Full Length Research Paper

The effectiveness of using multi devices technique on testing grammar of university students

Said Fathy Abdul Fatta
Qassim Private Colleges, Saudi Arabia.

Received 17 August, 2015; Accepted 07 October 2015

The present study was an attempt to determine the effectiveness of using multi-devices technique to develop students’ grammatical ability. Participants were 40 Preparatory Year Treatment students from Qassim Private Colleges in Saudi Arabia. The experimental group (N 20) used multi devices technique for testing grammar; online testing, mobile testing, machine testing (auto test) and pen and paper testing technique whereas the control group (N20) was tested through two exams; midterm and final exams in a paper and a pen format. The pre-post-test comprised two parts, the first part includes multiple-choice questions, the second part includes fill in gaps questions. Results of the t-test analysis revealed that using multi devices technique yielded significant effects on students’ grammar, i.e. the experimental group outperformed the control group.

Key words: Multi-devices, Testing grammar, mobile testing, online testing.

INTRODUCTION

Qassim Private Colleges’ students come to their classes equipped with electronic devices, including smartphones, tablets and laptops. In addition, they use these devices frequently during class for non-course-purpose like chatting, emails, WhatsApp and calling. There are also some uses during course-related like dictionary for translation, grammar checker and spelling checker. It is difficult to deprive these students from using such available technology in their classes. Therefore, we should benefit from such devices in our teaching process. Testing as a sine qua non of teaching process shows teachers how successful, their teaching has been. It provides wash back for them to adjust and change course content and teaching styles where necessary according to the test result. In addition, it provides us with a full picture about the effectiveness of the course book, materials, and methods. At the same time, it represents the starting point of identifying areas for remedial work. Whereas testing has this effective role in language-teaching process, course designers should provide learners with different techniques for language testing. It will be difficult to determine course or treatment validity on just on form exam. Developing a test, however, is often done by individual teachers, which is laborious work, with the quality of the questions depending on the individuals’ expertise (Hoshino, 2009).
The purpose of language tests

Brown (1996) identifies four purposes for language test. They can be summarized in the following points:

1. Proficiency Test:
   It aims to determine the level of individual in relation to other testees.
2. Placement Test:
   It aims to determine closely the level of individual.
3. Achievement Test:
   It aims to determine the achievement of individual in the program.
4. Diagnostic Test:
   It aims to determine the strength and weakness of individual in the program.

Testing grammar

According to Larsen-Freeman (1997), grammar comprises three dimensions: form, meaning and use and hence the goal of grammar testing is to test learners to use grammar accurately, meaningfully, and appropriately. So, it means when we test our students we test learners’ grammatical knowledge from all three aspects (morphosyntax, semantics and pragmatics). Therefore, one or two exams during term are not enough to assess students’ grammatical ability.

The need for multi-devices testing technique

The need for multi-device testing technique could be summarized in the following points:

Device availability

Multiple devices have multiplied learners’ expectations. The mobile masses demand both flexibility and convenience and expect content to be available at the point-of-need, and time-of-interest.

Bringing your own device (BYOD)

Schools and universities have been encouraging students to bring their own devices to the campus to better exploit the potential learning, interaction and even collaboration experiences that these devices can afford. Companies too support this trend and promote it to save both time and money. The practice of bringing your own device, popularly known as BYOD, allows individuals to take eLearning courses on their phone, for example, while on their way to the office or a class, continue the course at their desk and then complete it on their tablet while returning home.

Computing changing

Computing trends have changed; from huge mainframes to small personal desktop computers and mobile devices, from personal networking to the Internet and cloud computing. Learning solutions today need to keep up with the latest technologies and be served up on the latest devices.

An increasing number of millennial entering the workforce

There is a clear increase in the number of college students all over the world entering the work market every year. Therefore, this workforce should be qualified with new technologies (Khairajani and Bellare, 2015).

