Author Details:

1. PhD Graduate of TEFL, Allameh Tabataba'i University, Tehran, Iran. Omidostad007@gmail.com

2. PhD Graduate of TEFL, IKIU, Qazvin, Iran. (Corresponding Author)

hheidari67@gmail.com

3. MA in TEFL, University of Guilan, Rasht, Iran. Khoshsorour@gmail.com

Applying Classical Test Theory to Test Critical Period Hypothesis: The Case of Iranian EFL Learners' Grammatical Performance in a Subject-Verb Test on Agreement

Omid Ostad[1] ; Hadi Heidari[2*] ; Azadeh Khoshsoroor[3]

Abstract: The current paper pursues two aims: to employ Classical Test Theory (CTT) to develop and refine a test, and to determine the impact of the age factor on L2 grammar among Iranian EFL learners, providing evidence for the claim that CTT best evaluates the consistency of measurements in age-related studies. The study examines the performance of adults or teenagers as pre- and post-puberty learners, in the grammar test devised through CTT, to conclude whether younger learners or older ones perform better on subject-verb agreement. Based on random sampling, twenty-three male participants, including 12 adults and 11 adolescents, upper-intermediate native speakers of Persian, took part in the study. They were tested on a 60-item test of grammar (on subject-verb agreement), which has been developed using the CTT model, as it fits small sample sizes and can be used for understanding complex issues (e.g., age-related issues). Other than devising a test on agreement using CTT as a model, the results revealed that post-puberty learners scored better than their younger counterparts, partly because they are cognitively more mature. However, age is not the one and only factor in second language learning, and other factors like individual differences, contextual factors, and methodological rigor should be taken into account.

Keywords: Designing Test; EFL; Pre- and Post-puberty; Subject-verb Agreement; Tense Agreement.

1. Introduction

According to Bachman (1990), Classical test theory (CTT) is a body of related psychometric theory that predicts outcomes of psychological testing, such as the difficulty of items or the ability of test-takers. In other words, it is a theory of testing stating that a person's obtained score from a given test is the sum of the true score and the error score in that test. CTT aims to understand and improve the reliability of psychological tests. As Fulcher and Davidson (2007) contend, despite criticisms, CTT remains widely used. The reason is that perhaps because CTT can provide us with detailed information regarding the function of each test item and the distribution of the test scores. It can help us to monitor and improve the test reliability. Brown (2007) argues that Classical test theory assumes that each person has a *true score*, which would be obtained if there were no errors in measurement. A person's true score is defined as the expected number-correct score over an infinite number of independent administrations of the test. It is assumed that the *observed score equals true score* plus some *error*: $X = T + E$.

CTT was first developed and predominated in measurement circles during the early to middle twentieth century and is "a measurement theory which consists of a set of assumptions about the relationships between actual or observed test scores and the factors that affect these scores, which are generally referred to as error". At the heart of CTT is the assertion that an observed score is determined by the actual state of the unobservable variable of interest plus error contributed by all other influences on the observable variable. In CTT, observed scores (i.e., the examinees' actual scores on a test) contain errors (unsystematic effects due to factors not being measured), which in turn contribute to error variance (unsystematic variation in the scores that is due solely to random errors). True scores are hypothetical representations of the scores that would result if there were no errors in measurement. CTT fits small sample sizes and it can be applied to understanding complex issues, including age-related challenges. CTT focuses on the reliability and validity of test scores, which can be crucial in assessing psychological, cognitive, or health-related factors that evolve with age (DeMars, 2018; Hambleton & Jones, 1993). For example, Geisinger (2013) indicated that it can help evaluate the consistency of measurements in age-related studies, such as cognitive decline or physical health assessments, with a focus on age-related issues.

Age has been one of the decisive factors in L2 learning and acquisition. Most researchers agree that age can cause differences in L2 learning; however, other factors

such as motivation, aptitude, intelligence, and style and strategies are to be taken into account. What is still controversial is the performance of the different age groups, that is believed not to be similar.

The relationship between age and learning has been widely recognised by the Critical Period Hypothesis (CPH), first brought up by Lenneberg (1967). CPH has been defined as 'the period during which a child can acquire language easily, rapidly, perfectly, and without instruction' (Richards & Schmidt, 2002, p.145). It extends from birth up to somewhere around puberty, and learning during this period is much faster and easier than post post-puberty period (Larsen-Freeman & Long, 2014).

The focal idea around CPH, originally proposed by Lenneberg, suggests that there is a certain age around puberty, which is known as the optimal age for language learning, after which language acquisition ability declines. This idea has led to the widely accepted belief that "the younger, the better" for language acquisition (Rahman et al., 2017; Wang, 2015). However, some researchers (e.g., Liu, 2023) worked on different language-related areas to challenge this notion and stated that while younger learners may initially acquire language skills more quickly, older learners often achieve higher levels of proficiency over time. Such findings indicated the complex relationship between age and language acquisition, which violated CPH theory and showed the need for more studies in the field.

Hyltenstam (2003) distinguishes the critical period from the 'sensitive period'; in CPH, maturation is thought to take place and come to an end within an early phase of the life span, abruptly set off from the rest at a specific age (puberty or earlier) (p. 556). However, in the sensitive period, 'the sensitivity does not disappear at a fixed point; instead, it is thought to fade away over a longer period, perhaps covering later childhood, puberty and adolescence' (p. 556). Kumaravadivelu (2006) says the critical period represents a well-defined 'window of opportunity,' whereas the sensitive period represents a progressive inefficiency of the organism. SLA theories and research have explained the impact of age in second language acquisition. Interestingly, research by Hartshorne (2018) indicated that younger learners better acquire grammatical structures due to heightened neural plasticity during this critical period, and as their age increases, the ability to learn grammar declines, which is often attributed to reduced neurobiological adaptability.

One of the areas of grammar acquisition in the process of language acquisition for Iranian learners is the domain of agreement. It is a notion that is present in English but

not in the Persian language. In acquiring English as a second/foreign language, learners whose first languages do not have such a rule as subject-verb agreement or other types of agreements show a lot of errors in this regard. Quirk et al (1985) define the notion of subject-verb agreement as the concordance between the subject and the verb, in that in the relation between a subject and a verb, one of them shows a special characteristic that is in agreement with the displayed feature of the other. However, there exist other forms of agreement in the English language. Generally, the notion of agreement can be defined as two or more grammatical elements agreeing with one another (Svartvik & Sager, 1986).

Regarding the acquisition of grammatical features and the impact of maturational constraints in SLA, Birdsong and Molis (2001) provided some evidence to note that the acquisition of grammatical features tends to decline with age for those who begin learning a second language after puberty. This was a significant finding which confirmed the findings of CPH theory. However, some contrasting data came from Hakuta et al. (2003), who analyzed second language attainment among immigrants and observed a steady decline in language acquisition ability throughout the lifespan, which questioned the existence of a critical period. Such contrasting findings showed that although language learners' age may influence their language learning process, the effects are not as absolute as the CPH implies, and it needs more investigation.

