.74, p = .000), and 21 (χ^2 = 11.65, p = .015).

Table 15.

Differential Item Functioning Vocabulary by Ethnicity

	Chi-square	p-value	Adjusted p-value	Significant DIF
Item11	6.655	0.036	0.067	
Item12	4.726	0.094	0.157	
Item13	1.325	0.516	0.552	
Item14	9.271	0.010	0.036	*
Item15	13.367	0.001	0.009	**
Item16	0.250	0.883	0.883	
Item17	3.319	0.190	0.259	
Item18	3.519	0.172	0.258	
Item19	31.749	0.000	0.000	***
Item20	2.171	0.338	0.422	
Item21	11.659	0.003	0.015	*
Item22	7.434	0.024	0.067	
Item23	6.797	0.033	0.067	
Item24	6.872	0.032	0.067	
Item25	1.691	0.430	0.496	

Table 16 displays the results of DIF which compare four ethnicity groups' performance on five items of the sentence structure test. The results showed a single significant DIF on item 30; i.e. (χ^2 = 15.58, p = .002).

Table 16.

Differential Item Functioning Sentence Structure by Ethnicity

	Chi-square	p-value	Adjusted p-value	Significant DIF
Item26	0.418	0.811	0.811	
Item27	4.086	0.130	0.324	
Item28	2.588	0.274	0.457	
Item29	1.803	0.406	0.507	
Item30	15.589	0.000	0.002	**

Table 17 displays the results of the DIF for the language function sub-section of NUEEFL. The results indicated that there were not any significant DIF's on ten items of language function.

Table 17.

Differential Item Functioning Language Function by Ethnicity

	Chi-square	p-value	Adjusted p-value	Significant DIF
Item31	0.807	0.668	0.833	
Item32	0.365	0.833	0.833	
Item33	3.659	0.161	0.321	
Item34	6.412	0.041	0.135	
Item35	2.727	0.256	0.365	
Item36	3.853	0.146	0.321	
Item37	7.641	0.022	0.112	
Item38	2.851	0.240	0.365	
Item39	0.413	0.814	0.833	
Item40	7.601	0.022	0.112	

Table 18 displays the results of DIF which compare ethnicity groups' performance on 15 items of the cloze test. The results showed there except for the first two items; item 43 to 55 showed significant DIF across three ethnicity groups.

Table 18.

Differential Item Functioning Cloze Test by Ethnicity

	Chi-square	p-value	Adjusted p-value	Significant DIF
Item41	0.000	1.000	1.000	
Item42	0.000	1.000	1.000	
Item43	42.857	0.000	0.000	***
Item44	22.118	0.000	0.000	***
Item45	49.738	0.000	0.000	***
Item46	49.566	0.000	0.000	***
Item47	17.922	0.000	0.000	***
Item48	61.123	0.000	0.000	***
Item49	5.949	0.015	0.017	*
Item50	31.741	0.000	0.000	***
Item51	23.346	0.000	0.000	***
Item52	38.598	0.000	0.000	***
Item53	130.979	0.000	0.000	***
Item54	80.333	0.000	0.000	***
Item55	64.650	0.000	0.000	***

Finally, Table 19 displays the results of DIF which compare ethnicity groups' performance on 15 items of the reading comprehension test. The results showed there were not any significant DIF's on 15 items of reading comprehension section of NUEEFL.

Table 19.

Differential Item Functioning Reading Comprehension by Ethnicity

	Chi-square	p-value	Adjusted p-value	Significant DIF
Item56	3.840	0.147	0.459	
Item57	3.391	0.184	0.459	
Item58	2.352	0.309	0.490	
Item59	1.364	0.506	0.632	
Item60	4.467	0.107	0.459	
Item61	0.537	0.765	0.819	
Item62	2.238	0.327	0.490	
Item63	2.927	0.232	0.490	
Item64	1.184	0.553	0.638	
Item65	4.977	0.083	0.459	
Item66	6.773	0.034	0.459	
Item67	1.505	0.471	0.632	
Item68	0.004	0.998	0.998	
Item69	2.268	0.322	0.490	
Item70	3.453	0.178	0.459	

5. Discussion

The results indicated significant DIF related to gender, school type, and ethnicity in the NUEEFL, highlighting serious issues with fairness, construct validity, and possible bias in high-stakes language evaluation. The results indicated that school type exhibited the most significant DIF, particularly in grammar and cloze test assessments, whereas gender DIF was mostly seen in grammar and language function. Moreover, ethnically differential item functioning was significant in vocabulary and cloze test assessments, with item 19 (vocabulary) exhibiting the most significant bias. Furthermore, reading comprehension was mostly impartial, with the exception of school type. The findings suggest that the grammar and language function sections may include elements that preferentially benefit one gender, necessitating more examination. The cloze test and certain grammar components may disfavor learners from both public and private schools, prompting issues over fairness in test design. Moreover, the vocabulary and cloze test sections may include culturally or linguistically biased elements, thus influencing performance across different ethnic groups.

The most pervasive DIF was observed between public and private schools, particularly in grammar and cloze tests. Items such as grammar questions 5 and 10**,** exhibited extreme bias suggesting that students from different school backgrounds may interpret or respond to these items, differently. Private schools may emphasize certain grammatical structures or test-taking strategies not equally covered in public schools. Moreover, access to quality English

instruction, tutoring, or practice materials may vary significantly between school types. In addition, private school students may have more opportunities for immersive English exposure.

Significant DIF was found for male and female test-takers **in** grammar (items 1, 4, 7) and language function (items 34, 39)**.** Prior research suggests that males and females may employ different strategies in grammatical reasoning or pragmatic language use. Furthermore, certain topics or phrasing in these items may resonate differently across genders. Ethnicity-related DIF was most prominent in vocabulary (item 19) and the cloze test (items 43–55)**.** Some vocabulary items (e.g., item 19) may reflect concepts more familiar to certain ethnic groups. Moreover, cloze passages may contain structures more aligned with specific linguistic backgrounds.

Reading comprehension showed minimal DIF (only two items affected by school type), suggesting it may be the most robust section for fair assessment. Reading tasks assess skills less dependent on isolated grammar/vocabulary knowledge. This finding lend credence to those Gonzalez and Hinton (2018) who found that reading comprehension items had minimal DIF across various educational environments.

The findings of the current study revealed that the NUEEFL may fail to satisfy norms of fairness and dependability. The results are in line with other studies that have raised questions about the accuracy, fairness, and equity of high-stakes language tests in different settings (Alderson & Hamp-Lyons, 1996; Bachman, 1990; Shohamy & Eldar, 2002). In particular, aligning with Alderson and Hamp-Lyons (1996), the findings suggest that these tests may favor specific groups of learners, consequently challenging assertions of fairness. Bachman's (1990) framework of test usefulness aligns with the current study by illustrating that construct validity may be undermined when contextual and social factors, such as gender or ethnicity, disproportionately affect performance. In the same way, Shohamy and Eldar (2002) stressed the sociopolitical aspects of testing, showing how high-stakes exams may make things worse instead of giving everyone a fair chance. This is also true of the current findings.

In the Iranian context, the findings validate previous research that has emphasized pervasive inequities in the NUEEFL. For instance, Kamyab (2007, 2008) and Ghorbani (2012) recorded problems of construct underrepresentation and content imbalance, which are similar to the item-level biases found in this study. Safari (2016) and Safari and Rashidi (2018) highlighted the sociocultural and educational disadvantages encountered by students from public schools and rural regions, findings that correspond with the current study's evidence of differential item functioning across school types. Khodi et al. (2021) and Parviz (2023) have recently reported that there are still differences in access, performance, and fairness in the

NUEEFL based on gender and region. This is in line with what we found in the current analysis.

The results corroborate those of Khodi et al. (2021), who determined that although the test context and content align, the nature of the assessment is not a reliable predictor of participants' actual language proficiency owing to the absence of some essential language abilities in the evaluation. Furthermore, the results go against what Khodi (2020) found, which is that the national entrance exam does not discriminate against any group of people based on their level of education.

