

Examining the Suitability of Multiple-Choice Questions for Assessing Grammatical Ability in Foreign Language Testing

Naif Saeed Alsaedi

Department of Languages and Translation, Faculty of Arts and Humanities, Taibah University, Medina, Saudi Arabia
Email: nsaadi@taibahu.edu.sa

Abstract

Despite the widespread use of multiple-choice (MC) tests in language teaching and the recognized importance of grammar instruction in language acquisition and assessment, there remains a dearth of research examining the suitability of MC tests for evaluating second language (L2) or foreign language (FL) learners' grammatical ability – their proficiency in using grammar accurately and effectively across different contexts. This study aims to address this gap by investigating whether MC tests, which assess knowledge of grammar through the selection of correct answers from options, adequately reflect learners' ability to use grammatical structures accurately, meaningfully, and appropriately. Data were collected from 68 English FL learners using two assessment tools: an MC task and a gap-filling task (GF task). Results indicate significant performance disparities between the MC and GF tasks, suggesting that MC tests may not consistently gauge learners' true grammatical proficiency. This inconsistency underscores the limitations of MC testing as a singular measure of grammatical ability, particularly in assessing learners' capacity to apply grammar meaningfully across diverse contexts. Further research is warranted to explore this issue comprehensively.

Keywords: Foreign Language Teaching/Learning, Foreign Language Testing, Grammatical Knowledge, Grammatical Ability, Multiple-Choice Format, Linguistic Proficiency Level.

1. Introduction

The field of foreign language teaching (FLT) has been developing very rapidly. During the past 70 years or so, numerous approaches concerning how second language/foreign language (L2/FL) is taught have been proposed, e.g., the grammar-translation approach, the audiolingual approach, the cognitive approach, the ask-based approach and the communicative approach (for a detailed historical review, refer to Celce-Murcia, 2001; Hinkel and Fotos, 2001b; Larsen-Freeman and Anderson, 201; Richards and Rodgers, 2001). These teaching approaches differ greatly because they draw on insights from differing prolific theories regarding foreign language learning/second language learning (FLL/SLL), such as the generative theory, the sociocultural theory, the usage-based theory, the interactionist theory, the input processing theory, the processability theory and so on (for up-to-date reviews of the major contemporary approaches to SLL, refer to Mayo, Mangado and Adrián, 2013; VanPatten and Williams, 2015). For example, the aforementioned approaches hold different conceptions of grammar teaching: ‘some see no need to teach it; some see value in instruction that raises awareness. Some see grammar teaching as more of a cognitive process in guiding learners to figure out the rules inductively or in applying them deductively, others as more dependent upon social interaction’ (Larsen-Freeman, 2020, 117).

In recent years, however, these approaches have undergone many changes, and in keeping with current developments, the role of grammar instruction and how it should be taught in FL/L2 classrooms has changed. Many language-teaching professionals, even advocates of communicative language teaching (e.g., Ellis, 2001; Long, 2000; Savignon, 2005) who initially argued against explicit grammar instruction, have

become increasingly aware of the importance of grammar teaching. For example, Celce-Murcia and Larsen-Freeman (1999) and Wang (2010) argue that teaching language competencies cannot be efficiently achieved in the absence of proper grammar teaching in FL/L2 classrooms. In this regard, Norris and Ortega (2001) found that L2 grammar teaching results in substantial gains in the target grammatical structures and that the gains are sustained over time. Moreover, Long (1991) argues that ‘a systematic, non-interfering focus on form produces a faster rate of learning and (probably) higher levels of ultimate [L2] attainment than instruction with no focus on form’ (47). Focus on form here refers to grammar instruction that takes place within communicative contexts. (For an overview of the changes in place of teaching grammar over the years, see Hinkel and Fotos, 2001a; Larsen-Freeman, 2001, 2003; Nassaji and Fotos, 2011).