Context of the problem

Based on the researcher’s current experience as an EFL instructor for about 2 years at Qassim Private Colleges (QPC), the following observations were made. These include the fact that the evaluation process in Qassim Private Colleges in teaching English language Department relies on two criteria; a midterm exam graded with 20%, attendance and home assignment graded with 20% of the total mark and 60% for the final exam. The midterm and final exams brought for the students in pen and pencil format. This way is unfair in evaluating and judging students all over the term in specific course. Many problems happen in adopting this criterion of evaluation like the only chance for the test takers, specific task types and students’ low achievement. One of these courses that follow the same criterion is grammar course for fresh students in Preparatory Year Treatment. This was the first time to the students to be exposed for specific English courses. The students do not perform well in their midterm exam or the final exam. This was clear in their results during the previous terms. The teacher plays the role of examiner or grade-giver; being satisfied with marking in students’ grammatical errors in the midterm or final exams. When he supplies students with feedback on their errors, the time is over as it is one time exam. So, the students feel disappointed about their learning process.

Statement of the problem

Thus, the study problem is identified in Preparatory Year Treatment (PYP) students’ poor mastery of grammar. Therefore, the current study attempted to develop these
students’ grammar through adopting multi devices techniques. To reach this aim, the study tried to answer the following main question:

What is the effectiveness of adopting multi devices techniques in developing PYP students’ performance in grammar?

This main question is divided into the following sub-questions:

1. What are the theoretical bases of testing grammar in the light of the multi devices technique?
2. What are the appropriate task types that will be included in testing grammar through using multi devices technique?
3. What steps should be undertaken for designing grammar task types in the light of the multi-devices technique?

Hypotheses of the study

Hypotheses comparing the experimental groups and control groups mean scores on the post test:

1. There is a statistically significant difference between the mean scores of the experimental groups – receiving multi devices testing techniques – and the control groups – receiving regular testing techniques – on the post-test in overall performance in grammar in favor of the experimental group.
2. There were statistically significant differences between the mean scores of the experimental group on the pre-test and the post-test in overall performance in grammar in favor of the post test.
3. There were statistically significant differences between the mean scores of the experimental groups and the control groups in their final exam at the end of the term in favor of the experimental group.

Aim and objectives of the study

The study attempted to employ the multi-devices technique in testing grammar to develop PYP students at QPC in overall performance in grammar.

Definition of terms

Testing

A test, in simple terms, is a method of measuring a person’s ability knowledge, or performance in a given domain. (Brown, D. 2003).

Technique

It is the way of doing some tasks or performing something

Device

It is an object or a piece of equipment that has been designed to do a particular job.

Multi- devices

Using more than one device

Grammar

The science teaches the correct forms of English language.

RELATED STUDIES

Through reviewing literature, the research dealt with two parts of literatures; the first is types of testing grammar and the other is using more than one technique and device in teaching.

Formats of testing grammar

There are many types of testing grammar like

1. Multiple choice format

There is a sentence called a head and a number of choices. This number of choices may vary, but most of Multiple choice items have four choices because this number is considered to be the best exchange between lower the guessing level and the work of creating distractors.

2. Cloze test format

It requires from students to fill in the gap with one word. In the cloze procedure words are deleted from a text after allowing a few sentences of introduction. The deletion rate is mechanically set, usually between every fifth and eleventh word. Candidates have to fill each gap by supplying the word they think has been deleted.

3. Error recognition items

It requires from students to correct one or more mistakes found in the given sentence.