Despite the fact that the designers of Konkour took into mind the concept of linguistic competency, fully operationalizing it proved to be difficult owing to practical limits and test-related issues. This is reflected in the style and structure of the items. To be more specific, the productive and receptive abilities of speaking, writing, and listening were not included in the evaluation. This was primarily due to the fact that there were regional differences in availability to training materials and the logistical issues that were involved in delivering such components throughout the country. According to Khodi et al. (2021), their absence was also justified on the basis of eliminating subjectivity in scoring. This was due to the fact that it is difficult to ensure inter-rater reliability and fairness in large-scale testing. In light of this, Konkour has placed an emphasis on skills that can be objectively measured, which has resulted in an increase in reliability and practicality. However, this has come at the expense of construct representativeness. With this trade-off, some logistical and equitable problems are addressed; but, at the same time, the breadth of language ability that is being evaluated is narrowed, which increases the danger of creating graduates who have limited communicative competence. The current analysis validates the use of Differential Item Functioning (DIF) analyses to give empirical evidence of fairness. These design restrictions also heighten the need of assessing whether test items work fairly across subgroups, which is why the present study used these analyses.

6. Conclusion

The current research sought to investigate whether examinees' performance on the NUEEFL varied based on background, specifically, gender, school type, and ethnicity, rather than language competency. DIF analyses revealed that the majority of test items performed equitably across the groups, however several items had differential functioning, indicating possible sources of bias. These findings highlight the need of overseeing and enhancing high-stakes language tests to guarantee they accurately assess language competency and provide fair possibilities for all test-takers.

This study highlights the complex interplay of gender, school type, ethnicity and English

language proficiency. Addressing these factors through equitable testing practices and inclusive curricula can enhance language learning outcomes for all students. Future research should continue to explore these dynamics to inform more just and effective educational assessments. As educational environments become increasingly diverse, it is essential to develop and implement policies and practices that recognize and support this diversity, ensuring that all students have the opportunity to achieve high levels of language proficiency.

The study provides important insights into the fairness of the NUEEFL. These findings suggest that improvements are needed to ensure that the test is reliable, valid, and fair for all examinees, regardless of their gender, school type, or ethnicity. This study has also implications for policymakers, test developers, and educators who need to address these issues and ensure that the test meets international standards of fairness. The findings underscore the need for educational policies that promote fairness and inclusivity in testing. Test developers must consider the diverse backgrounds of test-takers to ensure equitable assessment practices. Educators should adopt culturally responsive teaching methods to support all students, particularly those from minority backgrounds. Policymakers should advocate for curricula that integrate multicultural perspectives and provide adequate resources for public schools to offer high-quality language education. Training programs for teachers should emphasize the importance of cultural competence and gender sensitivity in language instruction.

The present study offers several important implications for Konkour test designers too. It emphasizes meticulously analyzing test questions for gender, school type, and ethnicity fairness. The content should be checked to ensure that it does not bias or disadvantage certain groups. Second, the test constructors should provide a more complete picture of language abilities to better assess English proficiency. This should encompass information production and reception. This study shows that conventional test validation procedures require statistical methods, like DIF. This would enable empirical bias detection and correction. The test designers must be proactive about equality to boost Konkour's legitimacy, validity, and social acceptability as a high-stakes language exam. This may be achieved by matching item selection and scoring with global fairness norms.

The university entrance examination is clearly a high-stake test whose results have grave consequences for the test takers. Failure on the test cannot be easily ignored. This failure might be due to failure on a single item. Therefore, it is the responsibility of the test developers and users to make sure that each and every item included in the test is fair. Our results show that this is not the case. All in all, by addressing the issues presenting in the present study, the NUEEFL can better serve its purpose as a fair and valid measure of English proficiency for all test-takers, regardless of background.

This study proposes some solutions to enhance the fairness and validity of the NUEEFL. Test designers should use DIF analysis on a frequent basis to find and fix biased items. This will make sure that all things operate the same for all genders, school types, and ethnicities. Second, inclusive educational policies and curricula should provide suitable resources for public schools and minority groups. In addition, educators should employ culturally responsive teaching approaches and train other teachers about cultural competence, gender sensitivity, and fair language instruction so that they can fulfill the needs of all of their students. The test should be piloted in a real-world setting to make sure it is reliable, legitimate, and fair for all test-takers, regardless of their background.

This research encompassed several limitations that could have influenced the outcomes. Initially, the participants in the study might not accurately reflect the diverse population of NUEEFL test-takers in Iran, as the data collection was limited to a selected group of schools due to practical constraints. Secondly, obtaining comprehensive demographic and performance data was not always feasible, complicating the exploration of the interconnections among specific subgroups. Thirdly, a more comprehensive understanding of the factors contributing to unequal item functioning could have emerged if the research had incorporated additional qualitative perspectives from educators or test-takers alongside the quantitative test data. Finally, while the sample size was adequate for DIF analysis, it was insufficient for a thorough examination of less prevalent ethnic groupings, potentially limiting the statistical capacity to detect minor biases. To address these challenges and yield more reliable and broadly applicable findings, subsequent research should consider employing larger sample sizes, diverse data sources, and more extensive sampling methods.

References

Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing, 13*(3), 280–297. https://doi.org/10.1177/026553229601300304

Ary, D., Jacobs, L. C., Irvine, C. K. S., & Walker, D. (2018). *Introduction to research in education*. Cengage Learning.

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

Baniasadi, A., Salehi, K., Khodaie, E., Bagheri Noaparast, Kh., & Balal Izanloo, B. (2023). Fairness in classroom assessment: A systematic review. *Asia-Pacific Education Researcher, 32*(1), 91–109. https://doi.org/10.1007/s40299-021-00636-z

Chalhoub-Deville, M. (2015). Testing context: Test-taker characteristics. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 405–418). Routledge.

Dadvar, N., & Tabatabaee-Yazdi, M. (2023). EFL learners' perceptions of fairness in classroom assessment and their cognitive test anxiety. *Journal of Research in Techno-based Language Education, 3*(4). https://doi.org/10.22034/jrtle.2023.413067.1095

Darabi Bazvand, A., & Rasooli, A. (2022). Students' experiences of fairness in summative assessment: A study in a higher education context. *Studies in Educational Evaluation, 72*, 101118.

https://doi.org/10.1016/j.stueduc.2021.101118

DeLuca, C. (2012). Preparing teachers for the age of accountability: Toward a framework for assessment education. *Action in Teacher Education, 34*(5–6), 576–591. https://doi.org/10.1080/01626620.2012.730347

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability, 28*(3), 251–272. https://doi.org/10.1007/s11092-015-9233-6

French, B. F. (2020). Test bias. In *Encyclopedia of quality of life and well-being research* (pp. 1–4). Springer. https://doi.org/10.1007/978-3-319-69909-7_3261-1

Geranpayeh, A., & Kunnan, A. J. (2007). Differential item functioning in terms of age in the Certificate in Advanced English examination. *Language Assessment Quarterly, 4*(2), 190–222. https://doi.org/10.1080/15434300701375758

Ghorbani, M. R. (2012). Controversy over abolishing Iranian university entrance examination. *Asian Education and Development Studies, 1*(2), 139–152. https://doi.org/10.1108/20463161211240181

Ghorbani, V., & Kianifard, S. (2024). Exploring ways of assessing intercultural competence: Introducing the bimodal assessment model. *Curriculum Research, 5*(4), 27–48. https://doi.org/10.71703/cure.2024.1129165

Haertel, E., & Herman, J. (2005). Historical perspective on validity arguments for accountability testing (CSE Report 654). University of California.

Hosseini, S. M. H. (2007). ELT in higher education in Iran and India: A critical view. *Language in India, 7*, 1–11.

Kamyab, S. (2007). Flying brains: A challenge facing Iran today. *International Higher Education, 47*. https://doi.org/10.6017/ihe.2007.47.7955

Khodi, A. (2020). *An appraisal of validity and dimensionality of B.A. Iranian University Entrance Examination* (Unpublished doctoral dissertation). University of Tehran.

