International Journal of Learning, Teaching and Educational Research
Vol. 21, No. 10, pp. 265-279, October 2022
https://doi.org/10.26803/ijlter.21.10.14
Received Jul 28, 2022; Revised Oct 4 2022; Accepted Oct 20, 2022

The Impact of Metalanguage on EFL Learners' Grammar Recognition

Abuelgasim Sabah Elsaid Mohammed
Prince Sattam bin Abdulaziz University, Hawtat Bani Tamim, Saudi Arabia

Abdulaziz B Sanosi*
Prince Sattam bin Abdulaziz University, Hawtat Bani Tamim, Saudi Arabia

Abstract. It is common knowledge that grammar instruction is a prime factor in developing EFL students' proficiency. However, the argument revolves around the best method of teaching grammar, whether implicitly or explicitly. Metalanguage (ML), which stands for the technical terms that describe language, is essential for explicit instruction. As this teaching method is widely acknowledged, much research has focused on the relation between ML and language acquisition and proficiency. The focus has always been on the role of ML in developing students' language skills and uptake when taught through such a method. However, there is a relative paucity of studies that explore the exact impact of ML-based instruction on students' grammatical recognition, especially in the Arabian EFL context. Accordingly, the current study investigated the issue through an experimental method. The participants' (n = 73, 35 in the experimental group, and 38 in the control group) scores on the pre- and post-tests were analysed using t-test and item-analysis to test the research hypothesis. The results revealed that ML positively affects students' grammatical recognition. However, this improvement was evident only in some grammatical structures. These results imply that ML is beneficial for grammar instruction; however, a mixture of teaching methods should be applied to account for all grammatical constructs. The study findings may contribute to enriching the literature on ML and provide evidence for its importance in grammar instruction.

Keywords: metalinguistic knowledge; explicit instruction; focus on form; grammar; language knowledge

1. Introduction

Grammar teaching methods are among the widely discussed topics in English language teaching (ELT) literature. In this regard, a fundamental issue to investigate is whether to teach grammar explicitly or implicitly. Both methods have their advocates and opponents. Explicit instruction based on rule

* Corresponding author: *Abdulaziz B Sanosi, a.assanosi@psau.edu.sa*

presentation and error correction is believed to promote students' accuracy (Alderson et al. 1997) though it may impede spontaneous interaction (Krashen, 1982). On the other hand, implicit instruction is suggested with the advent of the Communicative Language Teaching Method (Savignon, 1987) to foster the meaningful acquisition of a second language (L2). However, it was revealed later that teaching grammar implicitly might lead to inaccurate usage of grammatical structure as students "cannot easily attend to both meaning and form at the same time" (Basturkmen et al., 2002, p. 1). Accordingly, explicit instruction remains the norm for most EFL teachers as the prime objective of most of them is accuracy.

Explicit instruction requires students' knowledge of a collection of technical terms that describe linguistic units, structures and relations. These terms are known as metalanguage (ML). Knowledge of ML is believed to be of essential importance to EFL learners. It helps students to comprehend grammar rules better (Alderson et al., 1997) and it improves their language proficiency (Hu, 2010). To this end, metalinguistic knowledge and its relation to EFL learners have been the focus of considerable research over the past decades. Previous research focused on the effect of ML on different language skills and aspects (Daffern, 2016; Dong, et al., 2020), their application of what they study in their actual performance (Basturkmen, et al., 2002), and the relationship between linguistics and metalinguistic units (Alipour, 2014). Nevertheless, a relatively neglected area in this field is identifying how ML-based instruction can affect grammar recognition.

Recent years have witnessed a revitalized consideration of the topic. Accuracy levels of EFL (English foreign language) and ESL (English second language) learners have deteriorated. This represents a problem as one of the essential aims of EFL programmes is to foster correct language usage. Therefore, "more recently, the consensus seems that some form of grammar instruction is useful" (Nunan, 2003, p. 157). Notwithstanding, sufficient research that equals the urgency of these findings could not be found. In efforts to deal with this renewed importance, this study attempted to measure the impact of ML-based instruction on developing students' grammar recognition. It was aimed at validating the claim that ML can foster the grammatical performance of EFL learners. The researchers investigated this area by testing the following null hypothesis.

H_0 *There is no impact of ML on EFL learners' grammar recognition*.

2. Literature Review

Metalanguage

Metalanguage is the language teachers use to speak about the language when they teach (Ellis, 2016). As Ellis (2016) points out, three uses of ML are distinguished. The first one is utilizing ML to talk about primary grammatical phrases such as *objects, nouns, verbs, or sentences*. Second, ML refers to technical linguistic terms including *phonotactics, adjunct, or suffixes*. Finally, ML is employed for non-technical terms such as *means, say, or correct*. These three uses of ML to describe essential grammatical items, technical and non-technical terms, are referred to by Fortune (2005) as *ML A, ML B, and ML C*, respectively. This classification is beneficial in limiting the range of ML and is also useful in categorizing its various functions (Ellis, 2016). Nevertheless, Bastrukmen et al. (2002) explain that this classification is not that simple as it does not incorporate assessing remarks such as *excellent* or *correct*.