4. True or false items

It requires from students to tick if this sentences is grammatically right or wrong.
Related studies to Testing grammar

Chunyi et al. (2014) stated that with the advanced development of mobile technology and portable devices, learners can do their learning activities anytime and everywhere when Internet access is available. Integrating pedagogical and technical strengths of mobile technology into learning settings proves imperative in previous research. Thus, this study aims to investigate the effects of implementing a ubiquitous multimedia message-transmitting platform (C&U-Message) on college students' English learning. A total of 26 college students participated in a 6-week experiment used client-side application system C&U-Message (C&U-Msg) system for English learning through Android-based mobile phones. Data collected from the pre- and post-tests and a learning satisfaction survey questionnaire were analyzed. The findings of this study reveal that C&U message application on mobile English learning, learning content for mobile English learning, practical use of C&U message, user satisfaction with learning achievements, use of digital learning materials, and user's attitude toward learning language through mobile devices. Finally, the C&U-Msg system can be effectively utilized for English learning.

Mueller and The pen is mightier than the keyboard: Advantages of longhand over laptop note taking (2014) compare between taking notes on laptops and in longhand. He noticed that taking notes on the laptops are rather than in longhand is increasingly common. Many researchers have suggested that laptop note taking is less effective than longhand note taking for learning. This research suggests that even when laptops are used solely to take notes, they may still be impairing learning because their use results in shallower processing. The findings proved that taking notes on laptops performed worse on conceptual questions than students who took notes longhand. Whereas taking more notes can be beneficial, laptop note takers' tendency to transcribe lectures verbatim rather than processing information and reframing it in their own words is detrimental to learning.

Aleimi (2012) did a comparison between adopting cloze and multiple-choice test in testing grammar. To investigate the research, the candidates were screened through a NELSON test. This test enabled the researchers to have homogeneous subjects; therefore, among 84 subjects. The two groups were provided with 10 consecutive weeks, each session 100 minutes long. Every week one experimental group was treated by a 25-item cloze test of grammar while the other experimental group was provided with a 25-item multiple choice one. The analysis of data indicated that there was no significant difference between the average performances of the two groups. In another words, the findings proved that the cloze procedure did not improve the subjects' knowledge of grammar any more than the multiple-choice grammar test did.

Khoi and Shamsi (2012) investigated the effects of different test formats on the measurement of grammatical knowledge by comparing the construct validities of two different formats of error-identification grammar tests (one with four options for each item and one with a no error option as the fifth option) and a multiple choice grammar test. After administering all the three tests to 131 Iranian EFL learners, it was found that while the TOEFL error-identification test enjoyed the highest level of construct validity, the no-error option in the SAT error-identification test reduced its construct validity to a considerable degree, indicating that including no error options in admissions test does not allow an accurate evaluation of examinees' grammatical knowledge. Moreover, it lent support to the idea that the test method facet could be a strong source of bias in language testing affecting the fairness of the decisions in relation to admitting students to specific academic programs.

Mehrgan (2012), in his study, investigated the effects of adopting CALL on improving EFL learners' grammatical ability. To this aim, a multiple-choice test of grammar, of which the reliability was 0.79, was administered to 83 available TEFL students out of whom 60 students were selected as homogeneous and randomly divided into two groups of 30: a control group and an experimental group. The experimental group was taught English grammar through CALL and the control group received a non-CALL instruction on grammatical structures. The results of the study through a posttest revealed the fact that the experimental group outperformed the control group. Therefore, CALL appeared to be useful in developing English grammar of the TEFL students.

A study by Ophir et al. (2009) has examined the relationship between undergraduates' reported levels of "media multi-tasking"—or performing multiple, simultaneous activities in different media, including print, television, computer-based video, music, text messaging, instant messaging, web-surfing, email—and their performance on cognitive tasks measuring attention. The findings revealed that undergraduates in the study who were identified as "heavy media multi-taskers" were significantly more susceptible to distractions than were those who were identified as "light media multi-taskers."

Hoshino (2009) proposed a method for automatically generating Multiple-Choice (MC) Fill-In-the-Blank (FIB) questions for English grammar and vocabulary. He developed three AOG methods and conducted evaluations in terms of the proposed viewpoints. First, a semi-automatic question generation that allows the test author to compose questions just with some clicks on the screen. Second, a CAT (Computer Adaptive Testing) system that administers automatically generated questions. Third, he evaluated randomly-generated questions with a group of students. AOG method for grammar questions is as efficient as workbook questions. Evaluation on
Table 1. Questionnaire results.