Khodi, A., Alavi, S. M., & Karami, H. (2021). Test review of Iranian university entrance exam: English Konkur examination. *Language Testing in Asia, 11*(14), 1–10. https://doi.org/10.1186/s40468-021-00125-6

Kim, K. H., & Zabelina, D. (2015). Cultural bias in assessment: Can creativity assessment help? *International Journal of Critical Pedagogy, 6*(2), 130–148.

Kunnan, A. J. (2018). *Fairness and justice in language assessment*. Cambridge University Press. https://doi.org/10.1017/9781139649484

Mauldin, R. K. (2009). Gendered perceptions of learning and fairness when choice between exam types is offered. *Active Learning in Higher Education, 10*(3), 253–264. https://doi.org/10.1177/1469787409343191

Mazor, K. M., Hambleton, R. K., & Clauser, B. E. (1998). Multidimensional DIF analyses: The effects of matching on unidimensional subtest scores. *Applied Psychological Measurement, 22*(4), 357–367. https://doi.org/10.1177/014662169802200404

McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Blackwell Publishing.

Mislevy, R. J. (2018). *Sociocognitive foundations of educational measurement*. Routledge. https://doi.org/10.4324/9781315708302

Parviz, M. (2023). Reflecting on the consequences of the Iranian university entrance examination: A systematic-narrative hybrid literature review. *Discover Education, 2*(22). https://doi.org/10.1007/s44217-023-00046-x

Rasooli, A., Rasegh, A., Zandi, H., & Firoozi, T. (2022). Teachers' conceptions of fairness in classroom assessment: An empirical study. *Journal of Teacher Education, 74*(3), 260–273. https://doi.org/10.1177/00224871221130742

Rasooli, A., Zandi, H., & DeLuca, C. (2018). Re-conceptualizing classroom assessment fairness: A systematic meta-ethnography of assessment literature and beyond. *Studies in Educational*

Evaluation, 56, 164–181. https://doi.org/10.1016/j.stueduc.2017.12.008

Razmjo, S. A. (2006). Content analysis of specific questions of the English language test group of the national entrance exam of the country's universities. *Shiraz University Journal of Social Sciences and Humanities, 1*(46), 465–480.

Rezai, A., Alibakhshi, G., Farokhipour, S., & Miri, M. (2021). A phenomenographic study on language assessment literacy: Hearing from Iranian university teachers. *Language Testing in Asia, 11*, 1–18. https://doi.org/10.1186/s40468-021-00142-5

Russell, M., Madaus, G., & Higgins, J. (2009). *The paradoxes of high stakes testing: How they affect students, their parents, teachers, principals, schools, and society*. LAP.

Safari, P. (2016). Reconsideration of language assessment is a MUST for democratic testing in the educational system of Iran. *Interchange, 47*(3), 267–296. https://doi.org/10.1007/s10780-016-9282-4

Safari, P., & Rashidi, N. (2018). Democratic assessment as scales of justice: The case of three Iranian high-stakes tests. *Policy Studies, 39*(2), 127–144. https://doi.org/10.1080/01442872.2018.1428748

Shohamy, E., & Eldar, S. (2002). High stakes exams and washback: The case of the Bagrut in Israel. *Assessment in Education, 9*(3), 307–333. https://doi.org/10.1080/0969594022000027646

Tierney, R. (2016). Fairness in educational assessment. In M. A. Peters (Ed.), *Encyclopedia of educational philosophy and theory* (pp. 1–6). Springer. https://doi.org/10.1007/978-981-287-532-7_171-1

Tofighi, S., & Ahmadi Safa, M. (2023). Fairness in classroom language assessment from EFL teachers' perspective. *Teaching English as a Second Language Quarterly, 42*(2), 81–110. https://doi.org/10.22099/tesl.2023.46825.3173

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods, 18*(3), 320–334. https://doi.org/10.1037/a0032125

Xu, Y., & Brown, G. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education, 58*, 149–162. https://doi.org/10.1016/j.tate.2016.05.010

Yang, B. W., Razo, J., & Persky, A. M. (2019). Using testing as a learning tool. *American Journal of Pharmaceutical Education, 83*(9), 7324. https://doi.org/10.5688/ajpe7324\

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF)*. National Defense Headquarters.

Curriculum Research

Volume 6, Issue 3
Sept. 20, 2025

Identifying and Analyzing the Components of Educational Games Based on Social Skill Development for Preschoolers

Article Type:

Original Research

Authors:

Shilan Aslani[1]

Khalil Ghaffari[2]

ORCiD:0000-0003-1307-4925

Esmaeil Kazempour[3]

ORCiD:0000-0001-5546-2921

Article History:

Received: 2025.06.02

Accepted: 2025.09.08

Published: 2025.09.20

Abstract

In recent years, many educational experts have concluded that game is one of the key elements in learning various skills in children, especially social skills. The aim of the present study was to identify the components of games based on social skill development for preschools in order to design a desirable model. The present study was conducted using a qualitative approach, specifically the grounded theory method. Data were collected through 25 semi-structured interviews with preschool teachers and educators who had valuable experiences related to the research topic using purposive sampling (snowball), and were analyzed in three main stages: open, axial, and selective coding. Interview data analysis revealed five main categories which affect educational games based on social skill development model; namely, causal factors, contextual factors, intervening factors, strategies, and advantages. It is concluded that proper implementation of a game-based social skills curriculum and paying attention to its various elements can lead to a child's physical, mental, and sensory development.

Key Words: Educational Games, Preschoolers, Social Skill Development

1. Department of Educational Studies and Curriculum Planning, AR.C, Islamic Azad University, Arak, Iran. Email: shilan.aslani@iau.ac.ir
2. Department of Educational Studies and Curriculum Planning, ALG.C, Islamic Azad University, Aligoodarz, Iran (Corresponding Author). Email: ghaffari@iau.ac.ir
3. Department of Educational Studies and Curriculum Planning, TON.C, Islamic Azad University, Tonekabon, Iran. Email: e.kazempour@toniau.ac.ir

1. Introduction

Preschool refers to a two-year period that covers children aged 4 to 6 years under educational programs (Fathi, 2016). It is an early childhood education program for children before they enter formal schooling (Duan,2025). Preschool aims to prepare children for the transition to formal schooling by developing their social, emotional, cognitive, and physical skills through play-based learning and structured activities (Xovoxon,2024). This program is usually a child-centered program that follows a play and activity approach (Robinson, 2021). Education during this period has a significant impact on the child's all-round development, providing intellectual, verbal, social, emotional, and physical development (Tomé-Fernández et al., 2024). Attending preschool as an educational environment can have a significant impact on overcoming fear of unfamiliar and educational environments (Carneiro et al., 2024). The main focus in preschool is on developing and strengthening the child's important skills, and children are taught to learn basic concepts in a game-based manner. Social communication skills, motor skills, speech skills, listening skills, attention, and problem-solving skills are among the educational skills in preschool (Nurhasanah et al., 2025). Preschool education increases flexibility, cognitive skills, and discipline in the child (Ramírez et al, 2021). A child's emotional and cognitive development in preschool requires opportunities for the child to gain experience by encountering a variety of objects and places, encountering adults, other groups, and peers, and engaging in meaningful and enriching interactions with them (Jafari, 2014). The importance of the preschool years as a critical stage in learning and mental development is one of the important principles that has been repeatedly emphasized in the preschool period (Supriatna et al., 2024).

The issue of healthy mental development in the preschool period is very important (Mpella et al., 2019). Participating in preschool teaches children how to interact with their teachers and peers, and prepares them for basic concepts in elementary school. Attending preschool before entering elementary school helps children accept independence, eases the separation process from their parents, and strengthens their friendship skills (Haza et al., 2024). Social skills are taught to children during this period (Michelson et al., 2013). In the present era, a child who has proper mental development guarantees the future of any society (Maleki et al., 2019). Therefore, paying attention to the mental development of children is essential (Zheng et al., 2021).