Because teaching grammar has recently become an integral part of FLT, and because teachers have always acknowledged the inextricable link between teaching and testing, grammar testing has accordingly become an indispensable aspect of L2 teaching and classroom pedagogy. Teachers do need to determine if the grammatical rules they taught have been mastered by their students. They do so using different types of assessment tasks, e.g., multiple-choice (MC), matching, gap-fill and word formation. Of these, MC is the most common way of testing grammatical knowledge. This is because MC tests can cover many grammatical points and at the same time is easy to score using technologies, yet it also offers greater objectivity in grading compared to other testing formats. MC questions (MCQs), in their simplest form, require the selection of a correct answer from a set of alternatives (for more detail about MC tests, see Bachman and Palmer, 1996; Hughes, 2003; Brown and Abeywickrama, 2010; Bailey and Curtis, 2015).

Despite the popularity of MC assessment tasks among language educators, and despite the numerous debates on L2 grammar teaching and its role in language learning, there

is quite limited research on whether this test format is suitable to test L2 learners' grammatical ability. I am only aware of two studies that have indirectly addressed this question. Liao (2009) and Vafaei et al. (2012) examine the relationships between knowledge of grammatical form and semantic meaning in the context of L2 use by using MC tasks. These studies found that L2 learners' knowledge of grammatical forms is not only related to their knowledge of the semantic meanings, but it is also related to the ability to use these forms meaningfully in a variety of contexts. This finding, since it was obtained from studies that used MC tasks as elicitation techniques, suggests that the MC test format is suitable to test L2 learners' grammatical ability. Note that researchers (e.g., Keck and Kim, 2014; Purpura, 2004) draw a distinction between grammatical knowledge and grammatical ability; grammatical knowledge refers to knowledge of language forms (phonological, morphological, syntactical, etc.), whereas grammatical ability refers to the capacity to use these forms accurately and meaningfully in communication.

It should be mentioned at this point that there is a plethora of studies that have examined the validity of using MC tasks to assess grammatical knowledge. These studies have produced inconclusive results. While some researchers claim that MC testing format is not only valid to assess knowledge of grammar, but it also has high reliability since it provides a relatively consistent result, others argue that this testing format is not fully appropriate to assess grammatical knowledge (see among many others Abosalem, 2016; Bachman and Palmer, 1996; Bailey, 1998; Currie and Chiramanee, 2010; Dávid, 2007; Haladyna, 2004; Haladyna, Downing and Rodriguez, 2002; Mansory, 2020; Quansah, 2018).

The little research on the suitability of MC testing to assess students' grammatical ability, in my view, could be attributed to the fact that the term grammar is viewed by many language-teaching professionals as a set of rules governing the structures in a language, e.g., in English, the verb must agree with its subject in person and number,

so native speakers of English say ‘she teaches English’ and not ‘she teach English’. Such a theoretical view of grammar indeed influences the assessment of L2 grammatical ability. This is because the adopted theoretical view/framework of grammar defines what it means to have L2 grammatical ability, which in turn affects the assessment goal and procedure.

An important question to be raised on the basis of the above discussion is the following: Does knowledge of grammar inferred from the ability to select a grammatically correct answer from several options on a MC test provide adequate information about learners’ ability to use the target grammatical form(s) accurately, meaningfully and appropriately? In other words, is the MC test format suitable for properly measuring L2 learners’ grammatical ability? This is the research question that will be addressed theoretically in the next section and empirically in the sections that follow.

2. Defining Grammar for the Purpose of L2 Teaching and Assessment

The term grammar is polysemous; this is because different linguistic theories have different definitions for grammar, and it follows that they have different explanations for its learning and teaching. Yet ‘most theorists agree that no one theory will suffice to cover everything about grammar, its learning, and its teaching’ (Larsen-Freeman, 2020, 117). They argue that each definition has its own limitations to provide a theoretical and practical basis for grammar teaching, learning and testing (for detailed criticism and discussion, refer to Larsen-Freeman, 2020; Purpura, 2004).

To address this need to have a precise definition, at least for pedagogical purposes, in her seminal works, Diane Larsen-Freeman (2001, 2003, 2009) defines grammar as ‘a system of meaningful structures and patterns that are governed by particular pragmatic constraints’. (2009, 521). This definition proposes a pedagogic framework for grammar; it encompasses three components: form, meaning and use.