<table>
<thead>
<tr>
<th>Question No.</th>
<th>The first choice</th>
<th>The second choice</th>
<th>The third choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>30%</td>
<td>30%</td>
<td>40%</td>
</tr>
<tr>
<td>2.</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3.</td>
<td>95% no</td>
<td>5% yes</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>25%</td>
<td>15%</td>
<td>60%</td>
</tr>
<tr>
<td>5.</td>
<td>45%</td>
<td>30%</td>
<td>25%</td>
</tr>
<tr>
<td>6.</td>
<td>25%</td>
<td>5%</td>
<td>70%</td>
</tr>
</tbody>
</table>

validity shows some evidence that the pattern-generated questions measure intended grammar targets.

METHODS

Design
The quasi-experimental design was used in this study, where a sample of two classes of the Preparatory Year Treatment (PYP) students was assigned to the experimental and control group. A pre/post test was conducted to measure the effectiveness of the treatment.

Participants
The subjects of this study were composed of (40) students, in Qassim Private College. The students were in the PYP in Qassim Private Colleges. These students belong to two different majors; English language Major and Computer Science Major. They study together four English courses; reading 001, writing 001, listening 001 and grammar 001. The code 001 refers to the first term in the preparatory year treatment. Their ages ranged between 18to 24 years. The students were then divided into two groups – the control group and the experimental group. Both groups were administered to the Pre-test which showed uniformity in the results with very little variation that shows the two groups are similar in their grammatical ability. Both the groups were instructed by two of the teachers for a full semester in grammar course that was designed for them.

Instruments
The researcher used two tools for collecting the data of the study:

The students' questionnaire
This questionnaire was adopted over 20 students, the experimental group, to know their attitudes towards using multi device technique in testing grammar. The questionnaire simply consisted of six questions (see appendix 1).

Analyzing the questionnaire results
It is clear that:

1. Nearly, most of the students preferred the online testing. That was clear in their results in online test. They saw it as a new experience through which they could relax and hid their fear of exams. Therefore, they represented forty percent of the sample.
2. When the students were asked to rearrange their preference test techniques, they chose online testing as the first preference followed by the mobile learning and a paper and a pen test.
3. Just only five percent of the students experienced online testing before. Therefore, they saw it as an attractive and enjoyable experience.
4. Online learning gave the students chance to feel they were good at grammar. But, the students confessed for the teacher that they asked for help from their friends or another teacher to answer some questions.
5. Mobile tests came as a second preference for the students. They saw as a new technique for them. They preferred it to traditional testing.

Finally, this questionnaire showed that there were more interest with online testing and mobile test for these students in return for the traditional testing.

b. The pre/post test
A pre/posttest was designed in the light of English for fresh students at PYP. This test was basically based on the content of the course that prescribed for the students to study during the first semester of the academic year 2014-2015. Then, it was administered to the students of both the experimental and control groups to make sure that they were at the same level of performance before starting the experiment; and hence the progress achieved by the experimental group students could be attributed to using multi device technique that they received through the whole semester. The pre-test was also used as a post-test to investigate the effectiveness of adopting multi device technique to develop grammar.

Description of the test
The translation pre-posttest was constructed in the light of the following sources:

1. Reviewing previous studies concerned with testing grammar.
2. The following points were taken into account when writing the test items:
   - The words and sentences used in the test items were simple, clear, and familiar to the students.

The Content of the test
It consisted of three parts;
1. Multiple Choice Questions.
The students are asked to answer 20 MCQ. These questions were related to the first eight units of grammar book. It included verb be, present simple tense, present progressive tense, past simple tense, and past progressive tense. Exam questions were divided equally between these topics. Each topic represented by four questions in this part.