One of the issues that has recently been raised in relation to the mental development of children is the issue of acquiring social skills, which is one of the essential elements in improving the mental development of children (Paz et al., 2021). Social skills are a large group of psychological, social, and interpersonal skills that help children make informed decisions, communicate effectively, develop coping and personal management skills, and lead healthy

lives (Heydari, 2017). Social skills develop a complex set of skills in children, including communication, problem-solving, decision-making, self-management, assertiveness, and peer interaction (De Mooij et al., 2020). It can be argued that social skills are a set of abilities that lead to the child being able to initiate positive and useful relationships and increase his or her friendship with others in order to adapt him or herself to environmental conditions (Barzegar, 2018; Sadri et al., 2016). These skills are very necessary for maintaining interpersonal relationships and satisfying these relationships (Lee et al., 2023).

Learning social skills socializes children, and through this, children learn and form norms, skills, motivations, attitudes, and behavior (Dryburgh et al, 2020). Social skills enable children to develop social relationships at different stages of life (Günindi, 2022). Social skills play an important role in developing children's ability to communicate with others and to behave in specific social situations (Zhu et al., 2021). Improving children's emotional and social communication largely depends on their social skills (Jabeen et al., 2022). Such skills lead to social adaptation in the child and have a lifelong impact on the child's life; therefore, childhood, especially preschool, is a sensitive period for developing social skills in children (Mulvey & Jenkins, 2021).

Acquiring social skills helps children begin positive social interactions and relationships with others and maintain these skills throughout their lives. People with poor social skills experience more difficulty in seeking social support and coping with stress when they are under stress (Symeonidou, 2023. Conversely, people with strong social skills usually have strong support networks that are called upon in times of crisis (Dong et al., 2023). Children's development is influenced by numerous factors, including the environment and society. The process of socialization is one of the most fundamental axes of personality development (Caron et al., 2023). Social skills are especially useful for children to develop good relationships with others, follow social rules, be responsible, participate and cooperate, and help others enjoy their rights (Aksoy & Baran, 2020).

Given the role of social skills as a powerful catalyst for the development of positive behavior, building life skills in the early years of life helps children navigate their social and emotional challenges, such as coping with emotional pain, conflict, peer pressure, and relationship issues (Fikri & Tegeh, 2022). Therefore, it is clear that the preschool period is a crucial period in the development of social skills. During this period, one of the essential elements is game. Children in preschool have a great desire to game (Tersi & Matsouka, 2020). Game plays a fundamental role in the development and upbringing of children and many cognitive and emotional issues can be transferred to children through game (Downey & Gibbs, 2020).

According to Yuniar (2021), game to children is what speech is to adults. Game is a natural, enjoyable, surprising, and mysterious activity. Game is a means of expressing feelings, building relationships, describing experiences, revealing desires, and self-fulfillment. (Józsa et al., 2023). Because children often have less verbal and cognitive ability to express their feelings, game is a natural and objective means of communication for them to cope with the world. The importance of using game in education is so great that countries such as Malaysia and Singapore have established academic disciplines in this field (Zamata et al., 2023). The necessity of using games, considering its effects and its use in the educational process to achieve easy, attractive and stimulating learning, seems more and more necessary than ever (Habibi et al, 2022).

The use of games involves a variety of students' senses, including visual and auditory, and provides an attractive learning environment for them (Saygin & Karakas,2021). Many social characteristics, including responsibility, communication with others, decision-making, and managing emotions, can be transferred to children through game (Caron et al., 2023).Through game, the child takes the first steps towards socialization (Van Berkel & Bosman, 2023). Through game, children learn how to be accepted by the group, discover the extent of their abilities needed to succeed in the group, and also become aware of how to use the abilities of others to meet their own needs (Ke et al., 2022).

Various studies have been conducted on games and social skills, including Tomé-Fernández et al., 2024, Nurhasana et al (2025), Haza et al. (2024), Dong et al. (2023), an Józsa et al. (2023). A brief review of the above research indicates that they can be categorized into 3 areas: (1) the effectiveness of game-based education in learning social skills in a specific group of students, including children with learning disabilities, autism, and hyperactivity ; (2) the effect of game-based education in learning social skills in adolescents and high school students; and (3), theoretical study of the components of game-based education in learning social skills. As is clear from the above classification, no independent research has been conducted to examine the components and elements of game-based education in learning social skills for preschool children. Considering this, it seems necessary to design and develop an appropriate model through which social skills can be taught to children using games. Therefore, the purpose of this research is to identify and analyz the components of game-based social skills education for preschools.To serve that end, the following research question was formulated:

- How do educational games contribute to the development of social skills among preschoolers, considering their causal and contextual factors, intervention-related elements, applied strategies, and resulting consequences?

2. Methodology

2.1. Design

In the present study, the grounded theory approach was used. The reason for using this method is that researchers usually use it to develop a specific theory about the phenomenon under study.

2.2. Participants

The participants in the present study were individuals who had valuable lived experiences related to the research topic. Data collection continued until theoretical saturation and 25 interviews were conducted. The participants consisted of professors and instructors who could provide the researcher with valuable information in the field of social skills training. In this study, purpusive sampling method (snowball) was used. The characteristics of the participants are presented in Table 1:

Table 1.

Characteristics of participants in the research

Number	Field of Study	Education	Gender	Age (Year)	Experience (year)
1	Psychology	Ph.D.	Female	50	12
2	Philosophy of education	Ph.D.	Male	61	23
3	Psychology	Ph.D.	Female	46	10
4	Curriculum Development	Ph.D.	Male	47	9
5	Elementary education	Ph.D.	Male	39	5
6	Educational psychology	Ph.D.	Female	55	18
7	Curriculum Development	Ph.D.	Male	48	9
8	Elementary education	Ph.D.	Female	57	19
9	Educational psychology	Ph.D.	Male	42	7
10	Elementary education	Ph.D.	Female	46	9
11	Elementary education	Ph.D.	Male	53	18
12	Curriculum Development	Ph.D.	Male	49	10
13	Curriculum Development	Ph.D.	Female	58	19
14	Elementary education	Ph.D.	Male	41	6
15	Elementary education	Ph.D.	Male	49	8
16	Curriculum Development	Ph.D.	Female	54	16
17	Curriculum Development	Ph.D.	Male	40	5
18	Philosophy of education	Ph.D.	Male	52	14

19	Philosophy of education	Ph.D.	Female	44	9
20	Curriculum Development	Ph.D.	Female	58	19
21	Curriculum Development	Ph.D.	Female	49	12
22	Psychology	Ph.D.	Male	42	5
23	Psychology	Ph.D.	Female	53	14
24	Educational psychology	Ph.D.	Male	56	17
25	Educational psychology	Ph.D.	Female	43	7

2.3. Instruments

Semi-structured interviews were used to collect the data. The interviews were conducted by one of the researchers, face-to-face and online. Each interview lasted between 40 and 45 minutes. All interviews were conducted in Persian. Before the interviews, consents were obtained from the interviewees to participate in the interview and they were assured that their information would remain confidential. To assess the validity and reliability of the interview questions, the questions were checked, modified, and approved by six professors in the relevant field. As a result, five The interview prompts were developed, as follows:

- What are the causal factors influencing the use of educational games for preschoolers' social skill development?
- What are the contextual factors involved in educational games for preschoolers' social skill development?
- What are the intervention-related factors in educational games designed to enhance preschoolers' social skills?
- What strategies are used in educational games to support preschoolers' social skill development?
- What are the consequences of using educational games for preschoolers' social skill development?

The interviews were audio-recorded and carefully recorded by the researcher. The "members' check" method was used to confirm the validity of the interview trnscripts. In this way, the written text was sent to the participants to confirm the accuracy of the content and to correct it if necessary.

2.4. Data Analysis

Data analysis was performed based on the instructions of Strauss and Corbin (2011), which includes three main stages of open, axial and selective coding. For validation, the coding and analysis of the interview data were reviewed by a colleague in curriculum studies, and agreement was observed regarding the results. Finally, the qualitative model of the

research was presented. To further ensure validity, the member-checking method was employed to validate the obtained model.