The use of ML falls in the argument of whether to teach grammar implicitly or explicitly. (Ellis, 2005, in Roshan & Elhami, 2016b). The former involves no awareness and could be used in fast and natural communication. In comparison, the latter is confirmatory and focuses on the accurate knowledge of a foreign language's phonological, lexical, grammatical, pragmatic, and socio-critical characteristics, along with employing ML to mark this knowledge.

A plethora of research suggests that ML can be useful for language learning. For instance, Ellis (2016) proposes that it has prolific uses in communicative language teaching (CLT). Roshan and Elhami (2016a) point out that using ML assists students in communicating through the language, explaining and analysing it. Consequently, this supports learners to possess correctness, self-correction, and competence. In CLT, ML raises learners' awareness of the language, resulting in accurate and fluent use of the target language (Ellis, 2016). Moreover, research has shown that ML helps enhancing students' attention to English grammatical items (Abdollahzadeh, 2016; Roshan & Elhami, 2016a; Rizqan, 2020); improving learners' general language proficiency (Berry, 1997 in Berry 2010; Berry 2009a; Hanson, 2013; and Randen, 2022); and raising students' awareness of the language (Hu, 2010).

Furthermore, Hu (2010) reports that ML is beneficial in several ways. First, it improves learners' knowledge of forms and functions. These two aspects are crucial in grammar learning. Second, it draws on the contrastive analysis between the foreign language and the mother tongue. This contrastive analysis is influential in predicting students' mistakes and raising their awareness to avoid them. Third, it facilitates grammar instruction originating from students' knowledge of forms and functions, increasing students' consciousness of their learning. Hence, their metacognitive abilities will be activated and employed. Fourth, Hu (2010) suggests that another benefit of ML is concerned with

> *the explanatory precision with which linguistic generalization can be made and the efficient delimitation of the contexts to which the generalization applies. ML that is appropriately used can pre-empt both under- and over-generalization of the rules in question.* (p. 66).

According to Roshan and Elhami (2016a), this may lead to false generalizations. Another advantage of ML is its importance in teaching/learning grammar (Roshan & Elhami, 2016b; Rizqan & Rohmah, 2020) since it raises students' attention to the language while enabling them to investigate and recognise it. Consequently, ML activates students' attention to and recognition of grammatical items.

Attention, Noticing, and Recognition

Noticing theory hypothesizes that 'input' becomes 'intake' if it is noticed and deliberately archived (Schmidt, 2010). Schmidt (2010) states that language acquisition is initiated by what students recognize and are conscious of in the language input. What Schmidt refers to is that recognition plays an integral role in understanding language input of all kinds. Schmidt (2010) maintains that "some level of attention to form is required" (p.723).

Awareness can be classified into three ranks (Schmidt, 2010). The first one is conscious, which is a cognitive operation involving formulating inner images of exterior incidents. The second level of consciousness is recognition or noticing. Recognition refers to the variation between implied and recognized data (Bowes, 1984, as cited in, Schmidt, 1990). Understanding is the last level of awareness. It entails noticing and thinking about something to dissect and contrast it with other things noticed in various opportunities.

Conscious raising aims to assist language students in attending to forms in the input while simultaneously providing the required input for acquiring the language (Benati, 2021and East, 2021). Conscious raising could be achieved using ML, which facilitates learning grammatical items (Roshan & Elhami, 2016a).

Previous Studies

Targeting investigating the ML categories production in grammar and exam adopting content analysis, Rizaqan & Rohmah (2020) found that their students produced both technical and non-technical ML, with the former occurring more frequently than the latter. The study concluded that both ML types affected students' grammar recognition. Bakhshandeh & Jafari (2018) disclosed that overt teaching of the present and past simple passive voice forms improved the participants' direct knowledge of these forms as the emprical group's students performed better than those in the control group in the untimed grammatical judgment post-test.

Djahimo (2018) reported two results. The first was that the participants expressed their satisfaction with learning English grammar through consciousness-raising tasks. It enabled them to understand the conditional sentences, simple present and past tenses, which were targeted by the study. The second result was that the students performed well in the post-test compared to their performance in the pre-test. Roshan & Elhami (2016 a & b) investigated the impact of ML on Iranian EFL students' grammar noticing and learning, respectively. The studies found that ML enhanced the students' grammar recognition and learning by raising the participants' consciousness of grammatical items. Bouziane & Harrizi (2014) examined Moroccan EFL learners' grammatical competence after explicit and implicit grammar teaching. The study adopted Swan and Baker's *Grammar Scan* expert tests. They found that the participants' scores were high in some grammatical items such as the infinitive, relatives, sentence structures, passives, determiners, *-ing* forms, and nouns, among other items. They made use of explicit and implicit grammar teaching. Mallia (2014) concluded that most of his participants preferred deductive grammar learning. Mallia (2014) revealed that those who preferred explicit grammar learning did better in forming, applying, and recognizing the past perfect tense.