Larsen-Freeman (2001, 252–253) defines these components as follows:

- Form refers to ‘how a particular grammar structure is constructed and how it is sequenced with other structures in a sentence or text.
- Meaning refers to ‘what a grammatical structure means.
- Use addresses questions such as ‘when or why does a speaker/writer choose a particular grammatical structure over another? And when or why does a speaker/writer vary the form of a particular linguistic structure?’

Larsen-Freeman (2009, 521) illustrated these three components of grammar with a common structure in English – the passive voice. The way of forming passive construction requires, for example, some form of the ‘be’ verb and the past participle. It has the grammatical meaning of communicating something about/to which something happens/happened. The passive is used when the agent is unknown, should be concealed or when the use of the passive reflects the preferred word order for marking given and new information, among other uses. The speaker must also be able to distinguish contexts in which the passive voice is more appropriate than the active voice and make a choice between them when variation is possible in contexts when the two structures have approximately the same meaning.

I think this definition of grammar is sufficiently focused to fulfil its teaching and testing functions. This is because it explains what it means to have the grammar of a language. To put it another way and to set the discussion in the context of FL, this definition explains what exactly it means to have FL grammatical ability – which is ‘the ability to use grammar structures accurately, meaningfully, and appropriately’ (Larsen-Freeman, 2003, 6).

With this definition in mind, we arrive at the conclusion that the primary goal of FL classroom-based grammar testing is to measure learners’ grammatical ability – that is, to test if the students can use the target grammatical form(s) accurately, meaningfully

and appropriately. To realise this goal, it is not sufficient for an effective test to measure if L2 learners know grammatical structures; it also needs to assess if they can use them meaningfully in a variety of contexts – namely, if they know which forms are appropriate and typical in a given situation. This is because if L2 learners do not know how to apply grammatical knowledge in real-life communication, knowledge of grammar is not going to be useful, because the goal of SLL is ‘to develop communicative competence and to enable learners to use language accurately and fluently for real communicative purposes’ (Nassaji and Fotos, 2011, 14).

It should be mentioned at this point before completing this section that although Larsen-Freeman’s model of L2 grammar teaching provides the most plausible answer to this question of what it means to have L2 grammatical ability, ‘very little has been said about the assessment of grammatical ability, and unsurprisingly [. . .] not much has changed since the 1960s. In other words, for the past fifty years, grammatical ability has been defined in many instances as morphosyntactic form’ (Purpura, 2004, 253). Thereby, grammatical ability is still tested using mainly MCQs.

3. Methodology

This study was conducted in English as a foreign language (EFL) grammar classes at the Department of Languages and Translation of Taibah University. Originally, a total of 92 first-year undergraduate students were recruited for this study. Their levels of English varied, ranging from elementary to lower-intermediate to upper-intermediate to advanced, according to their scores on the Oxford Quick Placement Test. The three advanced participants were excluded/removed from the study’s final analysis by the researcher, who functioned as the tutor, because they had lived in an English-speaking country for a long period during childhood and therefore had acquired English natively. Some participants were also excluded because they did not complete both tasks as expected and instructed, e.g., left some of the experimental questions

unanswered. Table 1 offers an overview of the participants involved in the study and their proficiency levels (PLs).

Table (1): Participants involved in the study and their proficiency levels

| PL* | Participants who originally participated in the study | Participants included in the study analysis |
|------------------------------|---|---|
| Elementary (Elem) | 39 | 28 |
| Lower-intermediate (LI) | 24 | 18 |
| Upper-intermediate (UI) | 26 | 22 |
| Advanced | 3 | 0 |
| Total number of participants | 92 | 68 |

*See Appendix 1 for more information about the Oxford Quick Placement Test and its proficiency-levels classifications scale.