3. Findings

In order to answer the research questions, after operationalizing a total of 25 interviews with the research participants who were expert professors and directly related to the subject, the collected data were analyzed in the form of open coding. Doing the open coding process on these data led to the achievement of many concepts, which the researcher tried to reduce and categorize based on similarities and conceptual commonalities. The results of this coding are presented in the following tables:

- **What are the causal factors influencing the use of educational games for preschoolers' social skill development?**

The causal factors that are effective in educational games based on the development of social skills can be presented in Table 2.

Table 2.

Causal facors related to educational games based on social skills development

Causal Factors	1. The importance of the role of game in education	1.1. Effectiveness of game-based training 1.2. Institutionalization of game-based training among students 1.3. Objective and tangible results of game-based social skills training 1.4. Stakeholder demand for the use of games in social skills training
	2. Changing the identity of education	2.1. The disappearance of traditional education in teaching social skills to students 2.2. Attention to current realities and students' needs 2.3. Emphasis on educational innovations in social skills

Effectiveness of game-based training. In general, training that is based on entertainment is more effective and lasting. One participant stated that: *"In preschool, children have more objective thinking, so if we want to teach them social skills effectively, using games is very effective"* (Participant 4).

Institutionalization of game-based education among students. Today, society's expectations of education have changed; it is no longer limited to simply educating students as in the past. One of the participants noted that: *"The society we live in no longer seeks to send their children to school to learn addition and subtraction. Parents want their children to become good citizens and to grow up to be good human beings"* (Participant 24).

Objective and tangible results of game-based social skills training. Paying attention to skills is one of the most important tasks of education in the current era. One of the participants remarked that:

> *I would like to give an example: suppose we sent our child to school, and after several years of continuous education, he still lacked the skills to communicate with others. In my view, such an education should be criticized, because our goal is not merely to produce scientists and mathematicians, but to educate well-rounded individuals — something that is not possible without attention to a variety of essential skills* (Participant 18).

Stakeholder demand for the use of games in teaching social skills. The needs of today's society are very different from the past, and this has doubled the attention to life skills. One of the participants highlighted that:

> *In today's society, interpersonal problems have increased, and many people lack even the basic ability to communicate effectively with one another. Moreover, possessing knowledge alone is no longer sufficient as it may have been in the past; therefore, we need to reconsider and reform the very foundations of education* (Participant 9).

The disappearance of traditional education in teaching social skills to students. Traditional education is no longer responsive in the current situation and fundamental changes are needed. A participant argued that:

> *Traditional classroom education often focuses on the teacher presenting information while students passively receive it. In today's world, this approach is increasingly ineffective, as it does not equip individuals with the skills needed to function successfully in society* (Participant 11).

Education based on existing realities and students' needs. Effective education is one of the essential components in the current century. One of the participants acknowledged that:

> *Education can be sustainable only if it responds to the needs of its stakeholders. Many parents argue that the current system fails to adequately address their children's needs, partly due to insufficient emphasis on developing students' life skills* (Participant 21).

Emphasis on educational innovations in social skills.

One characteristic of knowledge is its perishability in today's age of incomplete or fragmented information. One of the participants mentioned that:

> *Knowledge is becoming obsolete rapidly, so relying solely on it is a major strategic*

*mistake in education. Instead, we must invest in lifelong skills, such as social skills. But can a skill like empathy, one of the core social skills, ever become obsolete (*Participant 19)?

- **What are the contextual factors involved in educational games for preschoolers' social skill development?**

The contextual factors for implementing an educational game based on the development of social skills in preschool can be presented in the following table:

Table 3.

Contextual factors *related to educational games based on social skills development*

Contextual factors	4. Values, beliefs, and attitudes	4.1. Teachers' values, beliefs, and attitudes 4.2. Parents' values, beliefs, and attitudes 4.3. Staff values, beliefs, and attitudes 4.4. Children's values, beliefs, and attitudes

Teachers' values, beliefs, and attitudes. Teachers' beliefs and attitudes play an important role in teaching game-based social skills to children. One of the participants stated that:

*Teachers and coaches, as facilitators of education, play a crucial role in whether children accept or reject certain forms of training. Fortunately, most coaches now recognize that developing social skills in children is only possible through creative play (*Participant 8*).*

Parents' values, beliefs, and attitudes. Parents, as key pillars of education, play a vital role in supporting game-based learning. One of the participants acknowledged that:

*Implementing training successfully is impossible without parental support. When teaching social skills through play, it is essential to consider the values and beliefs of the parents involved (*Participant 20*).*

Values, beliefs, and attitudes of staff. School staff and their attitudes and beliefs can play a major role in advancing game-based education. One participant stated admitted that:

*The school must approve any training program to be implemented, and this is especially important when teaching social skills through play, as even minor inconsistencies in this area can render the program ineffective (*Participant 14*).*

Children's values, beliefs, and attitudes. Children, as the primary beneficiaries of social skills, should be given special attention. One of the participants stated in this regard:

Any form of social skills training or games in this field must be grounded in children's beliefs and opinions; otherwise, it will be ineffective. If a child perceives that

*a game conflicts with their beliefs, they are unlikely to engage deeply with it. (*Participant 22*).*

- **What are the intervention-related factors in educational games designed to enhance preschoolers' social skills?**

The intervention-related factors in implementing educational games based on the development of social skills in preschool are presented in the following table:

Table 4.

Intervention-related factors *related to educational games based on social skills development*

Intervention-related factors	5. Environmental factors	5.1. Empathetic atmosphere in the school towards the use of games in teaching social skills
		5.2. Availability of facilities and infrastructure for the use of games in teaching social skill
	6. Programs and policies	6.1. Reality-based planning in the use of games in social skills education
		6.2. Dynamic and flexible policies in games and social skills
		6.3. Continuity and in-depth review of programs related to the use of games in social skills education

Empathetic school environment towards the use of games in teaching social skills. The school atmosphere has a significant impact on teaching social skills through games. One participant mentioned that:

*The prevailing atmosphere in a school can either facilitate or hinder social skills training based on modern educational methods such as games. When a school fails to embrace this approach and lacks coordination and synergy, success cannot be expected (*Participant 3*).*

Availability of facilities and infrastructure for the use of games in teaching social skills. Availability of facilities plays a significant role in teaching social skills through games. One participant emphasized that: "*Teaching social skills through games requires modern facilities and equipment, including laptops, tablets, appropriate hardware and software, and access to digital games" (*Participant 15*).*

Reality-based planning in using games in social skills training. One of the important principles in game-based social skills training is paying attention to existing realities. One of the participants believed that: "*Many goals set for teaching social skills are overly idealistic or unrealistic, making it difficult to effectively use games in developing these skills" (*Participant 17*).*

Dynamic and flexible policies on game and social skills. Policies that are defined in the

context of using play in teaching social skills should be flexible. One participant highlighted that:

> *By nature, games are flexible; therefore, any program designed in this field must incorporate that flexibility. A rigid, linear, top-down approach is unsuitable in this context (*Participant 9*).*

Continuity and in-depth consideration of programs related to the use of games in social skills training. The continuity of game-based social skills training programs is another crucial component. One of the participants stressed that:

> *Any program that is implemented sporadically and without continuity cannot achieve the desired effectiveness. When teaching social skills through games, the program must be sustained over a long period to be truly effective (*Participant 2*).*

- **What strategies are used in educational games to support preschoolers' social skill development?**

The strategies for implementing educational games based on the development of social skills in preschool can be presented in the following table.

Table 5.