These studies suffer from some limitations. Rizqan and Rohmah (2020) employed a limited number of participants. Roshan and Elhami's (2016a) study adopted a noticing test that included a few items compared to those adduced in Basturkman et al. (2002, pp. 7-8) and Berry (2010, p.141). Similarly, Roshan and Elhami's (2016b) research adopted a questionnaire that might not suffice such a study since it incorporates a number of items and grammatical points. Therefore, it limits the generalizability of its findings. Bouziane and Harrizi (2014) collected their data via a test a year after students had finished their grammar course. This delay might have affected the students' level of grammar knowledge. Thus, the results revealed by the study might not be valid.

3. Methods

The study reported here adopted a quasi-experimental design as it is the most suitable method of exploring the causal links between variables in applied linguistics (Dornyei, 2007). The selected method was used because grammar recognition can be influenced by many variables. Focusing on the effect of ML-based instruction can better be attained by an experimental approach that involves applying the treatment. Accordingly, the researchers administered pre- and post-tests to collect the research data.

3.1 Participants

Seventy-three female students participated in this study. They studied English Language and Literature at Prince Sattam Bin Abdulaziz University, Saudi Arabia. They had been automatically assigned to two different sections by the university registration system based on the priority of registration. The researchers took advantage of this categorization and designated one section (n = 38) as a control group and adopted the other group (n = 35) as experimental. Thus, the sampling method followed can be considered a clustering method. Such sampling was considered convenient and suitable since it maintains homogeneity of the groups and variability of its members according to normal distribution and individual differences. Two intact classes were chosen: one for the treatment and the other as a baseline for controlling the experiment.

The students' English Language levels were not measured for the purposes of the study, However, based on them having completed nine years of studying general English at public schools and three semesters' majoring in English, it was supposed that their levels would range from A1 to B1 according to the Common European Framework of Reference for Languages (CEFR). They studied introductory courses in English language, literature, and translation. The participants were informed about the study and notified that participation was voluntary and independent of course assessment.

3.2 Instruments

To collect data for this study, we adopted pre- and post- Grammatical Judgment Tests (GJT). GJT is famous as a "standard method of determining whether a construction is well-formed" (Rimmer, 2006, p. 246). GJT has been in use for more than 50 years to evaluate students' understanding of grammar and their language level (Tan & Noor Izzati, 2015). Considering the aim of this study, the GJT was thought to be relevant, as it enabled us to measure the participants' grammar correctness before and after the treatment.

Each of the two tests contained 25 questions divided into two sections. The first section consisted of 12 questions, eight of which were wrong and four were correct sentences. The students were required to evaluate (judge) the correctness of the sentences. However, the second section comprised 13 multiple-choice questions (MCQs) from which the participants were requested to choose the correct answer.

The questions were modified from Swan and Baker's (2008) Grammar Scan Test. The tests encapsulated specific grammar items that covered adjectives, adverbs, articles, nouns, pronouns, and verb tenses. These items were displayed in all types of sentences: positive, negative, and interrogative.

The rationale for selecting these grammatical points was based on the university's previous courses students had studied. Before choosing the items, the researchers reviewed the course specifications of *Grammar I and Grammar II* courses which are taught at levels two and three, respectively. Additionally, the same grammatical points were included in the course the students were studying. The course, *An Introduction to Translation*, deals with the problems that face English-Arabic translators due to differences between the grammatical systems of the two languages.

3.3 Test Reliability

Two university professors specializing in Applied Linguistics and TESOL (Teaching English to Speakers of Other Languages) were asked to review the test and evaluate its correctness, relevance and suitability for the study. They provided feedback on specific points employed to edit the final draft of the tests. The reliability of the test results was verified by the Blackboard item-analysis tool, which computes the discrimination power of the questions, their difficulty, the standard deviation of the scores and the standard error of measurement.

According to the Blackboard manual, discrimination values indicate the ability of a question to differentiate between students with a high and low level of understanding. A value can fall into one of three categories: *Good* (a score more than 0.3), *Fair* (a score amid 0.1 and 0.3), and *Poor* (a score lower than 0.1). On the other hand, difficulty indicates the proportion of participants who accurately responded to the test items. The higher the percentage, the easier the question. A question is considered *Easy* if its difficulty value is greater than 80 %, *Medium* if the value is between 30% and 80 %, and *Hard* if the value is less than 30%. Item analysis also analyses how the questions are grouped or deviate from the average score. Smaller Standard Deviation values suggest that the scores are tightly grouped around the average score, suggesting that the students' performance is homogenous. On the other hand, standard error values indicate the extent of accidental variability. The smaller the standard error of measurement, the more accurate the measurement provided by the test question.

3.4 Procedures

In the initial stage of the research, participants were notified of the objectives and importance of the study. They were also informed that participation in the study was voluntary and unrelated to the formal assessment of current courses and hence no consequences were expected if they were not willing to participate. As the two tests were available online through Blackboard, they were informed that logging into the system and starting the tests indicated their understanding and acceptance of the terms. All the students agreed to the terms and participated in the study.