Hence, the aim of this study may be defined as designing a good test on agreement based on the CTT model of testing, firstly, to measure the grammatical proficiency of language learners in an EFL context, and secondly, to investigate the effect of age on grammatical awareness of Iranian English language learners. The present study seeks to compensate for the scarcity of research in this area, especially in Iran. Thus, this study attempts to bridge this gap by devising a test on agreement using CTT as a model to develop and refine the test and to employ the test to weigh the difference in the performance of the two groups of learners. It can be used to shed light on the difference between younger and older learners' grammatical performance in a subject-verb test on agreement according to the theory of CPH.

Overall, the current body of research indicated that the connection between age-related factors and grammar acquisition remains underexplored, particularly in relation to how these factors influence test reliability and validity. Therefore, more studies are needed to analyze measurement errors and true score variance through CTT, which can guide the development of age-appropriate grammar tests. This study was aimed at

finding out if adults are better than younger learners regarding their grammatical knowledge of subject-verb agreement and tense agreement. Since it was claimed that post puberty cohorts can outperform pre puberty cohorts in the areas of L2 such as syntax and morphology. The effort in this paper, therefore, would be focusing on syntax, and more specifically agreement.

2. Literature Review

2.1. Theoretical Framework

Singleton (2004) indicated that one of the most popular of the notions in relation to age and L2 learning: "the younger, the better and the older, the better" (p. 31), focusing on pronunciation and linguistic features. Singleton (2003) indicated that young learners do better in the long run in L2 lexical acquisition. There are a number of studies in support of the younger the better, and others in favor of the older the better.

Some researchers believe that younger learners have some advantages over older ones in L2 learning. They claim that young children can learn more easily and more quickly compared to older ones (Ellis, 2008; Larsen-Freeman, 2009; Mayberry & Lock, 2003). Johnson and Newport (in Lightbown & Spada, 2008) conducted a study on immigrants and found that those who began learning later had less native-like language abilities. Patkowski (1982) studied 67 immigrants with high education in the US from different backgrounds. Thirty-three of the subjects had immigrated to the United States before the age of puberty, and 34 of them after the age of 15. He examined their spoken language ability and concluded that those who had immigrated before the critical period were better language learners (Patkowski, 1982, p. 59).

Also, Carroll (1963) hypothesized that the ability to reach a native-like accent decline towards puberty. Oyama (1976) studied 60 male Italian immigrants in the US, ranging in age from 6 to 20. She found that a native-like accent is better in young learners than older ones. In terms of listening comprehension, she stated that 'those who arrived after the age of 16 showed markedly lower comprehension scores than the native speaker' (Oyama, 1982, p. 23).

The view that younger language learners perform better in language acquisition is further supported by other researchers (e.g., Rahman et al., 2017; Wang, 2015).

There are a number of other researchers who advocate the older the better, namely short-term experimental researchers who drew their assumptions based on adults' gains in immersion programmes.

Fathman and Precup (1989) tested speaking in a study where children scored better in pronunciation, adults were better in syntax. Ekstrand (1982) tested pronunciation, diction, and listening, reading, written, and oral production and came to the conclusion that better mental development results in better L2 learning. Lightbown and Spada (2008) stated: older learners have a high level of problem-solving and metalinguistic abilities.

Krashen et al. (1979) studied the morphosyntax of children and claimed that older learners pass the early L2 grammatical stages faster than younger ones. Krashen's well-known 3 proposals regarding morphosyntax are as follows:

Adults proceed through early stages of syntactic and morphological development faster than children (where time and exposure are held constant); older children usually acquire the language at a faster pace than younger children (again here time and exposure are kept constant); acquirers who begin natural exposure to second languages during childhood generally achieve higher second language proficiency than those beginning as adults.

Meanwhile, Liu (2023) worked on different language-related areas and highlighted a major difference between younger and older language learners. He underscored the fact that while younger learners may initially acquire language skills more quickly, older learners often achieve higher levels of proficiency over time.

2.2. Empirical Studies on Agreement

Agreement is a common problem among second/foreign language learners. Many researchers have tried to show this issue from different perspectives. In a study on errors related to agreement, Tafida and Okunade (2016) showed that the main problematic areas involve pronouns, verb tense, nouns and errors involving intervening variables. In another study, Maros et al. (2007) examined a group of Malaysian school students' essays. The results showed that the most frequent grammatical error of students, after six years of learning English, were the use of articles, subject-verb agreement and copula 'be' respectively.

Hoshino et al. (2010) worked on the subject-verb processing in a second language by bilinguals and concluded that only highly proficient bilinguals are responsive to conceptual number in a second language.

Additionally, Johnson and Newport (1989) studied pre- and post-puberty second language learners by testing their grammar proficiency. The results showed that pre-puberty learners had an advantage over post-puberty learners. So, they concluded that

the critical period for language acquisition extends its effects to second language acquisition as well.

In a recent study, Major (2014) investigated the role of learners' age of onset of acquisition in second language proficiency and suggested that the younger the age of onset of acquisition, the greater the proficiency in L2.

2.3. CTT

There are numerous advantages to working within the CTT framework when doing test development as cited by Brown (in Fulcher & Davidson 2012).

First, the concepts of CTT are widely taught and understood. Second, CTT concepts are relatively easy to learn, use, and explain. The very useful descriptive statistics (e.g., the mean, standard deviation, range, etc.) and item analysis statistics (especially item facility and item discrimination) can all be calculated fairly easily in spreadsheets or widely available computer programs. Third, the CTT model fits many measurement needs, particularly for the development of proficiency and placement measures that can aid admissions, program comparisons, and placement decisions. Fourth, right or wrong, CTT accommodates and allows interpretation of examinee scores of 0% and 100% and item facility estimates of .00 and 1.00.

Meanwhile, Brown and Bailey (2008) point out that many educators in testing are trained in CTT, and numerous courses are within the topic area of CTT. Brown (cited in Fulcher & Davison, 2012) argues that CTT is easy to learn and implement, practical and accurate, and can provide the test developers with useful information regarding item analysis. Regarding the impact of CTT on recent models (e.g., IRT), Holland and Hoskens (2003) argue that CTT is not in direct contrast with IRT as IRT is a mere extension of CTT.

CTT is found to be instrumental in evaluating the reliability of language tests, particularly when considering complex issues like age differences among learners. CTT focuses on the consistency of test scores by analyzing measurement error and true score variance (Brown, 2016). In the context of age differences, CTT can help identify whether a language test is equally reliable across different age groups. For example, it can assess if older learners face more variability in test scores due to cognitive or memory-related factors, or if younger learners exhibit higher consistency due to greater neuroplasticity (Brown, 2016).