To examine the research question, the data was gathered through two elicitation tools that were part of a midterm exam: an MC task and a gap-filling task (GF task). The former is used to assess if the participants know the target grammatical forms (tense/aspect forms in English), whereas the latter is used to assess if they can use these forms meaningfully in their production. The MC task comprised 12 sentences. The participants must choose the correct answer from four options. Examples of the MCQs are given below (see Appendix 2-B for a complete listing of the test sentences presented in the MC test).

1. Look! That man _____ to open the door of your car.

| | | | |
|----|-----------|----|-------|
| a. | try | b. | tried |
| c. | is trying | d. | tries |

2. Ali: Why did you buy all this sugar and chocolate?
Ahmed: I _____ a delicious chocolate cake for dinner tonight.

| | | | |
|----|------------------|----|----------|
| a. | am going to make | b. | make |
| c. | will make | d. | has made |

The second elicitation task also comprised 12 sentences – ones that were the same to the ones used in the MC task but presented in the GF format (see Appendix 2-A for a complete listing of the test sentences presented in the FG test). Compare the following examples in 3 and 4 with the examples in 1 and 2:

| | |
|----|---|
| 3. | Look! That man (try) _____ to open the door of your car. |
| 4. | Ali: Why did you buy all this sugar and chocolate? Ahmed: I (make) _____ a delicious chocolate cake for dinner tonight. |

This task is designed to require learners to use certain grammar structures, i.e., tense/aspect forms. Note that the gaps are preceded by one lexical item which must be transformed in order to fill the gap correctly. The rationales behind this procedure – giving the participants a cue word which they need to fit into the sentence by changing its form – are to eliminate the possibility of having several possible answers and to eliminate the possibility that the learner would be unable to find the right L2 words to fill the gaps.

Both tasks were administrated on the same day. However, the order in which tasks appeared during testing was taken into consideration; the GF task was followed by the MC task. The GF task was to be completed before the MC task to minimise the possibility that the participant might use the options given under the MC sentences to fill in the blanks in the GF sentences, given that the two tasks have the same sentences but are presented in different formats.

The MCQs were marked as correct or incorrect. If the participant chooses the correct answer, they score one point, if they choose any of the distracters, they score a zero. The procedure used to mark the GF task was similar to that already used to mark the MC task – one point was given for each correct answer and zero points for each incorrect one. This marking method was carried out using Microsoft Excel 2016.

First, for the purpose of descriptive statistics, each proficiency group's results were reported, including information about the mean, minimum, maximum, standard deviation and percentage of correct answers for both tasks. Note that due to the large number of participants involved in the study and space limitations, only the group results were presented, examined, compared and discussed in this study.

Then, to prepare the data for the next descriptive and inferential statistical step, the Excel data sets on both tasks were imported into the R Statistical Programme. After that, using the Shapiro test, the sample's normality of distribution was checked to pave the way for running the appropriate statistical inferential procedures. If the data were normally distributed ($p > .05$), paired sample t-tests were used to compare two variables (e.g., to determine any statistically significant differences between the means of the two formats, namely, MC and GF) within the same group of proficiency. Alternatively, the Wilcoxon test was used for the same purpose if the data were not normally distributed. The alpha level was set at $p < 0.05$ for all of these tests, meaning that a result is considered significant if $p < 0.05$. In the following section, the data will be presented and analysed.

4. Data Analysis

This section is meant to present the results that emerged from the data elicitation tools – the MC task and the GF task. To properly address the research question stated in section one, it is important to first present general descriptive statistics of the empirical data and then compare the overall correct performance in the MC task to that in the GF task within each group of learners. Consider the following table:

Table (2): Descriptive statistics for the correctly answered items in the two tasks for the three participant groups of proficiency

| LP | Task | No. of participants | Total No. of items | No. of correctly answered items | % of correctly answered items | Min | Max | Mean | Median | St. Deviation |
|------|------|---------------------|--------------------|---------------------------------|-------------------------------|-----|-----|-------|--------|---------------|
| Elem | MC | 28 | 336 | 167 | 50% | 2 | 9 | 5.96 | 6 | 1.59 |
| Elem | GF | 28 | 336 | 46 | 14% | 1 | 3 | 1.64 | 1 | 0.81 |
| LI | MC | 18 | 216 | 158 | 73% | 5 | 10 | 8.78 | 9 | 1.47 |
| LI | FB | 18 | 216 | 104 | 48% | 4 | 7 | 5.78 | 6 | 1.03 |
| UI | MC | 22 | 264 | 238 | 90% | 8 | 12 | 10.82 | 11 | 1.07 |
| UI | FB | 22 | 264 | 192 | 73% | 8 | 10 | 8.73 | 8 | 0.86 |