The pre-test was administered in the second week of the second semester in January, 2022. The participant's scores and analyses were automatically acquired through the university Learning Management System (LMS) Blackboard and stored for the second analysis stage. Subsequently, the second researcher taught the selected grammatical points to the two groups for twelve weeks using two different methods. The researcher used an inductive approach for the control group, presenting the targeted grammatical points in context. The students were then guided to elicit the forms and asked to generate comparable forms in different contexts. Students were then asked to translate the forms into Arabic. The problems of translation were discussed thoroughly, and suggested solutions were presented.

For the experimental group, explicit grammar instruction, based on English and Arabic ML was used. Students were coached on parts of speech and tenses in both languages. This was done by comparing the related rules and functions in the two languages. Students were asked to practise the related forms and generate comparable forms in both languages. Throughout these stages, grammatical technical terms were used. After twelve weeks of teaching with both methods, the post-test was administered through Blackboard. The scores of the students on both tests were downloaded to an Excel spreadsheet. Scores of students who completed one test were excluded, and the remaining data were adopted for analysis.

3.5 Data Analysis

To test the central research hypothesis, researchers conducted an independent sample t-test. The T-test is a proper statistical method when the aim is to compare various groups of people (Dornyei, 2007). Accordingly, the researchers used SPSS to compute the mean scores of the two tests for both groups and the P-value to determine the significance of the difference between the scores. Furthermore, the experimental group scores on both tests were analysed to check the impact of the treatment on the specific grammatical points studied, that is, adjectives, adverbs, articles, nouns, pronouns, or verb tenses. To achieve this, the differences between the mean scores of each grammatical point in the pre-test and the post-test were computed.

4. Results and Discussion

4.1 Participants Scores in the Pre-test

The results of the first data analysis stage pertaining to the pre-test's item analysis are displayed in Table 1 below.

Table 1. Pre-test item analysis (Sorted by difficulty)

Q	Question Type	Discrimination	Difficulty	Average	Std Dev.	Std Error
13	Either/Or	0.3	0.85	0.86	0.36	0.06
5	Either/Or	0.4	0.83	0.84	0.38	0.06
23	Multiple Choice	0.3	0.82	0.83	0.39	0.07
24	Multiple Choice	0.4	0.80	0.80	0.41	0.07
7	Either/Or	0.2	0.79	0.80	0.42	0.07
15	Multiple Choice	0.4	0.78	0.79	0.42	0.07
20	Multiple Choice	0.3	0.77	0.78	0.43	0.07
25	Multiple Choice	0.3	0.74	0.75	0.45	0.07
2	Either/Or	0.3	0.70	0.70	0.47	0.08

21	Multiple Choice	0.4	0.70	0.71	0.46	0.07
4	Either/Or	0.4	0.67	0.68	0.47	0.07
10	Either/Or	0.3	0.65	0.66	0.49	0.08
6	Either/Or	0.4	0.64	0.65	0.49	0.08
1	Either/Or	0.4	0.63	0.64	0.50	0.08
18	Multiple Choice	0.3	0.62	0.63	0.50	0.08
22	Multiple Choice	0.3	0.61	0.62	0.50	0.08
17	Multiple Choice	0.5	0.56	0.57	0.50	0.08
19	Multiple Choice	0.4	0.54	0.55	0.51	0.08
16	Multiple Choice	0.4	0.51	0.51	0.51	0.09
12	Either/Or	0.2	0.49	0.50	0.50	0.08
14	Multiple Choice	0.3	0.48	0.49	0.51	0.09
11	Either/Or	0.4	0.46	0.47	0.51	0.08
9	Either/Or	0.3	0.39	0.40	0.50	0.08
3	Either/Or	0.4	0.28	0.29	0.46	0.07
8	Either/Or	0.4	0.27	0.27	0.45	0.07

In general, the results show that the question type did not affect the participants' performance. It is also noted that most questions have a significant discrimination power. Twenty-three test questions had discrimination values of 0.3 or above, i.e., they fall into the *Good* category, while two questions were *Fair* and no *Poor* question found. On the other hand, the test difficulty was normal. Overall, there were 4 *easy* questions (with values < 0.80), two *hard* ones (with values > 0.30) and 19 *medium* questions. The small standard deviation values of the questions suggested that the students were homogenous in their answers as their scores were clustered closely around specific values. Moreover, it is noted that standard error values also were small, suggesting an accurate measurement. These results support the reliability of the test scores.

4.2 Participants' Scores in the Post-test

As for the post-test, Table 2 presents the item analysis results of the test questions.

Table 2. Post-test item analysis (Sorted by difficulty)

Q	Question Type	Discrimination	Difficulty	Average	Std Dev	Std Error
25	Multiple Choice	0.2	0.85	0.86	0.37	0.06
23	Multiple Choice	0.3	0.82	0.83	0.40	0.07
22	Multiple Choice	0.3	0.78	0.79	0.43	0.07
13	Either/or	0.2	0.74	0.75	0.45	0.07
24	Multiple Choice	0.5	0.72	0.73	0.45	0.08
19	Multiple Choice	0.4	0.71	0.71	0.46	0.08
9	Either/or	0.4	0.69	0.70	0.47	0.08
12	Either/or	0.2	0.69	0.70	0.47	0.08
21	Multiple Choice	0.3	0.69	0.69	0.48	0.08
20	Multiple Choice	0.2	0.66	0.67	0.48	0.08
16	Multiple Choice	0.4	0.64	0.64	0.49	0.08