It is important to note that Stevens (2004) criticized the methodologies employed in previous studies, arguing that the sampling and estimation procedures could significantly

impact the findings related to the CPH. Therefore, the application of CTT to studies of the CPH necessitates careful consideration of the instruments used to measure language proficiency and the age of onset of language learning. This highlights the importance of not only ensuring that the tests used to evaluate language acquisition are both reliable and valid, but also necessitates conducting more studies in different subject matters and contexts. This can justify the findings as the conclusions drawn from such studies can vary dramatically based on these factors. For example, the role of individual differences, such as motivation and the degree of immersion in the target language environment, complicates the relationship between age and language acquisition.

Research by Singleton and Leśniewska (2021) emphasizes that factors beyond mere biological maturation, including psychological variables and the learner's context, play crucial roles in determining language learning outcomes. Though the study highlights the multifaceted nature of second language acquisition, considering sociocultural, psychological, and cognitive factors, and discusses naturalistic and formal learning contexts, it places greater emphasis on formal instruction, potentially overlooking nuances in naturalistic language acquisition. It also argues that the CPH is not proven and unfalsifiable, which may limit the scope for constructive debate and further empirical investigation

This perspective aligns with findings from Lin et al. (2023), who noted that the impact of the critical period may vary depending on the specific language skills being assessed, such as semantic awareness and pronunciation. However, focusing on semantic radicals and lexical tones in Chinese may limit the applicability of findings to other languages or linguistic features. The study also faced challenges in isolating the impact of the critical period from other factors like motivation, exposure, and cognitive strategies.

Overall, different scholars have expanded on the application of CTT by addressing how age-related cognitive and linguistic changes affect test reliability and validity, particularly in the domain of language learning and assessment. Among them, the works of Bachman and Brown (1996), Hyltenstam (2003), and Hakuta et al. (2003) have contributed significantly to understanding CTT and its application to age-related issues, particularly in the context of language acquisition and assessment. The study by Bachman and Brown (1996) emphasized the importance of test reliability and validity, and core principles of CTT in language assessment, with a focus on how age-related cognitive and linguistic differences can influence test performance, highlighting the need for age-

appropriate test designs. Hyltenstam's (2003) study delved into age-related differences in second language acquisition, particularly focusing on the CPH and how age impacts linguistic proficiency and how CTT can be used to measure these differences reliably. The work by Hakuta et al. (2003) explored the interplay between age, cognitive development, and language learning and its contribution to understanding how age-related factors influence test outcomes and how CTT can help in designing assessments that account for these variables.

However, an in-depth exploration of these studies shows that they have not addressed some important issues, like a focus on grammatical awareness in the context of CPH. While CTT has been widely used to assess test reliability and validity, there is a lack of targeted research that specifically applies CTT to measure grammatical awareness, and most studies focus on broader language skills rather than grammatical features like agreement. Moreover, many studies do not adequately differentiate between age groups when applying CTT to test CPH, which limits the understanding of how grammatical awareness evolves across different age ranges and how it aligns with the critical period. Finally, such studies often overlook the impact of cultural and linguistic backgrounds on grammatical awareness, which is evident in multicultural societies like Iran. This gap makes it challenging to generalize findings across diverse populations and also difficult to interpret how the nuances in CTT can be related to language learners' ability to process grammatical features (e.g., the agreement feature), impacted by age-related issues.

To answer such discrepancies regarding the findings on the age factor in grammar acquisition, the present study is guided by the following question in order to provide us with insights as to how different the results of the comparison using the test might be:

1. Which age group, the adults or the teenagers, will perform better in the grammar test devised through CTT?

3. Methodology

3.1. Participants

There were two groups of participants in the study. The first group (12 undergraduates of English literature from Guilan University) took part in the piloting phase of screening the test item, and the second group (23 EFL learners in a language institute in Rasht) took part in the second phase of the study. Based on random sampling, the selected sample consisted of 23 male students of English in language institutes, with 12 adults and 11 adolescents. Since the study was exploratory, the researchers aimed to recruit a

small sample size to have more control over participants' characteristics and some of the erroneous data to be able to build upon the existing knowledge and generate insights for future research rather than provide definitive conclusions. The participants were selected through random sampling; however, some of their main characteristics, like age, language proficiency, and exposure to the English language outside the classroom, were taken into account. Prior to the study, the participants took the OPT version 2, and out of 30 participants, 24 were selected as being at the intermediate level, in order to control the impact of language proficiency and have more homogenous participants. Their L1 was Persian, and they were at the upper intermediate level of L2 proficiency. They were randomly divided into two cohorts. Cohort 1's sample consisted of 12 students of 20 to 28 years, and Cohort 2's sample had 11 students from 10 to 13. The average age of cohort 1 was 23.4, and the average age of cohort 2 was 12.09. These age groups were chosen to differentiate the two groups based on the CPH. The test was originally designed for 24 students, but one of the participants withdrew from the second cohort. All participants had a maximum of 4 years of training in English before the test.

3.2. Materials

A test was designed on subject-verb agreement and tense agreement, which consisted of 60 items. the revised final version is in the appendix section of this study. The test was adapted and revised from TOEFL practice materials, which have shown to be a robust tool, and its reliability and validity were confirmed through rigorous research and development processes conducted by Educational Testing Service (ETS), which showed consistent results across different administrations and in academic and professional contexts. It was revised based on the participants' characteristics and their needs. Then, it was piloted by 10 participants, which showed to be reliable, and its validity was confirmed by two experts in the field. The first 10 items were sentences with one grammatical error in each. Three other words or phrases were underlined (distracters) along with the error (the right answer). Participants were asked to mark their answers in the answer sheet. Items 11 to 30 were multiple choice items, and items 31 to 50 were part of a cloze test with responses in multiple choice format. In the last ten items, participants were asked to supply the correct form of 'to be' verbs.

3.3. Procedures and analysis

The test was first administered as a pilot at Guilan University. The subjects of the test were freshmen. The allocated time for the test was 45 minutes. The test was composed of 30 multiple-choice items. 20 items in a cloze passage, and 10 short-answer questions.

The 12 students, 2 females and 10 males, took the test, and the results and mathematical computations and tabulations are listed in detail as follows:

First, the measures of central tendency were obtained, and then the measures of variability.

Next, the item facility choice distribution was conducted. Item discrimination and finally reliability was calculated. Ultimately the results were interpreted and the problematic test items were either removed or modified. There was no pretest, because the researcher believed that firstly the participants were homogenized through the placement test. Secondly, the questions in the test would have been exposed and the memory effect could have come into play.