This table reveals that the Elem learners, as a group, performed much better in the MC task than they did on the GF one. They correctly answered far more sentences with MC options than sentences with blanks to be filled in with appropriate words. The means of the correctly answered items are 5.79 (StdDev 1.93) and 1.64 (StdDev 0.81) on the MC task and the GF task, respectively.

This trend holds true throughout all the LI and UI groups as well, meaning that these learners exhibit similar patterns of behaviour to their Elem-level counterparts, but of course with higher degrees of accuracy. As groups, they score much higher in the MC task, with mean scores of 9 (StdDev 1.70) and 10.82 (StdDev 1.07), than on the GF one, with means of 5.78 (StdDev 1.03) and 8.73 (StdDev 0.86) for the LI group and the UI group, respectively.

For better comparative visualisation, the results in Table 2 are summarised below in Figure 1 in terms of percentage of correctly answered items in both tasks for each group of learners.

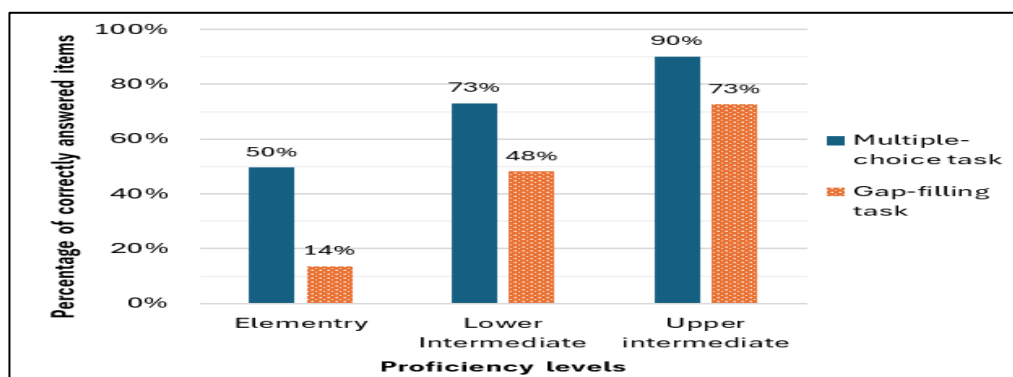


Figure (1): Pairwise comparisons between the MC task and the GF task performances

The differences among these results suggest that learners' performances vary from task to task. This suggestion was proved for all three groups of L2 learners through the results of inferential tests. The differences in performance within each group of learners across the MC task and the GF task were all shown to be statistically significant by Wilcoxon tests ($P\text{-value} = 0.000 < 0.05$, $P\text{-value} = 0.000 < 0.05$ and $P\text{-value} = 0.000 < 0.05$ for the Elem learners, LI learners and UI learners, respectively). The results of the normality tests applied to select the appropriate parametric or nonparametric statistical techniques (paired samples t-tests vs. paired Wilcoxon tests) are found in Appendix 3.