11	Either/or	0.5	0.63	0.64	0.50	0.08
3	Either/or	0.1	0.57	0.57	0.50	0.08
8	Either/or	0.1	0.55	0.55	0.51	0.08
17	Multiple Choice	0.4	0.55	0.56	0.50	0.08
10	Either/or	0.4	0.53	0.54	0.50	0.08
15	Multiple Choice	0.5	0.53	0.53	0.51	0.08
5	Either/or	0.2	0.48	0.49	0.51	0.08
7	Either/or	0.2	0.48	0.48	0.51	0.08
6	Either/or	0.2	0.47	0.47	0.51	0.08
14	Multiple Choice	0.4	0.47	0.48	0.51	0.08
18	Multiple Choice	0.2	0.47	0.47	0.48	0.08
4	Either/or	0.4	0.46	0.46	0.51	0.08
2	Either/or	0.3	0.42	0.42	0.51	0.08
1	Either/or	0.2	0.32	0.33	0.48	0.08

Regarding discrimination power, the post-test had fourteen *Good* questions and eleven *Fair* ones. The test included no *Poor* questions. Most of the questions, that is, twenty-three, were of *medium* difficulty, while two questions were in the *Easy* category. The small standard deviation values indicate a consistent performance by the participants, and the low standard error values mean that the measurement [What does this mean?] is unlikely to be due to chance, as all the values were less than 0.1, which means a slight probability of chance.

4.3 The Impact of ML on EFL learners' grammar recognition

The central research hypothesis was tested after checking the final data for reliability. An independent T-test was performed to compare the scores of students who received ML-based instruction (Experimental group) to the score of the control group which received regular instruction. The results are displayed in Table 3 below.

Table 3. T-tests results

		Mean	**N**	**Std. Deviation**	**t**	**df**	**Sig (2-tailed)**
Pair 1	CTRL_pre	14.5	38	4.70	-.718	37	.477
	CTRL_post	15.1	38	4.10			
Pair 2	EXP_PRE [Comparing the performance of both groups in the pre-test is important?]	13.86	35	3.25	-9.586	34	.000
	EXP_POST	18.11	35	4.02			

The findings show that no significant difference occurred between the performance of the two groups in the pre-test. The difference between the mean scores of the group is only 0.6; however, a considerable difference is observed in their scores in the post-test as displayed in Figure 1 below.

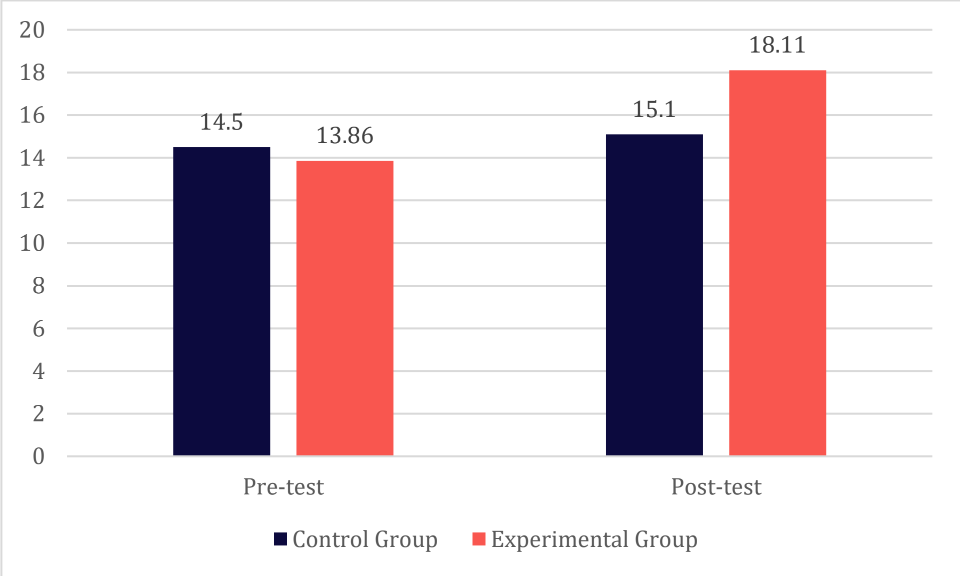

Figure 1: Overall participant scores in the two tests

Furthermore, the results indicate that there was a significant difference between the performances of the experimental group in the post-test (M=18.11, SD = 4.02) and the pre-test (M=13.86, SD =3.25); t(34) = -9.586, p = .000). This result is supported by the insignificant difference in the performance of the control group in the post-test (M=15.1, SD = 4.10) and the pre-test (M=14.5, SD= 4.70); t(37) = -.718, p = .477. Accordingly, the null hypothesis of the research is rejected, and the alternative hypothesis, that is., there is a significant relation between ML-based instruction and grammatical accuracy, is supported.

4.4 The Impact of ML on the Selected Grammatical Points

Further analysis was conducted to check the impact of ML on each grammatical point. The result of this analysis is presented in Table 4.