2. Do as shown in brackets.
Students are asked to do specific change that is required and shown to them. It included twenty sentences.
   a. He does his homework at ten. (use: yesterday)
   b. The students were asked to change the sentences into past simple.

3. Fill in the space
The students were asked to fill in the space with the correct form of the verb. This part included 20 questions. The total mark of this test was (60) marks.

Validity of the test
To measure the test content validity, the first version of the test was given to 2 EFL assistant professors to evaluate appropriateness of each item in measuring students’ mastering these grammatical categories. Moreover, they were requested to evaluate the test as a whole in terms of: (a) correctness, (b) number of items, and (c) suitability of the test items to fresh students at PYP level at Qassim Private Colleges. The test proved to be mostly valid as the jury approved the questions and considered them suitable for their level as PYP students.

Piloting the test
Piloting the test aimed at (a) determining reliability of the test; (b) determining the suitable time to be allotted for the test; and (c) obtaining item analysis results, including item difficulty and item discrimination. Therefore, 10 students were selected for the pilot study. These students were randomly selected. Students of the pilot study belonged neither to the experimental group nor to the control group. They were excluded from the whole treatment.

Test reliability
The test was administered to a randomly selected group of 10 students. Then, the test was administered one more time after two weeks to the same 10 students. The Pearson correlation coefficient between the test/re-test results was calculated. The reliability coefficient was 0.75, which is relatively high. Therefore, the test could be considered a reliable one for the purpose of the current study.

Test time
It was estimated that one classroom period (approximately 40 min) would provide sample time for the students to read the test questions and write their answers. This time was calculated in the following way:

The time taken by the fastest student + the time taken by the slowest student.

The treatment
Duration
Adopting multi using techniques continued for one hundred days during the first term.

1. Aims of the treatment
These aims can be summarized in the following points:
   a. Providing students with more opportunities to have eight exams each term.
   b. Decrease students’ fear from exam or exam phobia in testing grammar.
   c. Content of the treatment
This treatment included nine exams through using different techniques. They were divided as follows; three online exams thorough using goggle forms, three exams through using WhatsApp application on the students’ mobiles and three pen and paper exams.

Validity of the treatment
To measure the test content validity, the first version of the tests was given to 2 EFL assistant professors to evaluate appropriateness of the treatment in the light of their objectives. Nearly, they approved all of them.

RESULTS
Results of this study will be reported in terms of the study hypothesis.

Hypothesis one
1- It stated: There is a statistically significant difference between the mean scores of the experimental groups – receiving multi devices testing techniques – and the control groups – receiving regular testing techniques – on the post-test in overall performance in grammar in favor of the experimental group.

To test this hypothesis, t-test for independent samples was conducted to compare the means between the experimental and control groups.

2- Then the t. test was computed between the two groups. Observed t. value, (2.36) with the value of the tabulated t at the level (0.05) was .003. It showed that t value obtained from the posttest was more than t table value. This means that there were a statistically significant difference between the mean scores of the experimental groups – receiving multi devices testing techniques – and the control groups – receiving regular testing techniques – on the post-test in overall performance in grammar in favor of the experimental group (Table 2).

Hence, the first hypothesis stated that there would be significant differences in posttest mean scores of both the
control and the experimental group in overall performance in grammar in favor of those of the experimental group, was verified.

Hypothesis Two

There were statistically significant differences between the mean scores of the experimental group on the pre-test and the post-test in overall performance in grammar in favor of the post test. To test the above hypothesis, the paired samples t-test was used, and the results are shown in Table 3.

Then the t. test was computed between the pretest and the posttest. Observed t. value, (−19.195) with the value of the tabulated t at the level (0.05) was .000. It showed that t value obtained from the posttest was more than t table value. Hence, the second hypothesis stated that there were statistically significant differences between the mean scores of the experimental groups on the pre-test and the post-test in overall performance in grammar in favor of the post test was verified.