The main research methods used in CTT are to include at least item analysis (item facility, item discrimination, and distractor efficiency analysis), reliability estimates, the standard error of measurement, and various validity analyses (Bachman & Palmer, 1996). The current study followed the same procedure.

3.3.1. Calculating the measures of central tendency

- The mode:

  (17.25, 17, 16, 15.50, 14, 13.50, 13.25, 12.50, 12.25, 11.50, 11, 8.50)

  Since the mode is the most frequent score, but no score is repeated in this set of scores, *there's no mode*.

- The median:

  The score at the 50th percentile is the median, so *the median is 13.50*.

- The mean, or the mathematical average: $\bar{X}$.

-

$$\bar{X} = \frac{\sum X}{N} = \frac{176.75}{13} = 13.59. \qquad (1)$$

3.3.2. Calculating measures of variability

- Range:

  The difference between the biggest and the smallest score: $17.25 - 8.50 = 8.75$

Table 1.

Scores and calculated variance.

Square of the difference $(X - \bar{X})^2$	Difference from the mean $(X - \bar{X})$	Variance X
$(3.66)^2 = 13.39$	$17.25-13.59 = 3.66$	17.25
$(3.41)^2 = 11.62$	$17- 13.59 = 3.41$	17
$(2.41)^2 = 5.80$	$16- 13.59 = 2.41$	16
$(1.91)^2 = 3.64$	$15.50- 13.59 = 1.91$	15.50
$(0.91)^2 = 0.82$	$14.50- 13.59 = 0.91$	14.50
$(0.41)^2 = 0.16$	$14- 13.59 = 0.41$	14
$(-0.09)^2 = 0$	$13.50- 13.59 = -0.09$	13.50
$(-0.34)^2 = 0.11$	$13.25- 13.59 = -1.34$	13.25
$(-1.09)^2 = 1.18$	$12.50- 13.59 = -1.09$	12.50
$(-1.34)^2 = 1.79$	$12.25- 13.59 = -1.34$	12.25
$(-2.09)^2 = 4.36$	$11.50- 13.59 = -2.09$	11.50
$(-2.59)^2 = 6.7$	$11- 13.59 = -2.59$	11
$(-5.09)^2 = 25.9$	$8.50- 13.59 = -5.09$	8.50

SUM $= 0.08$ SUM $= 74.87$

- Variance:

$$\frac{\sum(X - \bar{X})^2}{N - 1} = \frac{74.87}{13 - 1} = 6.23 \quad (2)$$

- Standard Division:

$$S = \sqrt{V} = \sqrt{6.23} = 2.49 \quad (3)$$

3.3.3. Item facility

There were 60 items whose facility are listed in the following table:

Table 2.

Item Facility of the designed test's items

items	IF	items	IF	items	IF
1	0.61	21	0.84	41	0.84
2	0.69	22	0.67	42	0.46
3	0.61	23	0.84	43	0.76
4	0.38	24	0.53	44	0.76
5	0.76	25	0.69	45	0.46
6	0.46	26	0.46	46	1
7	0.46	27	0.76	47	0.92
8	0.69	28	0.69	48	0.53
9	0.76	29	0.69	49	0.46
10	0.53	30	1	50	0.84
11	0.84	31	0.84	51	0.92
12	0.84	32	0.92	52	0.92
13	0.53	33	0.64	53	0.69
14	1	34	0.69	54	0.30
15	0.61	35	0.84	55	0.69
16	0.92	36	0.84	56	0.84
17	0.53	37	0.69	57	0
18	0.46	38	1	58	0.30
19	0.76	39	1	59	0.76
20	0.61	40	0.84	60	0.07

Items below 0.37 and above 0.63 needed to be revised.

3.3.4. Choice distribution

The choice distribution related to the designed test was listed in the following table.

Table 3.

Choice distribution of the designed test.

Item	key	Choice distribution 1	2	3	4	Item	key	Choice distribution 1	2	3	4
1	2	0	8	3	1	26	3	7	0	6	0
2	3	1	1	9	1	27	2	0	10	0	3
3	3	1	0	8	1	28	1	9	3	0	0
4	3	0	6	5	0	29	1	9	4	0	0
5	2	1	10	1	1	30	1	13	0	0	0
6	3	1	1	6	3	31	2	11	1	0	1
7	3	2	2	6	1	32	2	1	12	0	0
8	1	9	2	1	0	33	2	6	6	0	1
9	1	10	0	1	1	34	2	0	9	0	4
10	4	1	1	3	7	35	2	2	11	0	0
11	2	0	11	0	2	36	2	2	11	0	0
12	4	0	1	1	11	37	2	3	9	0	0
13	4	0	0	6	7	38	1	13	0	0	0
14	4	0	0	0	13	39	1	13	0	0	0
15	1	8	1	1	3	40	1	11	2	0	0
16	3	1	0	12	0	41	1	11	0	0	0
17	4	1	4	0	7	42	1	6	6	0	0
18	3	0	5	6	0	43	1	10	1	0	2
19	3	0	0	10	3	44	4	2	1	0	10
20	4	0	3	2	8	45	2	6	6	0	1
21	1	11	1	0	0	46	2	0	13	0	0
22	4	1	1	2	8	47	1	12	1	0	0
23	1	11	1	0	0	48	2	3	7	0	1
24	4	1	3	2	7	49	4	0	5	2	6
25	2	1	2	3	0	50	1	11	2	0	0

3.3.5. Item discrimination

The participants were divided into a high and a low group. They obtained scores as follows:

Table 4.

High and Low groups of the designed test's results

	score	Number of cornet response
High group	17.25	52
	17	51
	16	48
	15.50	47
	14.50	44
	14	42
Low group	13.25	39
	12.50	38
	12	36
	11.50	35
	11	33
	8.5	26

The item discrimination of every one of the items was computed and listed as follows (in this case, a very promising ID):

$$e.g : ID_{(1)} = \frac{5-2}{(1/2)/2} = \frac{3}{6} = 0.50 \quad (4)$$

So this item, as an example, does discriminate between the two groups, based on the following threshold for good item discrimination is often set as follows (Angoff, 1972):

Minimal Quality: An ID of around 0.10 is considered minimally acceptable.

Good Quality: An ID of 0.20 or above is generally regarded as good.

Strong Quality: An ID of 0.30 or higher is considered excellent.

Table 5.