5. Discussion

This study investigates whether the MC test format can measure L2 learners' grammatical ability. To examine this question, the study tries to indicate if the participants, having selected grammatically correct answers from several options on a MC test, could use the target grammatical form(s) accurately, meaningfully and appropriately in a variety of contexts. The combined results of the two data sources made it evident that the participants, regardless of their proficiency levels, did not perform consistently across the different task types – their performance in the GF task

was much poorer than their performance in the MC task. Most participants, including those of lower linguistic proficiency levels, can select, to a great extent, a grammatically correct answer from several options on a MC test, yet they have great difficulty using the target grammatical form(s) accurately and meaningfully to fill in the blanks in the GF test. The results of the statistical tests reveal that the learners' scores obtained on the MC test differ significantly from the GF one, regardless of their proficiency levels. Learners' MC scores are significantly better than their GF scores. Therefore, the results indicated that learner answers in a MC test do not accurately reflect the learners' true ability to use grammar. This result is inconsistent with the finding of the two studies done by Liao (2009) and Vafae et al (2012), which suggest the MC test format is suitable to test L2 learners' grammatical ability. However, this finding is consistent with that of the studies conducted to investigate another element of language competency, namely, L2 learners' lexical knowledge and use (see Amini and Ibrahim-González, 2012; Kılıçkaya, 2019; Kremmel and Schmitt, 2016). These studies also show that learner answers in an MC test do not reflect learners' true competency that the test is claiming to measure.

Given this finding, the author can suggest caution in using MC testing as a single data source to assess L2 learners' grammatical ability. The problem with this testing format is that it focuses solely on grammatical accuracy when assessing grammatical ability. Therefore, it does not provide information on whether learners can use target grammatical form/rule meaningfully in a variety of contexts.

It should be stated at this point that it is too early to generalise this finding, not only because there is quite limited research on whether MC testing can measure grammatical ability, but also because the results have shown that the learners, as their proficiency in English increases, are becoming increasingly better at applying their grammatical knowledge to complete the GF task (refer back to Table 3 and Figure 1). Therefore,

future studies with larger sample sizes including advanced learners are required to investigate this question in depth.

6. Conclusion

This study sought to investigate the efficacy of MC tests in assessing L2 or FL learners' grammatical proficiency. Through comparative analysis of MC test and GF task administered to 68 English FL learners, significant disparities in performance were observed. The findings suggest that while MC tests are commonly used to evaluate grammatical knowledge, they may not consistently capture learners' ability to apply grammar accurately and meaningfully across varied contexts.

The observed discrepancies underscore the limitations of MC testing as a sole indicator of grammatical proficiency. Specifically, MC tests primarily assess recognition rather than production of grammatical structures, potentially overlooking learners' capacity to use grammar in communicative settings. This study highlights the need for a nuanced approach to language assessment, one that integrates diverse evaluation methods to more comprehensively measure learners' grammatical competence.

Further research is essential to delve deeper into the implications of these findings. Future studies could explore alternative assessment strategies that better align with the multifaceted nature of grammatical proficiency in language learners. By refining assessment practices, educators and researchers can better support language teaching and learning initiatives aimed at fostering robust grammatical skills across diverse linguistic contexts.

Funding: This research is not funded.

Acknowledgments: I wish to express my deepest gratitude to Professor Hamza Alshenqeeti (Taibah University) for many fruitful remarks. Special thanks also go to

the students who participated in this study. However, no one but me can be held responsible for remaining errors.

Conflicts of Interest: The author declares no conflict of interest.

Authenticity: This manuscript is an original work.

Artificial Intelligence Statement: AI and AI-assisted technologies were not used.

References

1. Abosalem, Y. (2016). Assessment techniques and students' higher-order thinking skills. *International Journal of Secondary Education*, 4 (1), 1-11.
2. Amini, M., and Ibrahim-González, N. (2012). The washback effect of cloze and multiple-choice tests on vocabulary acquisition. *Language in India*, 12(7), 71-91. Retrieved March 8, 2018, from <http://www.languageinindia.com/july2012/v12i7july2012.pdf>.
3. Bachman, L. F., and Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
4. Bailey, K. (1998). *Learning about Language Assessment: Dilemmas, Decisions, and Directions*. Heinle & Heinle, Pacific Grove, CA.
5. Bailey, K. M., and Curtis, A. (2015). *Learning about language assessment: Dilemmas, decisions, and directions* (2nd ed.). Boston, MA: National Geographic Learning.
6. Brown, H. D., and Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). New York, NY: Pearson.
7. Celce-Murcia, M. (2001). *Language Teaching Approaches: An Overview*. In M. Celce-Murcia (Ed.), *Teaching English as a Second or Foreign Language* (3rd Ed.). Boston: Heinle & Heinle Publishers.
8. Celce-Murcia, M., and Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course* (2nd ed.). Boston, MA: Heinle and Heinle.