Strategies for implementing educational games based on the development of social skills

Strategies	7. Attention to the importance of game and new methods	7.1. Using group and collaborative games in teaching social skills 7.2. Using games based on role-playing, modeling, and empathy in teaching social skills
	8. Negligence and inattention	8.1. Using stereotyped games 8.2. Using one-dimensional games 8.3. Using games devoid of creativity and originality

Using group and collaborative games in social skills training. Incorporating group games is one of the strategies commonly used in social skills training. One of the participants confirmed that: "*I believe group games are highly important, as playing games that require cooperation and participation can strengthen social skills such as teamwork and turn-taking" (*Participant 3*).*

Using games based on role-playing, modeling, and empathy in teaching social skills. Another effective strategy in teaching social skills is the use of empathy-based and role-playing games. One of the participants mentioned that: "*To effectively teach students social skills, they must be exposed to real-life situations. For example, through role-playing, each student can learn about social responsibilities (*Participant 19*).*

Using stereotyped games. One ineffective strategy in teaching social skills is the use of stereotyped or rigid games. One of the participants noted that:

> *Repetitive and uncreative games can lead to student boredom. Games should actively engage students, stimulating their emotions and feelings, which are essential components of social skills (*Participant 7*).*

Using one-dimensional games. Another negative strategy in the field of teaching social skills is the use of one-dimensional games. One of the participants admitted that:

> *Games that lack clear principles and steps, or that focus solely on a single skill, are not well-suited for teaching social skills. Effective games should engage the student's understanding, emotions, and cognition (*Participant 16*).*

Using games devoid of creativity and originality. Another ineffective strategy in teaching social skills is the use of games that lack originality. One of the participants acknowledged that:

> *A game that does not allow students freedom of action and flexibility cannot effectively teach social skills. The game should provide opportunities for both individual and collective student initiative (*Participant 17*).*

- **What are the consequences of using educational games for preschoolers' social skill development?**

The consequences of implementing an educational game based on the development of social skills in preschool can be presented in the following table:

Table 6.

Consequences of implementing an educational game based on the development of social skills

Consequences	9. Individual	9.1. Emotional and Emotional Development of Children
		9.2. Development of Communication Skills of Children
	10. Social	10.1. Development of Responsibility and Accountability in Children
	11. Behavioral	11.1. Attention and Respect for Differences in Society
	12. Cultural	12.1. Avoidance of Abnormal Behaviors and Lawfulness
		12.2. Development of Empathy and Solidarity Skills with Others
		12.3. Cultivation of Positive Thinking and Critical Thinking

Children's emotional and sentimental development. One of the positive outcomes of teaching social skills based on game is children's emotional development. One of the

participants stressed that:

> *When children are immersed in a game, they are compelled to respond to the opinions, beliefs, and values of their peers in various situations. These responses help them recognize their own feelings and emotions, promoting self-awareness, adjustment, and personal growth (*Participant 22*).*

Development of children's communication skills. One of the direct effects of game-based social skills training is the development of communication skills. One of the participants highlighted that:

> *During play, children learn to establish various forms of eye contact, verbal, and nonverbal communication with their peers, as well as how to coordinate within a communication network and interpret messages. In essence, play conveys the complexities of communication to children (*Participant 6*).*

Development of responsibility and accountability in children. Development of responsibility is another outcome of teaching social skills through playing. One participant stated that:

> *Group and role-playing games primarily teach children how to be responsible toward their peers and accountable for the tasks assigned to them during the game. In doing so, they learn accountability to others and the importance of adhering to established rules (*Participant 20*).*

Attention to and respect for differences in society. Another positive outcome of teaching social skills through game is understanding and respecting the differences of others. One of the participants emphasized that:

> *When children participate in social skills training through games, they gradually recognize that their peers differ in language, race, intelligence, talent, and emotions. Through this process, they learn that such differences should be respected (*Participant 4*).*

Avoiding abnormal behaviors and respecting the rules. Reducing and avoiding abnormal behaviors is another result of game-based social skills training. One of the participants stressed that:

> *Many individuals who commit crimes often lack social skills and have not learned how to interact effectively with others. Teaching social skills through games helps convey these essential interaction skills to individuals (*Participant 1*).*

Developing empathy and solidarity skills with others. Developing a sense of

interdependence with others and society is one of the outcomes of teaching social skills through games. One of the participants mentioned that:

> *When children are taught to play in groups and with diverse peers, they are essentially learning that a fulfilling life involves sharing their experiences with others, listening to others' stories, and engaging empathetically with those experiences (*Participant 17*).*

Cultivating positive thinking and critical thinking. Cultivating positivity and constructive criticism of others is another result of game-based social skills training. One of the participants stated that:

> *During play, children learn to identify the flaws and shortcomings of others while simultaneously preparing to cooperate constructively with them. In this process, they develop the ability to critically evaluate others while maintaining a positive attitude toward collaboration (*Participant 22*).*

Based on the research findings, the conceptual model presented in Figure 1 shows the relationships between the main categories resulting from the qualitative analysis process.

4. Discussion

The results of the research indicated that the elements of educational games based on the development of social skills in preschool can be explained in the form of five factors: causal, contextual, intervention-related, strategies, and consequences.

As indicated, issues such as the effectiveness of game-based education, the institutionalization of game-based learning among students, and the objective, tangible results of game-based social skills training are critical areas for consideration. In recent years, societies have increasingly recognized the importance of social skills and have mobilized resources to enhance various individual competencies. In this regard, Barzegar (2018) reported that education related to social skills has grown more than fourfold in the past decade. This type of education is no longer limited to universities but has expanded to other institutions, including schools, many of which have developed and implemented specialized programs to equip students with social skills.

Figure 1.

Model of game-based social skills education for preschools

Another key factor in game-based social skills training is its effectiveness and measurable outcomes. Education in recent decades has shifted toward gamification. For example, Scott et al. (2022) found that training delivered to children through games is twice as effective as traditional methods. Indirect training through games is particularly beneficial for children with an object-oriented mindset. One important issue raised in the present study is the decline of traditional training. The reality is that modern society is not receptive to superficial instruction or the mere transfer of information, especially in crucial areas such as social skills. Given the nature of social skills, the ultimate goal is skill acquisition, which can be best achieved through practical approaches such as games.

The second influential category in educational games aimed at developing social skills in preschool is contextual factors, which include the values, beliefs, and attitudes of teachers, parents, school staff, and children. Among these, teachers play a particularly important role. Teachers' beliefs and attitudes can either facilitate or hinder the use of educational games for social skills development. As facilitators, teachers with a positive attitude toward educational games can significantly influence their successful implementation. In this regard, Heydari et al. (2017) conducted a study examining teachers' willingness to teach social skills through games and found that more than two-thirds supported this approach.

The next important factor is the attitudes and beliefs of parents, who complement and reinforce in-school education. Ohara et al. (2019) argued that implementing social skills training through play is not possible without parental support, as some school-based games are extensions of activities at home and remain ineffective without parental cooperation. Another key element is the school's educational staff. As the implementers of educational programs, they play a crucial role in determining the effectiveness—or ineffectiveness—of game-based social skills training. Smolkowski et al. (2022) emphasized that complete coordination among all school stakeholders is essential for teaching social skills through games, ensuring that they share a unified vision on the matter.

Issues such as fostering an empathetic school atmosphere toward the use of games in teaching social skills, ensuring the availability of facilities and infrastructure, and adopting reality-based planning are critical to the success of game-based social skills training. An empathetic school atmosphere is particularly influential. The prevailing environment in the school and the quality of relationships among its members play a vital role in teaching social skills, especially through games. In this regard, Jafari (2014) found that the synergy created by an empathetic atmosphere in schools increases the participation of various educational stakeholders in games related to social skills, thereby enhancing the program's success.

Another factor influencing game-based social skills training is the availability of

facilities and equipment. Since the implementation of various games requires both hardware and software resources, any shortage of equipment can hinder such training. Polat et al. (2022) identified the lack of necessary facilities and infrastructure as one of the primary causes of failure and ineffectiveness in educational programs. Finally, idealism and disregard for practical realities represent another major challenge in social skills training. Efforts in this field must be grounded in real-world conditions to ensure feasibility and effectiveness.

Issues such as the use of group and collaborative games, as well as games based on role-playing, modeling, and empathy, are central to teaching social skills. The findings of the present study indicate that educational games aimed at developing social skills in preschool can be approached in two ways: acceptance or rejection. The adoption of innovative methods—such as group and cooperative games, role-playing, modeling, and empathy-based activities—falls within the accepting approach to educational games for social skills development.