Table 4. Overall participants' performance on the two tests in terms of grammatical points

Group →	Control			Experimental		
↓Gram. Point	*Pre. Av.*	*Post_Av.*	*Diff.*	*Pre.Av.*	*Post_Av.*	*Diff.*
Adjectives	0.54	0.57	0.03	0.54	0.55	0.01
Adverbs	0.59	0.62	0.03	0.59	0.63	0.04
Articles	0.66	-0.65	-0.01	0.66	0.64	-0.02
Nouns	0.66	0.61	-0.05	0.67	0.48	-0.18
Pronouns	0.70	0.64	-0.06	0.71	0.65	-0.06
Verb Tense	0.50	0.56	0.06	0.51	0.60	0.09

The findings show a considerable discrepancy in the gains of students in different grammatical points. They hence suggest varied effects of ML instruction on different grammatical aspects. Figure 2 below illustrates the gains achieved by each group regarding each grammatical point.

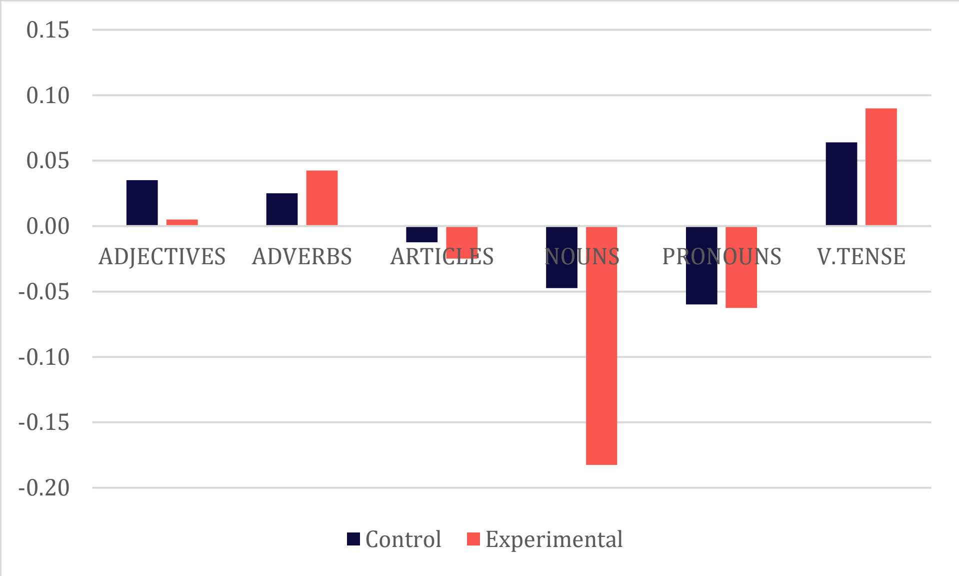

Figure 2: Gains achieved by the two groups in each grammatical point

It is evident that while the treatment was effective in improving the participant's performance in adverbs and verb tense, it had no effect on adjectives, articles, nouns, and pronouns.

In general, the result is in line with the findings of most of the previous studies reported in literature. The positive impact of using ML-based grammar instruction was supported by Bakhshandeh and Jafari (2018), Roshan and Elhami (2016a & b), Djahimo (2018), and Sanosi (2022), who noticed an improvement in the knowledge and use of grammatical forms after receiving explicit instruction based on ML. This improvement is represented in several forms and labelled by different terms such as noticing (Roshan & Elhami, 2016a & b), Knowledge (Bakhshandeh & Jafari, 2018), consciousness (Djahimo, 2018), and accuracy (Sanosi, 2022). All of these, nevertheless, revealed that ML instruction affects students' performance in grammar in one way or another.

An interesting point to investigate, however, is that the development in the participants' performance is not observed at all levels of grammatical points studied. A probable justification for this finding is that metalanguage, although ML knowledge is helpful in improving grammatical recognition, it is not the exclusive factor that determines improving it. Several extrinsic and intrinsic factors can determine students' recognition of grammar, including teaching methods, syllabi, and L1 (first language) transfer. As evident in Table 4 above, the improvement is detected in the participants' scores regarding only two grammatical points, which were *adverbs* and *verb tense*, respectively. In contrast, scores in adjectives, articles, nouns and pronouns were not affected by the treatment, suggesting no impact of ML-based grammar instruction. Implications of these results are that using ML-based instruction is likely to be impactful on specific grammatical points, which means that other grammatical structures require different teaching techniques.

This result supports the findings of Bouziane and Harrizi, (2014), who found that explicit grammar instruction effectively improved some grammatical structures but not all of them. Also, tenses, especially past and present perfect, were some of the structures acquired more properly after students were taught explicitly. It is also comparable to the findings of Bakhshandeh and Jafari (2018), which suggested that ML classes were influential in developing students' performance on past and present passive voice.