9. Currie, M., and Chiramanee, T. (2010). The effect of the multiple-choice item format on the measurement of language structure. *Language Testing*, 27(4), 471–91.
10. Dávid, G. (2007). Investigating the performance of alternative types of grammar items. *Language Testing*, 24 (1), 65–97.
11. Ellis, R. (2001b). Introduction: Investigating form-focused instruction. *Language Learning*, 51, 1–46.
12. Haladyna, T.M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Mahwah, NJ: Lawrence Earlbaum Associates.
13. Haladyna, T.M., Downing, S.M., and Rodriguez, M.C. (2002). A review of multiple choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309–334.
14. Hinkel, E. and Fotos, S. (2001a). *New Perspectives on Grammar Teaching in Second Language Classrooms* New York: Routledge.
15. Hinkel, E. and Fotos, S. (2001b). From Theory to Practice: A Teacher's View. In E. Hinkel, & S. Fotos (Eds.), *New Perspectives on Grammar Teaching in Second Language Classrooms* (pp. 1-12). New York: Routledge.
16. Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.
17. Keck, C. and Kim, Y. (2014). *Pedagogical Grammar*. Amsterdam: John Benjamins.
18. Kılıçkaya, F. (2019). Assessing L2 vocabulary through multiple-choice, matching, gap-fill, and word formation items. *Lublin Studies in Modern Languages and Literature*, 43 (3), 155-166.
19. Kremmel, B., and Schmitt, N. (2016). Interpreting vocabulary test scores: What do various item formats tell us about learners' ability to employ words? *Language Assessment Quarterly*, 13(4), 377-392. DOI: 10.1080/15434303.2016.1237516

20. Larsen-Freeman, D. (2001). Teaching grammar. In M. Celce-Murcia (Ed.), Teaching English as a second or foreign language (3rd ed., pp. 251–85). MA: Heinle & Heinle.
21. Larsen-Freeman, D. (2003). Teaching language: From grammar to grammaring. Boston: ThomsonHeinle.
22. Larsen-Freeman, D. (2009). Teaching and testing grammar. In M.H. Long & C. J. Doughty (eds.) The Handbook of Language Teaching (p.518-542). Oxford: Blackwell.
23. Larsen-Freeman, D. (2020). Teaching and Researching Grammar skills: Theory-and Research-Based Practices. In N. Polat, T. Gregersen & P. MacIntyre (eds.), Research-Driven Pedagogy: Implications of L2A Theory and Research for the Teaching of Language Skills (pp. 97-124). New York, NY: Routledge.
24. Larsen-Freeman, D., and Anderson, M. (2011). Techniques and principles in language teaching (3rd ed.). Oxford, UK: Oxford University Press.
25. Liao, Y.-F. A. (2009). Construct validation study of the GEPT reading and listening sections: Re-examining the models of L2 reading and listening abilities and their relations to lexico- grammatical knowledge (Unpublished dissertation). Columbia University.
26. Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. DeBot, R. Ginsberge, & C. Kramersch (Eds.), Foreign language research in cross- cultural perspective (pp. 39–52). Amsterdam: John Benjamins.
27. Long, M. (2000). Focus on form in task-based language teaching. In R. D. Lambert, & E. Shohamy (Eds.), Language policy and pedagogy: Essays in honor of A. Ronald Walton (pp. 179–92). Philadelphia, PA: John Benjamins.
28. Mansory, M. (2020). The significance of non-traditional and alternative assessment in English language teaching: Evidence from literature. International Journal of Linguistics, 12(5), 210-225.