Item Discrimination of the Designed Test

Items	ID	Items	ID	Items	ID
1	0.50	21	0	41	0.33
2	0.33	22	016	42	0.33
3	0.16	23	0	43	0.33
4	-	24	0.50	44	0
5	0.16	25	0.16	45	0.33
6	-	26	0.33	46	-1
7	0.16	27	0.16	47	0.16
8	0.50	28	0	48	0.16
9	0.50	29	0.50	49	-
10	0.33	30	0	50	0.16
11	0.16	31	0	51	0.33
12	0.33	32	0.16	52	0.16
13	0.33	33	0.50	53	0.16
14	0.33	34	0.33	54	0.66
15	0.33	35	0.16	55	0.33
16	0	36	0.33	56	0.33
17	0.16	37	-	57	0.33
18	0.16	38	0.16	58	0
19	0	39	0	59	0
20	0.83	40	0	60	0.50
	0		0		-
	-				0.16
	0.16				

3.3.6. Reliability

In order to calculate the reliability of the test, we should first obtain the coefficient of correlation between the two sets of scores:

$$r = \frac{N(\Sigma \times y) - (\Sigma x)(\Sigma y)}{\sqrt{[N(\Sigma x^2) - (\Sigma x)^2][N(\Sigma y^2) - (\Sigma y)^2]}}$$

$$= \frac{13(2553.88) - (176.75)*183.75)}{\sqrt{13(2478.38) - (17675)^2][13(2493.8) - (183.75)^2}}$$

$$= \frac{722.62}{\sqrt{(978.38)(1344.66)]}} = \frac{722.62}{1146.99} = 0.63$$

(5)

Thus, the coefficient of correlation is <u>0.63.</u> Since this mathematical value is close to 1, we can say there is a high degree of positive correlation.

3.4. Data analysis

According to the results shown in the item facility table, items such as items 6, 7, 10, and 24 had an ideal degree of facility ranging from 0.46 to 0.53. However, there were items which were too easy or too difficult, such as 14 or 30 whose facility equalled 1, which is too easy. And items such as 57, whose facility equals zero, show too much difficulty. Therefore, these problematic items were either revised through consultation with standardized tests or replaced with more standard ones.

Based on the choice distribution table, Item 6 is an ideal one since almost 50% of students selected the right option, which was number 3. However, there were items such as 38 or 39 in which 100% answered correctly which shows the distractors were not effective. Or in item 18, 50% chose the right answer while the other 50% chose a distractor, which indicates that the distractor was so close to the right choice that it attracted their attention. Based on the information obtained from the choice distribution table, the non–standard items were either revised or replaced with more standard ones.

Considering the table of item discrimination, some items are highly standard since they fully discriminated between low and high group. Amongst these are items 1, 33, 59, and 7 yet there are items which are not discriminating at all; for instance item 30 whose ID equals Zero. Items of such were modified or replaced with more discriminating ones.

In case of reliability it should be pointed out that two tests were administered to the same group. The results of comparing the two sets of scores show that the tests were at some acceptable degree of reliability.

4. Results

After the process of development and revison of the test, it was carried out through the piloting phase, and then, participants took the test and their scores are presented in table 6:

Table 6.

Both cohorts' scores

No	Section 1 (1-10)	Section 2 (11-30)	Section 3 (31-50)	Section 4 (51-60)	Aggregation
#1	3.5	2	1.5	3.5	10.5
#2	5	2.5	2	5	14.5
#3	3	3.5	3.5	4	13.5
#4	4.5	4.5	5	5	19
#5	4.5	5	4.5	5	19
#6	4	4	4	3	15
#7	4.5	4.5	4	5	18
#8	4	3.5	3.5	4	15
#9	4.5	5	5	5	19.5
#10	4	4	5	4	17
#11	3	2.5	4	3.5	13
#12	3	2.5	4	2.5	12
#13	4	4.5	4.5	4	17
#14	3.5	3	3	3.5	13
#15	4.5	5	2.5	4	15.5
#16	3	2	2	1.5	8.5
#17	3	3	3	2.5	11.5

No.					
#18	4	4.5	5	3.5	17
#19	4	5	5	4.5	18.5
#20	3	2.5	2	1.5	9
#21	3	3	4	2	12
#22	4.5	4	3	5	16.5
#23	4.5	4.5	3	3	15

Students were numbered in alphabetical order, given a number to each—to protect their right of anonymity, based on the issue of the confidentiality, provided along the informed consent to the participants—and it is shown in 'table 6'. The points received in each section were organised separately and then added up and calculated out of the total of 20 points.

Table 7.

Cohort 1 scores

No.	Section 1 (1-10)	Section 2 (11-30)	Section 3 (31-50)	Section 4 (51-60)	Aggregation
#3	3	3.5	3.5	4	13.5
#5	4.5	5	4.5	5	19
#6	4	4	4	3	15
#7	4.5	4.5	4	5	18
#9	4.5	5	5	5	19.5
#10	4	4	5	4	17
#11	3	2.5	4	3.5	13
#13	4	4.5	4.5	4	17
#18	4	4.5	5	3.5	17
#19	4	5	5	4.5	18.5
#22	4.5	4	3	5	16.5
#23	4.5	4.5	3	3	15

Table 7 shows the table in which, the points received by adults (cohort 1) is displayed. The average score of cohort 1 is 16.58.

Table 8.

Cohort 2 Scores

No.	Section 1 (1-10)	Section 2 (11-30)	Section 3 (31-50)	Section 4 (51-60)	Aggregation
#1	3.5	2	1.5	3.5	10.5
#2	5	2.5	2	5	14.5
#4	4.5	4.5	5	5	19
#8	4	3.5	3.5	4	15
#12	3	2.5	4	2.5	12
#14	3.5	3	3	3.5	13
#15	4.5	5	2.5	4	15.5
#16	3	2	2	1.5	8.5
#17	3	3	3	2.5	11.5
#20	3	2.5	2	1.5	9
#21	3	3	4	2	12

The average score of cohort 2 is 12.77 which is lower than cohort 1. And also, as shown in table 8, in learning vocabulary, grammar, writing, and speaking it was cohort 1, adults, who had a better performance. The following table (9) displays the section-based results of the two cohrots in average points.

Table 9.

Average scores

Average scores	Section 1 (1-10)	Section 2 (11-30)	Section 3 (31-50)	Section 4 (51-60)	Aggregation
Cohort 1	4.04	3.50	4.2	4.45	16.58
Cohort 2	3.63	3.05	2.9	3.18	12.77

As tables 9 and 10 show, adults in cohort 1 scored better in the test. Although teenagers were expected to be better in at least one or two sections of the test, the test results showed something completely different.

Table 10.

Descriptive Statistics

	Mean	SD	Stand Error Mean	Variance
Cohort 1	16.58	2.08	0.60	4.35
Cohort 2	12.77	3.07	0.92	9.46

As Table 11 shows, the level of significance was 0.02 and alpha was 0.05, which shows that there was a statistically significant difference between the groups.