Given that social skills are a relatively new and emerging area in education, traditional teaching methods are unlikely to be highly effective. Teaching social skills cannot be achieved merely by delivering information or lectures, as these skills must be applied across all areas of students' lives. Therefore, instruction in this area requires a multifaceted approach. Simply labeling a lesson as "social skills" will not produce meaningful results; rather, such training should be delivered through engaging methods, including digital games. Teaching social skills through games is both attractive and practical, fostering higher levels of student engagement and increasing interest in the subject.

The consequences of social skills training include children's emotional and sentimental development, enhanced communication skills, increased responsibility and accountability, greater attention to and respect for societal differences, avoidance of abnormal behaviors, and the promotion of law-abiding conduct. Numerous studies have emphasized these positive outcomes. For example, Maleki et al. (2019) found a direct and significant relationship between social skills training and the development of creativity, critical thinking, and positivity in students. Social skills provide students with various tools for thinking and reflection, contributing to their intellectual growth. Similarly, Bakola et al. (2019) reported that social skills training increases desirable behaviors, reduces delinquent tendencies, and fosters greater responsibility among students. Salavera et al. (2022) argued that social skills training is a key element in promoting empathetic living and fostering solidarity. Such training is also believed to enhance interactions between children, thereby improving their verbal and interactive skills.

5. Conclusion

The findings of this study suggest that the primary goals of game-based social skills education should include fostering responsibility, promoting children's socialization, and developing empathy and solidarity. Such training can influence an individual's sense of self-sufficiency, self-confidence, and self-esteem, thereby playing a significant role in encouraging normative behavior. By engaging in game-based social skills activities, children experience a greater sense of acceptance within their social environment, which in turn facilitates their integration into society.

The study also highlights that game-based social skills education is influenced by numerous interrelated micro- and macro-level components. Neglecting any of these components can undermine the effectiveness of the program. For instance, without a supportive and welcoming attitude toward such initiatives, even the most well-designed programs may fail. Similarly, overlooking the interactive and creative aspects of social skills can render a program ineffective. Game-based social skills training can be likened to a complex structure in which the proper arrangement of each component determines its functionality and impact.

Based on the results of this study and existing literature, the following implications are offered for policy and practice:

1. Shift from purely cognitive instruction to applied learning. Given the evolving nature of social skills education, programs should move beyond the delivery of abstract information and focus on creating practical, experience-based learning opportunities.
2. Consider stakeholder values and attitudes. The values and beliefs of teachers, parents, and other stakeholders play a critical role in the success of game-based social skills education. Program design should account for these contextual realities.
3. Avoid generalization and idealism. Due to the diversity of contexts in which social skills programs are implemented, educators and policymakers should focus on developing objective, context-specific, and operational strategies rather than overly broad or idealistic ones.
4. Strengthen teacher preparation. Since teachers have a central role in promoting game-based social skills education, training programs should ensure that educators are well-prepared and skilled in using games as a teaching tool.
5. Invest in digital infrastructure. Given the demonstrated value of digital games in teaching social skills, preschool institutions should be equipped with the necessary hardware, software, and technological support.

6. Incorporate diverse and multidimensional games. Social skills are multifaceted, and their instruction should draw on a variety of game types to address different skill dimensions effectively.
7. Enhance parental involvement. Parents play a vital role in reinforcing social skills outside of school. Strategies should be developed to actively engage parents in game-based social skills education initiatives.

This study faced several limitations that should be acknowledged when interpreting its findings. Access to some research participants proved challenging, which restricted the ability to conduct interviews with certain individuals relevant to the topic. Similarly, reaching parents of children and other stakeholders involved in social skills education was limited, narrowing the scope of perspectives included in the analysis. Another challenge arose from the cautiousness of some participants, who, due to personal or contextual reasons, refrained from disclosing certain realities and experiences related to social skills.

From a methodological standpoint, the study relied solely on the grounded theory approach without incorporating other complementary research methods that might have provided additional depth or triangulation. Furthermore, interviews were used as the primary data collection tool, with limited use of other methods for gathering information. These constraints may have influenced the breadth and richness of the data obtained.

References

Aksoy, P., Baran, G. (2020). The effect of story telling-based and play-based social skills training on social skills of kindergarten children: An experimental study. *Education and Science*, 45(204), 157-183. DOI:10.15390/EB.2020.8670

Bakola, L. N., Rizos, N. D., Drigas, A. S. (2019). ICTs for Emotional and Social Skills Development for Children with ADHD and ASD Co-existence. *International Journal of Emerging Technologies in Learning*, 14(5).121-137. DOI:10.3991/ijet.v14i05.9430

Barzegar, K.(2018) The impact of playing with Lego on social skills of preschool children. *Journal of Child Mental Health*. 2017; 4(3): 130-142. URL: http://childmentalhealth.ir/article-1-213-en.html

Carneiro, T., Carvalho, A., Frota, S., Filipe, M. G. (2024). Serious games for developing social skills in children and adolescents with autism spectrum disorder: A systematic review. *In Healthcare* .12(5), 508-517. DOI:10.3390/healthcare12050508

Caron, V., Barras, A., van Nispen, R. M., Ruffieux, N. (2023). Teaching social skills to children and adolescents with visual impairments: A systematic review. *Journal of Visual Impairment & Blindness*, 117(2), 128-147. DOI:10.1177/0145482X231167150

De Mooij, B., Fekkes, M., Scholte, R. H., Overbeek, G. (2020). Effective components of social

skills training programs for children and adolescents in nonclinical samples: A multilevel meta-analysis. Clinical child and family *psychology review*, 23(2), 250-264. DOI:10.1007/s10567-019-00308-x

Dong, X., Burke, M. D., Ramirez, G., Xu, Z., Bowman-Perrott, L. (2023). A meta-analysis of social skills interventions for preschoolers with or at risk of early emotional and behavioral problems. *Behavioral Sciences*, 13(11), 940-951. DOI:10.3390/bs13110940

Downey, D. B., Gibbs, B. G. (2020). Kids these days: are face-to-face social skills among American children declining? *American journal of sociology*, 125(4), 1030-1083. DOI:10.1086/707985

Dryburgh, N. S., Khullar, T. H., Sandre, A., Persram, R. J., Bukowski, W. M., Dirks, M. A. (2020). Evidence base update for measures of social skills and social competence in clinical samples of youth. *Journal of Clinical Child & Adolescent Psychology*, 49(5), 573-594. DOI:10.1080/15374416.2020.1790381

Duan, S. (2025). AI-Assisted Social Skills Training for Preschool Children with ASD: A Study on Early Psychological Interventions. *The Journal of Mind and Behavior*, *46*(1), 436-454. DOI:10.1037/h0100390

Fathi, K (2016). Basic Principles and Concepts of Curriculum Planning. *Tehran: Professors' Science Publishing* House.

Fikri, M. T., Tegeh, I. M. (2022). The impact of social skills training on the ability to cooperate in early childhood. *Indonesian Values and Character Education Journal,* 5(1), 32-41. DOI:10.23887/ivcej.v5i1.44227

Günindi, Y. (2022). The Effect of Online Education on Children's Social Skills during the COVID- 19 Pandemic. *International Electronic Journal of Elementary Education,* 14(5), 657-665. DOI:10.26822/iejee.2022.270

Habibi Asgarabad, M., Steinsbekk, S., Wichstrøm, L. (2023). Social skills and symptoms of anxiety disorders from preschool to adolescence: A prospective cohort study. *Journal of child psychology and psychiatry*, 64(7), 1045-1055. DOI:10.1111/jcpp.13787

Haza, B., Gosling, C. J., Ciminaghi, F., Conty, L., Pinabiaux, C. (2024). Research Review: Social cognition and everyday social skills in children and adolescents with attention-deficit/hyperactivity disorder: a meta-analysis of case–control studies. *Journal of Child Psychology and Psychiatry*, *65*(9), 1245-1254. DOI:10.1111/jcpp.14006

Heydari, R., Mohammadi, R., Radmehr, H. (2017). Evaluation the effectiveness of emotional self-efficacy training and problem-solving skills on increasing social adjustment and coping strategies in opioid abusers. Nova *Journal of Medical and Biological Sciences*, 6(1),28- 41. DOI:10.1016/j.sbspro.2013.06.499