-
29. Mayo MDPG, Mangado MJG and Adrián MM (2013) Contemporary approaches to second language acquisition. John Benjamins Publishing.
 30. Moreno, R., Martínez, R. J., & Muñiz, J. (2006). New guidelines for developing multiple-choice items. *Methodology*, 2(2), 65–72. <https://doi.org/10.1027/1614-2241.2.2.65>.
 31. Nassaji, H., and Fotos, S. (2011). Teaching grammar in second language classrooms: Integrating form-focused instruction in communicative context. New York: Routledge.
 32. Norris, J. M., and Ortega, L. (2001). Does type of instruction make a difference? Substantive findings from a meta-analytic review. *Language Learning*, 51, 157–213.
 33. Purpura, J. E. (2004). Assessing grammar. Cambridge: Cambridge University Press.
 34. Quansah, F. (2018). Traditional or performance assessment: What is the right way to assessing learners? *Research on Humanities and Social Sciences*, 8(1), 21–24.
 35. Richards, J. and Rodgers, T. (2001). Approaches and Methods in Language Teaching. Cambridge: Cambridge University Press.
 36. Savignon, S. (2005). Communicative language teaching: Strategies and goals. In E. Hinkel (Ed.), *Handbook on research in second language teaching and learning* (pp. 635–51). Mahwah, NJ: Lawrence Erlbaum Associates.
 37. Vafaei, P., Basheer, N., and Heitner, R. (2012). Application of confirmatory factor analysis in construct validity investigation: The case of the grammar sub-test of the CEP placement exam. *Iranian Journal of Language Testing*, 2(1), 1–19.
 38. VanPatten, B. and Williams, J. (2015). Theories in second language acquisition: An introduction. New York, NY: Routledge.
 39. Wang, F. (2010). The necessity of grammar teaching. *English Language Teaching*, 3(2), 78–81.
-

Appendices

Appendix 1: Oxford Quick Placement Test

A PDF version of the test can be downloaded from the following website:

<<https://talk.m2.res.zabanshenas.com/original/3X/2/3/23334e1d02a0d2533c91397eb44159d2ee30417a.pdf>>
> accessed 19/05/2022.

Appendix 2: The Experimental Tests

A. The Gap-Filling Task

- 1) Look! That man (**try**) _____ to open the door of your car.
- 2) The flight leaves Jeddah at 09:15 and (**arrive**) _____ in Riyadh at 10:40.
- 3) **Ali:** May I speak to Dr. Mohammed, please?
Ahmed: I'm sorry, he (**see**) _____ a patient at the moment.
- 4) I (**buy**) _____ a new car three days ago.
- 5) John phoned while we (**have**) _____ dinner.
- 6) What (**do**) _____ at this time yesterday?
- 7) The moon (**go**) _____ round the sun in about 28 days.
- 8) **Ali:** Why did you buy all this sugar and chocolate?
Ahmed: I (**make**) _____ a delicious chocolate cake for dinner tonight.
- 9) At this time tomorrow, I (**take**) _____ a test.
- 10) When I was young, I (**want**) _____ to be a pilot.
- 11) Mohammed and Fatima (**talk**) _____ on the phone every night.
- 12) **Ali:** Could someone get me a glass of water?
Ahmed: Certainly, I (**get**) _____ you one.

B. The Multiple-Choice Task

- 1) Look! That man _____ to open the door of your car.
a. tried
b. tries
c. trying
d. is trying
- 2) The flight leaves Jeddah at 09:15 and _____ in Riyadh at 10:40.
a. arriving
b. were arriving
c. arrive
d. arrives

Appendix 3: List of Shapiro Results

| PL | Task | P-value | Normality of distribution* |
|------|------|----------------|----------------------------|
| Elem | MC | $0.116 > 0.05$ | normally distributed |
| Elem | GF | $0.000 < 0.05$ | not normally distributed |
| LI | MC | $0.000 < 0.05$ | not normally distributed |
| LI | GF | $0.000 < 0.05$ | not normally distributed |
| UI | MC | $0.000 < 0.05$ | not normally distributed |
| UI | GF | $0.000 < 0.05$ | not normally distributed |

*The p-value > 0.05 implying that the distribution of the data is not significantly different from normal distribution. In other words, we can assume the normality.