While it seems intuitive that ML-based instruction is likely to foster students' grammar accuracy and knowledge, the partial impact is worth more investigation and further research. A sound justification of this phenomenon is probably the differences between students' L1 and English and their negative transfer. For example, it is widely acknowledged that Arab EFL students face difficulty regarding articles, pronouns, and nouns, among other points. Notwithstanding, even the improved grammatical points were also acknowledged as challenging for Arab EFL Learners. For example, verb tenses are proved to be complicated by many researchers e.g. (Yaseen, et al., 2018); nevertheless, they are found to be positively affected by ML-based instruction. Accordingly, it can be suggested that a mixture of explicit instruction based on ML and implicit instruction that adopts various techniques is the proper way of teaching grammar. These techniques can include contextualization by supplying socially-relevant examples (Mallia, 2018) and using the mother tongue ML (Rakab, 2021).

It is plausible that two limitations might have influenced the results obtained in this study. First, the size of the sample is small. Although all the necessary procedures were taken to guarantee proper sampling, it is thought that a larger sample size would generate more valid results. Another limitation is that the instrument did not take account of the ML knowledge per se. in other words, the test did not include questions on the knowledge and use of technical terms such as verbs, nouns, pronouns, and so forth. Having data about such usage might support or refute the claim that the improvement in the student's performance is because of their ML knowledge and not because of other factors such as the practice effect.

The present findings might have important implications for grammar instruction. As it is proved that ML-based instruction positively affects grammar accuracy in some structures, EFL (English first language) teachers are presupposed to apply it parallelly with other techniques that can account for the other undeveloped grammatical points. Teachers are asked to apply contexts, socio-communicative activities and inductive methods to teach grammar in this regard. To support this approach, it is expected that course developers will design grammar lessons in a way that can account for both form and functions with a probable room for students to practise through dialogues, role-play, games and other comparable activities.

5. Conclusion

Using ML for teaching grammar explicitly was a topic of debate for a considerable time. However, no decisive findings were found regarding the exact effect of this approach on developing students' grammatical accuracy. Therefore, the researchers aimed to revisit the issue and investigate the impact of ML-based instruction on developing students' grammatical accuracy. An experimental design of control and experimental group was adopted. Subsequently, the participants' scores in the pre-test and post-test were analysed to measure the impact of the treatment on the experimental group's performance. The findings support the idea that ML-based instruction fosters students' knowledge and use of grammar and hence improves students' grammatical accuracy. Notwithstanding, this development is proved to be only in specific grammatical structures such as tenses, adverbs, and adjectives, while other grammatical points remain unaffected. It can thus be suggested that EFL teachers should apply a blend of explicit and implicit instruction methods which can account for both grammatical rules and functions.

This study was limited by the small sample size and the instrument's structure, which contains no questions on ML knowledge per se. Accordingly, it is recommended that future research should employ a larger sample and administer tests that assess students' comprehension of the technical metalinguistic terms. Further analysis can then be conducted to measure the correlation between students' knowledge of ML and their actual grammatical accuracy.

6. Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

7. Funding Statement

This project was supported by the deanship of Scientific Research at Prince Sattam Bin Abdulaziz University under the research project number 2021/02/17916.

8. Acknowledgments

The authors would like to thank the Deanship of Scientific Research at Prince Sattam Bin Abdulaziz University for supporting this research project. Thank you too to the third-level students, the Department of English Language and Literature at the College of Science and Humanities, Hawtat Bani Tamim for participating in the study, and Dr Ahmad Saeed Rahma for his assistance in the statistical analysis of the research data.

9. References

Alderson, J., Clapham, C. & Steel, D. (1997). Metalinguistic knowledge, language aptitude and language proficiency. *Language Teaching Research*, 1(2), 93-121. https://doi.org/10.1177/136216889700100202

Abdollahzadeh, S. (2016). The effect of metalinguistic corrective feedback on EFL learners' grammatical accuracy. *Journal of Language Teaching and Research*, *7*(1), 185–191. https://doi.org/10.17507/jltr.0701.21

Alipour, S. (2014). Metalinguistic and linguistic knowledge in foreign language learners. *Theory and Practice in Language Studies*, *4*(12), 2640–2645. https://doi.org/10.4304/tpls.4.12.2640-2645

Bakhshandeh, S. & Jafari, K. (2018). The effects of input enhancement and explicit instruction on developing Iranian lower-intermediate EFL learners' explicit knowledge of passive voice. *Asian-Pacific Journal of Second and Foreign Language Education*, 3(18), 1 - 18. https://doi.org/10.1186/s40862-018-0060-

Bastrukmen, H., Loewen , S. & Ellis, R. (2002). Metalanguage in focus on form in the communicative classroom. *Language Awareness,* *11*(1), 1-13. https://doi.org/10.1080/09658410208667042

Benati, A. (2021). Focus on form in second language instruction. In H.M. Coombe, *Research questions in language education and applied linguistics* (pp. 63 - 67). Springer. https://doi.org/10.1007/978-3-030-79143-8_12

Bouziane, A. & Harrizi, M. (2014). Explicit grammar teaching pays off: the Case of Moroccan EFL university students. *Arab World English Journal*, 5(4), 64 - 83. https://awej.org/images/AllIssues/Volume5/Volume5number4Decmber/5.pdf

Berry, R. (2010). *Terminology in Language Teaching: Nature and Use* (Vol. 93). Peter Lang. https://doi.org/10.1075/term.18.2.06nie

Daffern , T. (2016). What happens when a teacher uses metalanguage to teach spelling. *The reading teacher, 70*(4), 423-434. https://doi.org/10.1002/trtr.1528

Djahimo, S. (2018). Applying consciousness-raising tasks in teaching grammar to EFL students in Indonesia. *Asian EFL Journal*, 20(11), 41 - 51.