Jabeen, A., Mahmood, Z., Saleem, S., Subhan, S. (2022). Development of social skills scale for

school children. *Journal of Postgraduate Medical Institute*, 36(2), 63-9. DOI:10.54079/jpmi.36.2.2946

Jafari, A (2014). The effect of educational games on the social development of preschool children. *Quarterly Journal of Educational Psychology*, 10.(33), 71-85. https://jep.atu.ac.ir/article_165.html?lang=en

Józsa, K., Oo, T. Z., Borbélyová, D., Józsa, G. (2023). Social skills development in children aged 4-8 years: a cross-country comparison of Hungary and Slovakia. In ICERI2023 Proceedings (pp. 3326-3335). IATED. DOI:10.21125/iceri.2023.0870

Ke, F., Moon, J., Sokolikj, Z. (2022). Virtual reality–based social skills training for children with autism spectrum disorder. *Journal of Special Education Technology,* 37(1), 49-62. DOI:10.1177/0162643420945603

Lee, J. H., Lee, T. S., Yoo, S. Y., Lee, S. W., Jang, J. H., jin Choi, Y., Park, Y. R. (2023). Metaverse-based social skills training programme for children with autism spectrum disorder to improve social interaction ability: an open-label, single-centre, randomised controlled pilot trial. *EClinicalMedicine*, 61(3).39-51. DOI:10.1016/j.eclinm.2023.102072

Maleki, M.; Chehrzad, M.M.; Kazemnezhad, E.; Mardani, A.; Vaismoradi, M. (2019). Social Skills in Preschool Children from Teachers' Perspectives. *Journal of Children education* 7(6), 64-83. DOI:10.3390/children6050064

Michelson, L., Sugai, D. P., Wood, R. P., Kazdin, A. E. (2013). Social skills assessment and training with children: An empirically based handbook. Springer Science & Business Media. DOI:10.5935/1980-6906/psicologia.v21n3p159-181

Mpella, M., Evaggelinou, C., Koidou, E., Tsigilis, N. (2019). The Effects of a Theatrical Play Programme on Social Skills Development for Young Children with Autism Spectrum Disorders. *International Journal of Special Education*, *33*(4), 828-845. DOI:10.31219/osf.io/pqktw

Mulvey, N., Jenkins, L. (2021). Language skills as predictors of social skills and behaviors in preschool children. *Contemporary School Psychology*, 25(4), 503-514. DOI:10.1007/s40688-020-00281-1

Nurhasanah, E., Sudatha, I. G. W., Santosa, M. H., Suartama, I. K. (2025). Group Guidance Strategies to Develop Social Skills in Adolescents: A Literature Review. *Journal of English Language and Education*, *10*(4), 799-812. DOI:10.31004/jele.v10i4.985

Ohara, R., Kanejima, Y., Kitamura, M., P. Izawa, K. (2019). Association between social skills and motor skills in individuals with autism spectrum disorder: a systematic review. *European journal of investigation in health, psychology and education*, 10(1), 276-296. DOI:10.3390/ejihpe10010022

Paz, T. R. U., Mosqueira, M. A. A., Ugarte, M. G. M. (2021). Systematic review of strategies for

developing social skills in education. *Sinergias educativas*, 6(3).41-62 https://biblat.unam.mx/en/revista

Polat, Ö., Sezer, T., Akyol, N. A. (2022). Collaborative learning with mind mapping in the development of social skills of children. *Participatory Educational Research*, 9(1), 463-480. DOI:10.17275/per.22.25.9.1

Ramírez-Montoya, M. S., Andrade-Vargas, L., Rivera-Rogel, D., Portuguez-Castro, M. (2021). Trends for the future of education programs for professional development. *Sustainability*, 13(13), 72-44. DOI:10.3390/su13137244

Robinson, P. F., Fontanella, S., Ananth, S., Martin Alonso, A., Cook, J., Kaya-de Vries, D., Saglani, S. (2021). Recurrent severe preschool wheeze: from prespecified diagnostic labels to underlying endotypes. *American journal of respiratory and critical care medicine*, 8(5), 523-535. DOI:10.1164/rccm.202009-3696OC

Sadri M, Zare Bahramabadi M, Ghiasi M. (2016) .Effectiveness of attachment-oriented therapy on the quality of attachment and social skills in children with oppositional defiant disorder. *Journal of Child Mental Health*. 3(2), 21-30. https://childmentalhealth.ir/article-1-423-fa.html

Salavera, C., Usán, P., Quilez-Robres, A. (2022). Exploring the effect of parental styles on social skills: The mediating role of affects. *International Journal of Environmental Research and Public Health*, 19(6), 32-95. DOI: 10.3390/ijerph19063295

Saygin, M., Karakas, H. (2021). The Effect of Social Skills-Based Activities on Primary School Students' Critical Thinking and Empathic Tendencies. *International Journal of Curriculum and Instruction*, 13(3), 2789-2812. https://ijci.net/index.php/IJCI/article/view/780

Scott, S., Gray, T., Charlton, J., Millard, S. (2022). The impact of time spent in natural outdoor spaces on children's language, communication and social skills: a systematic review protocol. *International Journal of Environmental Research and Public Health*, 19(19), 12- 38. DOI:10.3390/ijerph191912038

Smolkowski, K., Walker, H., Marquez, B., Kosty, D., Vincent, C., Black, C., Strycker, L. A. (2022). Evaluation of a social skills program for early elementary students: We Have Skills. *Journal of Research on Educational Effectiveness*, 15(4), 717-747. DOI:10.1080/19345747.2022.2037798

Supriatna, E., Hanurawan, F., Eva, N., Rahmawati, H., Yusuf, H. (2024). Analyzing factors affecting social skills development among students in Indonesian schools. *Islamic Guidance and Counseling Journal*, *7*(1). 35-61. DOI:10.25217/0020247447100

Symeonidou, S. (2023). Effects of Physical Activities on Social Skills and Well-being in Autistic Children: *A systematic literature review from* 2012-2023. *Journal of educational psychology.*6(3).71-84. DOI:10.1007/s40489-023-00418

Tersi, M., Matsouka, O. (2020). Improving social skills through structured playfulness program in preschool children. *International Journal of Instruction*, 13(3), 259-274. DOI:10.29333/iji.2020.13318a

Tomé-Fernández, M., Aranda-Vega, E. M., Ortiz-Marcos, J. M. (2024). Exploring social skills in students of diverse cultural identities in primary education. *Societies*, *14*(9), 158. DOI:10.3390/soc14090158

Van Berkel, A. J. M., Bosman, A. M. (2023). Sociable in Someone Else's Shoes: A Review of Drama as Social Skills Training for Elementary School Children. *International Journal of Research in Education and Science*, 9(1), 165-187. DOI:10.22219/raden.v4i2.84104

Xovoxon, F. (2024). Strategies for enhancing social skills in early childhood education through developmental learning approaches. *International Journal of Pedagogics*, *4*(12), 91-96. DOI:10.1186/s40359-024-02280-w

Yuniar, N. F. (2021). Childhood social skills and attachment styles: A systematic review. OPTIMA: *Journal of Guidance and Counseling*, 1(1), 36-54. DOI:10.4314/afrrev.v6i5.4189

Zamata, P. M. C., Saavedra, L. R., López, R. M. G. (2023). Educational program to strengthen school coexistence and social skills in student. *International Journal of Instruction*, 16(3), 77-94. DOI:10.29333/iji.2023.1635a

Zheng, L. R., Oberle, C. M., Hawkes-Robinson, W. A., Daniau, S. (2021). Serious Games as a Complementary Tool for Social Skill Development in Young People: A Systematic Review of the Literature. *Simulation & Gaming*, 52(6), 686-714. DOI:10.1177/10468781211031283

Zhu, Y. T., Li, X., Jiao, D. D., Tanaka, E., Tomisaki, E., Watanabe, T., Anme, T. (2021). Development of social skills in kindergarten: A latent class growth modeling approach. Children, 8(10), 870-891. DOI:10.3390/children8100870