Table 11.

t-test for Equality of Means of Adults and Teenagers

		Mean Difference	Std. Error Difference	Sig. (2-tailed)
Test	Equal variances assumed	3.81	-.99	.002
	Equal variances not assumed	3.81	-.99	.002

5. Discussion and Findings

Initially, the findings of the study contributed to devising a test on agreement using CTT as a model. Secondly, the findings revealed that post-puberty learners scored better than their younger counterparts partly because they are cognitively more mature. This may be due to the length of the period in which the students were supposed to learn the new language forms and structures; since adults have fewer problems learning in a short time, while teaching children/ teenagers requires more time. First of all, adults are faster learners due to some more advanced mental abilities, such as analysis or the ability in learning more abstract concepts. So we can conclude that their better performance could be at least because of their cognitive development.

Although the test showed that it was the adults who were better at learning language, it shouldn't necessarily mean that they will always remember what they have learnt. It is true that adults learn faster than children, but they might not be able to keep up with them in the long run. Teenagers, on the other hand, are better learners since they can maintain what they have learnt much longer, it takes longer to teach them though.

There were two major groups of researchers regarding age and language learning. The first group (including Ellis, 2008; Larsen-Freeman, 2009; Mayberry & Lock, 2003; Rahman et al., 2017; Singleton, 2003, 2004; Singleton & Leśniewska, 2021; Wang, 2015) stated that younger learners possess high learning potentials. On the other hand, some academics (including Dekeyser, 2012; Lightbown & Spada, 2008; Liu, 2023) favoured exactly the opposite. Post puberty learners are cognitively more mature, therefore better at skills requiring complex problem-solving abilities (e.g., vocabulary learning, grammar, writing, reading comprehension), whereas pre puberty learners performed significantly better at areas which require muscular flexibility, such as native-like accent, pronunciation, and prosodic features.

The findings can be attributed to the fact that adults and post-puberty learners often perform better than teenagers and pre-puberty learners in grammar tests, such as those on Subject-Verb Agreement, for several reasons: for example, Benítez-Burraco (2025) found that adults and post-puberty learners have more developed cognitive abilities, including better abstract reasoning and metalinguistic awareness, which are crucial for understanding grammatical rules.

Meanwhile, Dörnyei and Ushioda (2021) expanded on the role of motivation and learning Strategies and contended that adults often have higher motivation and employ

more effective learning strategies compared to younger learners, which can enhance their performance in grammar tests.

Moreover, Muñoz and Cadierno (2021) stated that adults may have had more exposure to the target language and opportunities to practice, leading to better performance in tests.

Hartshorne et al. (2018) acknowledged the role of CPH which suggests that younger learners are better at acquiring language naturally; however, it also indicates that older learners may excel in explicit learning tasks, such as grammar tests, due to their ability to focus on rules and patterns.

Research suggests that children and adults L2 learners pass through different developmental states in second language learning. Learning depends on the cognitive maturity and neurological factors as well. VanSickle and Ferris (2005) have shown the relation between age and SLA as "One of the dangers of the emphasis on critical periods is that it prompts us to pay too much attention to when learning occurs and too little attention to how learning might best occur" (p. 105).

Despite these findings, some opposing findings within the literature should also be stated. While DeKeyser (2000) acknowledges that cognitive maturity contributes to better performance in language learning among post-puberty learners, he provides some contrasting evidence, suggesting that adults and post-puberty learners may perform worse than teenagers or pre-puberty learners in grammar tests. This may be attributed to factors like reduced neuroplasticity, which affects implicit learning abilities, and the influence of the CPH, which posits that younger learners are better at acquiring language naturally (Hartshorne et al., 2018).

6. Conclusion

The study contributed to devising a test on agreement using CTT as a model. Moreover, it revealed that post-puberty learners scored better than their younger counterparts partly because they are cognitively more mature. The findings indicated that age is not the one and only factor in second language learning, and other factors like individual differences, contextual factors, and methodological rigor should be taken into account. As reported by Lightbown and Spada (2008), learning depends on learners' characteristics and the environment. Their findings suggested that older learners have a higher level of problem-solving and metalinguistic abilities than younger learners. However, there are contrasting views and perspectives regarding how children and adults acquire a foreign or second language. Adults naturally find themselves in such situations

that demand more complex language and expression of more complicated ideas, whereas children lack the pressure and maturity in second language learning. Although age is highly associated with the concept of critical period in many research studies, it is to be noted that many other factors are also in play. In summary, while the Critical Period Hypothesis provides a framework for understanding age-related effects in second language acquisition, its application must be nuanced. Classical Test Theory can aid in evaluating the reliability and validity of the measures used in this research, ensuring that conclusions drawn about the CPH are based on robust evidence.

There exist a number of implications for educationalists and test developers. Firstly, it is to be noted that every test in order to be fairly designed and well-developed requires a rigorous process of drafting and revising of individual items. Furthermore, caution must be employed in designing teaching and testing materials for different age groups as the display different performances and qualities in answering the same set of questions in the present study. Adults and post-puberty learners outperformed teenagers and pre-puberty learners in the grammar test on subject-verb agreement, due to some features like their advanced cognitive development, greater metalinguistic awareness, and ability to apply explicit learning strategies effectively. Their motivation and exposure to language can also contribute significantly to their success in such structured assessments. However, more studies are needed to provide a comprehensive picture of the different important factors in the study.

A number of controversial issues related to second language acquisition and the critical period hypothesis can be addressed. Age is not the one and only factor in second language learning. However, factors related to the age, for instance learning opportunities, individual differences, the motivation to learn, and learning styles, are also very determining variables that could affect the rate of second language learning in various developmental stages of the learners. It is hoped that future research will shed more light on the unknown areas of language learning and age, and also on other skills and components of language using tests, questionnaires, and other instruments concocted and refined based on CTT, or IRT. The interplay of age, individual differences, contextual factors, and methodological rigor underscores the complexity of language acquisition and the need for continued exploration in this field.

Future studies can develop grammar tests that account for the cognitive and linguistic strengths of different age groups, ensuring fairness and inclusivity. More longitudinal studies can be used to provide explicit grammar instruction and practice

opportunities for younger learners to bridge the performance gap. They can also account for the impact of diverse backgrounds on grammar test performance, which can affect generalizability and take linguistics, psychology, and education to refine testing methods and better understand age-related differences.

Appendixes

Appendix A.

A 60-item test on Agreement

A. Directions: Identify the one underlined word or phrase, (1), (2), (3), or (4), which must be corrected or rewritten. Mark your answers on the answer sheet.

1. The ability to conceal themselves by camouflage enable some otherwise defenseless animals to survive.

2. The information on the various types of wasps and bees in the area were useful to environmentalists who were fighting the use of pesticides.

3. Compared with the number of paid holidays enjoyed by most employees in the company, three weeks of vacation seem generous.

4. Sponges, with a structural organisation like that of a colony of one-celled animals, they stand at the lowest level of the animal kingdom.

5. The large bird house together with the numerous birdfeeders under the eaves attract a considerable number of different species in the summer.