Dong, Y., Peng, S.-N., Sun, Y.-K., Wu, S. & Wang, W.-S. (2020). Reading comprehension and metalinguistic knowledge in Chinese readers: A meta-analysis. *Frontiers in Psychology, 10*, 1-15. https://doi.org/10.3389/fpsyg.2019.03037

Dornyei, Z. (2007). Research methods in applied linguistics. Oxford University Press.

East, M. (2021). Attention to form in the communicative classroom. In H. Mohebbi, & C. Coombe, *Research questions in language education and applied linguistics* (pp. 11 - 16). Springer. https://doi.org/10.1007/978-3-030-79143-8

Ellis, M. (2016). Metalanguage as a component of the communicative classroom. *Accents Asia*, *8*(2), 143–153. http://dx.doi.org/10.1080/09658410208667042

Fortune, A. (2005). Learners' use of metalanguage in collaborative form-focused L2 output tasks. *Language Awareness, 14*(1), 21-38. https://doi.org/10.1080/09658410508668818

Hanson, J. (2013). The relationship between metalinguistic knowledge/learning contexts and language proficiency [Master's thesis, Colorado State University]. https://mountainscholar.org/bitstream/handle/10217/79057/Hanson_colostate_0053N_11580.pdf?sequence=1

Hu, G. (2010). Revisiting the role of metalanguage in L2 teaching and learning. *English Australia Journal*, *26*(1), 61–70. http://hdl.handle.net/10497/16146

Mallia, J. (2014). Inductive and Deductive Approaches to Teaching English Grammar. *Arab World English Journal*, 5(2), 221-235.

https://awej.org/images/AllIssues/Volume5/Volume5Number2June2014/17.pdf

Krashen, S. (1982). *Principles and practice in second language acquisition.* Pergamon Press Inc.

Nunan, D. (2003). Grammar. In D. Nunan (Ed.), *Practical English Language Teaching* (pp. 153 -172). McGraw Hill.

Rakab, M. (2021). The use of L1 metalanguage in L2 classrooms: The Case for Arabic. *Journal of English Language Teaching and Applied Linguistics*, 3(7), 60-69. https://doi.org/10.32996/jeltal.2021.3.7.5

Rimmer, W. (2006). Grammaticality Judgement Tests: Trial by error. *Journal of language and linguistics*, 5(2), 246-261. https://educationdocbox.com/Language_Learning/65964415-Grammaticality-judgment-tests-trial-by-error.html

Rizqan, M.D.A. (2020). Uncovering metalanguage in grammar exam and its implication on students' cognition. *Lingua Jurnal Pendidikan Bahasa*, 16(2), 32-45. https://doi.org/10.34005/lingua.v16i2.1036

Randen, G. (2022). The sound of asking a question': Metalanguage and crosslinguistic awareness in adults learning Norwegian as an additional language. In M. Monsen & G. Steien (Ed.), *Language Learning and Forced Migration* (pp. 139-156). Multilingual Matters. https://doi.org/10.21832/9781800412262-012

Roshan, A. & A. Elhami. (2016a). Metalanguage and grammar learning of Iranian EFL learners. *European Online Journal of Natural and Social Sciences*, *5*(2), 317–322. https://doi.org/10.20472/ahc.2016.001.015

Roshan, A. & A. Elhami. (2016b). The effect of metalanguage on grammar noticing of the Iranian EFL Learners. *International Journal of English Linguistics*, *6*(3), 194. https://doi.org/10.5539/ijel.v6n3p194

Sanosi, A. B. (2022). Correlation of EFL learners' metalinguistic knowledge and grammatical accuracy. *Studies in English Language and Education*, 9(3), 908-925. https://doi.org/10.24815/siele.v9i3.24615

Savignon, S. J. (1987). Communicative language teaching. *Theory into Practice, 26*(4), 235-242. https://doi.org/10.1080/00405848709543281

Schmidt, R. (1990). The role of consciousness in language learning. *Applied Linguistics, 11*(2), 129 - 158. https://doi.org/10.1093/applin/11.2.129

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In S. C. W. M. Chan (Ed.), *Proceedings of CLaSIC 2010* (pp. 721-737). National University of Singapore, Centre for Language Studies.

Tan, B. H. & Noor Izzati, M. N. (2015). Grammaticality Judgement Test: Do item formats affect test performance? *Pertanika Journal of Social Sciences & Humanities, 23*(S), 119-130. http://psasir.upm.edu.my/id/eprint/32880/1/09%20JSSH%28S%29-0064-2015.pdf

Yaseen, A., Ismail, K. & Yasin, M. (2018). Syntactic errors in an Arab EFL postgraduate student's spoken English during a thesis supervision sesstion. *Arab World English Journal, 9*(4), 392-406. https://dx.doi.org/10.24093/awej/vol9no4.29