Hypothesis three

This hypothesis was not directly included in the treatment. It was a comparison between students’ marks in grammar course at the final exam. This exam was held by one of the assistant professor at QPC. Both of control group and experimental group marks were analyzed by using t-test for independent groups.

Then the t. test was computed between the two groups at the final exam. Observed t. value, (3.989) with the value of the tabulated t at the level (0.05) was .000. It showed that t value obtained was more than t table value. Hence, the third hypothesis stated that there were statistically significant differences between the mean scores of the experimental groups and the control groups in their final exam at the end of the term in favor of the experimental group was verified (Table 4).

To sum up, the three hypotheses of the study were supported by the results. The statistical analyses of data indicate that the experimental group students who used multi-device techniques performed much better on the post-test in overall grammar than the students who received regular tests. Moreover, the experimental group students achieved significant progress in their final exam results after the treatment as compared to performance before the treatment. Hence, these positive findings proved the effectiveness of adopting multi testing techniques in developing university students’ grammar.

DISCUSSION AND CONCLUSION

Undoubtedly, grammar is an essential component of language. Linguists have been studying it for centuries, and it remains an object of learning for countless learners all over the world; it is an integral part of the language we

| Table 2. Post test results between control and experimental groups in overall performance in grammar |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Group           | N  | M   | S.D. | Std. Error Mean | t value | Sig. (2-tailed) |
| Experimental    | 20 | 29.50 | 10.58 | 2.36 | .2213 | .003 |
| Control         | 20 | 40.45 | 10.91 | 2.44 | -19.195 | .00 |

| Table 3. Test results comparing the experimental group students’ performance on the pre-test versus the post-test in overall performance in grammar |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Group           | N  | M   | S.D. | Std. Error Mean | t value | Sig. (2-tailed) |
| Experimental    | 20 | 18.90 | 10.52 | 2.35 | -19.195 | .00 |
| Control         | 20 | 40.45 | 10.91 | 2.44 | -19.195 | .00 |

| Table 4. t-test results comparing the experimental group and control group students’ performance at the final exam in overall grammar. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Group           | N  | M   | S.D. | Std. Error Mean | t value | Sig. (2-tailed) |
| pre-test        | 20 | 32.60 | 9.051 | 2.02 | 3.989 | .00 |
| post-test       | 20 | 44.70 | 10.10 | 2.25 | 3.989 | .00 |
people used technique in our courses through using Multi
devices, they got


The result of this study was in line with Chunyi (2014), in his findings when he stated that using C&U message application on mobile has an effect on English language learning. Moreover, the current study did not deal only with mobile applications to develop grammar but it dealt with other devices like online testing through using Google forms an also pen and paper tests.

In addition to Chunyi, this study was in line with Alemi (2012) when used two forms to test grammar; cloze test and multiple choices tests. His findings revealed that there is no significant effect between students' performance in each exam format. Moreover, the current study dealt with three formats for testing grammar. It also used three different devices.

Besides the previous studies, this study shares the importance of using more than one test formats with Khoii (2012). It supported the idea that the test method facet could be a strong source of bias in language testing. But this study added also that using one device for testing could affect our testing process and deprive the students from his chance to get better mark or better practice.

The result of this studies showed that when the students have more than one chance to be tested and have different test formats with different devices, they got better learning. This study found that most of the respondents in the experimental group showed improvement in their grammar performance. The experimental group shows a significant difference in its posttest and the final exam results.

There is another issue that is related to the questionnaire result. It revealed that students' fear of exam became less. They were not worried anymore. If they did not answer well in one test, they could do better again in the next test.

Recommendations

In light of the finding of the present study, the following recommendations seem pertinent:

1. Encouraging teachers to adopt to use multi device technique in developing other skills like writing and speaking.
2. Designing a part of our courses through using Multi-device technique.

3. Giving our students more than one test chance to be fair in our judgment upon them.

REFERENCES

Larsen-Freeman D (1997). Grammar and its teaching: challenging the myths. ERIC Identifier: ED406829