6. Since snake eggs are tough and baby snakes can survive alone at birth, neither the eggs or newborn snakes need protecting by the mother.

7. Digitalis is a drug which is prepared from the seeds and leafs of a plant with the same name and is used as a cardiac stimulant.

8. Benjamin Franklin, a famous American statesman, author, and scientist, he was in 1709 and lived to the age of eighty-four.

9. The most famous alumnus of the college were invited to participate in the graduation ceremony and related activities scheduled for late May.

10. At present, advertising is one of the most strictly regulated industry in the US.

B. Directions: choose the best answer (1), (2), (3), or (4). Then mark your answers on the answer sheet.

11. You may not see him now, and…….

    1) also Reza doesn't   2) neither may Reza

    3) so may Reza   4) Reza doesn't either

12. Ali played tennis and………

    1) so could his sister   2) so does his sister

    3) so was his sister   4) so did his sister

13. She doesn't know French and………

    1) her sons doesn't either   2) her sons aren't either

    3) neither don't her sons   4) her sons don't either

14. I don't drive too fast,………… my father does.

    1) since   2) and   3) neither   4) but

15. If it hadn't been for my parents, I …………to university

    1) would never have gone   2) will never go

    3) could never go   4) would never go

16. If you been there………, I would have seen you.

    1) have   2) would have   3) had   4) will have

17. The car will run well unless it ……….. overloaded.

    1) will be   2) had been   3) had   4) will have

18. "Should I boil the egg or fry them for you?" "……. way is good. I like eggs."

    1) all   2) both   3) either   4) neither

19) "The injured player could only watch the match"

    "He wished he ……………in the team."

    1) might have played   2) might play

    3) could have played   4) could play

20. Can you have another room…………in your house?

    1)build   2) to build   3)builded   4)built

21. Wallpaper in addition to new curtains …………been ordered.

    1) has   2) have   3) is   4) will

22. Algeria as well as Tunisia and Libya ……….sent students to the United States

    1) had   2) have   3) has   4) 1,3

23. Three hours ……………a long time to take on the homework right now.

    1) seems   2) seem   3) will seem   4) would seem

24. The company's earnings ……………. increased last year.

    1) have   2) has   3) had   4) ………….

25. Police …………….. that Thomas is in Brazil, although his exact whereabouts ……………… unknown.

    1)believes, remain   2) believe, remains

    3)believe, remain   4) believes, remains

26. Staff …………..that the new computer system …………..led to greater levels of stress in their work.

    1) says, has   2) say, have   3) say, has   4) says, have

27. None of the players…………. injured last week.

    1) is   2) was   3) are   4) were

28. The statistics in that report on oil production ……………… incorrect.

    1) is   2) are   3) be   4) been

29. Five and six ……………. eleven.

    1) is   2) are   3) be   4) been

30. Since five thousand miles ……………. too far to travel, the family have decided not to go to Florida and stay home during the holidays.

    1) is   2) are   3) isn't   4) aren't

C. Directions: In the following passage, select the correct form of the verbs. (Consider formal, written English)

    Written languages originate only in large, complex societies that …31…….. a great deal of information that ………32……. to be sorted. Even the Incas, who……33……. very advanced, never developed a written system. It is, therefore, astounding that

wooden tablets which ......34...... writing ....35...... discovered on tiny Easter Island in the 19th century. The writing was in Hieroglyphics, and amongst the characters........36....... figures of animals which ....37....... unknown in the island at that time.

For over a decade, scientists worldwide ......38........... tried o decipher the writing. However, any hope of doing this...........39........ probably lost in 1862, when one fifth of the island's population ......40............ carried off to slavery in Peru. Amongst the unlucky............41........... King Mauratu, along with his family, who.........42....... the last of the learnt on Easter Island. Political pressure, which ....43....... supplied by France and England, eventually ....44.......... the release of the captives. Unfortunately, by that time, over 80 percent ....45...... dead. And the survivors who.........46........ returned to the island brought smallpox, which ravaged the natives. Thus, by 1875, the population, which in 1860.........47....... About 5,000 had shrunk to 600. Of course, these victims did not die alone. With them died the knowledge of the oral history and written language that........48...... not been and perhaps will never be recovered. Today only a few of the tablets........49...... ; the rest .....50..... vanished.

31. 1) have   2) has   3) was   4) were

32. 1) need   2) needs   3) has to   4) have to

33. 1) was   2) were   3) are   4) is

34. 1) will contain 2) contain   3) containing   4) contains

35. 1) was   2) were   3) are   4) is

36. 1) was   2) were   3) are   4) is

37. 1) was   2) were   3) are   4) is

38. 1) have   2) has   3) will   4) would

39. 1) was   2) were   3) are   4) is

40. 1) was   2) were   3) are   4) is

41. 1) was   2) were   3) are   4) is

42. 1) was   2) were   3) are   4) is

43. 1) was   2) were   3) are   4) is

44. 1) brings   2) bring   3) will bring   4) brought

45. 1) was   2) were   3) are   4) is

46. 1) was   2) were   3) are   4) is

47. 1) was   2) were   3) are   4) is

48. 1) have   2) has   3) will   4) would

49. 1) will remain 2) remained   3) remains   4) remain

50. 1) have   2) has   3) will   4) would

D. Directions*:* Fill in the blanks with the most appropriate form of the verb "***to be".***

51. Restricted ownership and enforced registration ………… two major approaches to handgun control.

52. There ….……… no doubt in any one's mind about the guilt of the defendants.

53. A grey and white car …………….parked outside the building.

54. To a do-it-yourself handyman, a pair of pliers …………very useful.

55. Despite the development of a vaccine, measles ………..still a serious danger to adult victims.

56. Statistics as well as a course in research methodology …………required of all doctoral candidates.

57. The criteria for promotion …………clearly stated in our company last year.

58. Supernatural phenomena …..……… of great interest to many people.

59. The Portuguese ……………..fortunate to have such a beautiful coastline.

60. The alumni of the university ……………invited to the graduation ceremony.

Authors' Contribution:

The authors have written the article in a complete collaboration.

Conflict of Interests:

This research does not conflict with personal and/or organizational interests.

References

Angoff, William H. *Scales, Norms, and Equivalent Scores*. ETS, 1971.

Bachman, Lyle F. *Fundamental considerations in language testing*. Oxford University Press, 1990.

Bachman, Lyle F., and Adrian S. Palmer. *Language testing in practice: Designing and developing useful language tests*. Oxford University Press, 1996.

Benítez-Burraco, Antonio. "The cognitive science of language diversity: achievements and challenges." *Cognitive Processing* (2025): 1-18.

Birdsong, David, and Michelle Molis. "On the evidence for maturational constraints in second-language acquisition." *Journal of Memory and Language,* 44,2, 2001, pp. 235-249.

Brown, James Dean. "Multiple views of L1 writing score reliability." 2007.

Brown, James Dean, and Kathleen M. Bailey. "Language testing courses: What are they in 2007?." *Language Testing,* 25,3, 2008, pp. 349-383.

Brown, James Dean. *Testing in language programs: A comprehensive guide to English language assessment,* 2011.

DeKeyser, Robert M. "The Robustness of Critical Period Effects in Second Language Acquisition." *Studies in Second Language Acquisition*, vol. 22, no. 4, 2000, pp. 499–533. Cambridge University Press, https://doi.org/10.1017/S0272263 100004022..

DeKeyser, Robert. "Age effects in second language learning." *The Routledge handbook of second language acquisition*. Routledge, 2013, pp. 442-460.

DeMars, Christine E. "Classical test theory and item response theory." *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development* (2018): 49-73.

Dörnyei, Zoltán, and Ema Ushioda. *Teaching and researching motivation*. Routledge, 2021.

Ekstrand, Lars Henric. "Age and Length of Residence as Variables Related to the Adjustment of Migrant Children, with Special Reference to Second Language Learning." 1975.

Ellis, Rod. *SLA research and language teaching*. Oxford University Press, 198 Madison Avenue, New York, NY 10016-4314, 2008.

Fathman, Ann. "The relationship between age and second language productive ability." *Language learning,* 25,2, 1982, pp. 245-253.

Fulcher, Glen, and Davidson, Fred. "Language testing and assessment: An Advanced Resource Book." Oxford: Routledge, 2007.

Geisinger, Kurt F., Bruce A. Bracken, Janet F. Carlson, Jo-Ida C. Hansen, Nathan R. Kuncel, Steven P. Reise, and Michael C. Rodriguez. *APA handbook of testing and assessment in psychology, Vol. 2: Testing and assessment in clinical and counseling psychology*. American Psychological Association, 2013.

Hakuta, Kenji, Ellen Bialystok, and Edward Wiley. "Critical evidence: A test of the critical-period hypothesis for second-language acquisition." *Psychological science,* 14,1, 2003, pp. 31-38.

Hartshorne, Joshua K., Joshua B. Tenenbaum, and Steven Pinker. "A critical period for second language acquisition: Evidence from 2/3 million English speakers." *Cognition* 177 (2018): 263-277.

Holland, Paul W., and Machteld, Hoskens. "Classical test theory as a first-order item response theory: Application to true-score prediction from a possibly nonparallel test." *Psychometrika,* 68, 2003, pp. 123-149.

Hambleton, Ronald K., and Russell W. Jones. "An NCME instructional module on." *Educational measurement: issues and practice* 12.3 (1993): 38-47.

Hyltenstam, Kenneth. "Maturational Constraints in SLA." *The Handbook of Second Language Acquisition/Blackwell,* 2003.

Johnson, Jacqueline S., and Elissa L. Newport. "Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language." *Cognitive psychology,* 21,1, 1989, pp. 60-99.

Kumaravadivelu, Balasubramanian. *Understanding language teaching: From method to postmethod*. Routledge, 2006.

Krashen, Stephen D. *Issues in second language research*. London: Newbury House, 1979.

Krashen, Stephen D., Long, Michael A. and Scarcella, Robin C. "Age, Rate and Eventual Attainment in Second Language Acquisition." *TESOL Quarterly*, vol. 13, no. 4, (1979), pp. 573–82. *JSTOR*, https://doi.org/10.2307/3586451.

Krashen, Stephen D. "Second language acquisition." *Second Language Learning,* 3,7, 1981, pp. 19-39.

Larsen-Freeman, Diane, and Michael H. Long. *An introduction to second language acquisition research.* Routledge, 2014.

Larsen-Freeman, Diane. "Teaching and testing grammar." *The handbook of language teaching,* 2009, pp. 518-542.

Lenneberg, Eric H. "Biological foundations of language." 1967.

Lightbown, Patsy, Nina Spada, and Robert Wallace. "Some effects of instruction on child and adolescent ESL learners." *Research in second language acquisition,* 1980, 162-172.

Spada, Nina, and Patsy M. Lightbown. *How languages are learned.* Oxford University Press, 2006.

Lin, Siyan, Xia Wang, and Yumeng Zhang. "Evaluating the Critical Period Influence on the Development of Semantic Radicals Awareness and Lexical Tones Awareness in Chinese Character Recognition." *Lecture Notes in Education Psychology and Public Media,* 2, 2023, pp. 662-668.

Liu, Yanyan. "The influence of critical period hypothesis on the second language acquisition." *Lecture Notes in Education Psychology and Public Media, 11,* 2023, pp. 26-31. https://doi.org/10.54254/2753-7048/11/20230706

Mayberry, Rachel I., and Elizabeth Lock. "Age constraints on first versus second language acquisition: Evidence for linguistic plasticity and epigenesis." *Brain and language,* 87,3, 2003, pp. 369-384.

Muñoz, Carmen, and Teresa Cadierno. "How do differences in exposure affect English language learning? A comparison of teenagers in two learning environments." *Studies in Second Language Learning and Teaching* 11.2 (2021): 185-212.

Oyama, Susan. "A sensitive period for the acquisition of a nonnative phonological system." *Journal of psycholinguistic research,* 5, 1976, pp. 261-283.

Oyama, Susan. "The sensitive period and comprehension of speech." *Child-adult differences in second language acquisition,* 3951, 1982.

Patkowski, Mark S. "The sensitive period for the acquisition of syntax in a second language 1." *Language learning,* 30,2, 1980, pp. 449-468.

Quirk, Randolph, and David Crystal. *A comprehensive grammar of the English language*. Pearson Education India, 2010.

Rahman, Mohammad Mosiur, et al. "Effect of age in second language acquisition: A critical review from the perspective of critical period hypothesis and ultimate attainment." *International Journal of English Linguistics,* 7,5, 2017, 1-7.

Richards, Jack C. and Schmidt, Richard. *Longman dictionary of language teaching & applied Linguistics* (3rd ed.). NY: Pearson Education, 2002.

Singleton, David. "Critical period or general age, Age and the acquisition of English as a foreign language". *Clevedon,* 2003, pp. 3-22.

Singleton, David. "Language acquisition: The age factor." *Multilingual Matters,* 2004.

Singleton, David, and Justyna Leśniewska. "The critical period hypothesis for L2 acquisition: An unfalsifiable embarrassment?." *Languages,* 6,3, 2021, pp. 149-162.

Stevens, Gillian. "Using census data to test the critical-period hypothesis for second-language acquisition." *Psychological Science,* 15,3, 2004, pp. 215-216.

Wang, Meihua. "A study on the relationship between age onset of English learning and English achievement." *Theory and Practice in Language Studies,* 5,1, 2015, pp. 164-178.