



# STUDENT-GENERATED TESTS AND THEIR IMPACT ON EFL STUDENTS' LEARNING OF GRAMMAR

(ÖĞRENCİLERİN ÜRETTİĞİ TESTLER VE ONLARIN EFL ÖĞRENCİLERİNİN  
DİLBİLGİSİ ÖĞRENİMİ ÜZERİNDEKİ ETKİSİ)

**Sasan BALEGHIZADEH<sup>1</sup>**  
**Zahra ZARGHAMI<sup>2</sup>**

## ABSTRACT

The present paper reports on a study that examined the impact of student-generated tests on grammar learning of EFL (English as a foreign language) learners. Sixty-eight Iranian intermediate university students were randomly assigned to experimental and control conditions. At the end of the treatment period, the participants in both groups took two forty-item grammar tests: a student-generated one and a standardized one. The results showed that the participants in the experimental group significantly outperformed their peers in the control group. This suggests that the experience of test construction throughout the treatment period had a positive impact on grammar learning of the students.

**Key words:** Student-generated tests; Assessment for learning; Alternative assessments; classroom assessment techniques

## ÖZET

Bu çalışma, İngilizce'yi yabancı dil olarak öğrenen öğrencilerin ürettikleri testlerin onların dilbilgisi öğrenimlerine etkisi hakkında bir araştırmayı rapor etmektedir. Bu bağlamda, orta düzeydeki 68 İranlı üniversite öğrencisi deney ve kontrol gruplarına yerleştirilmiştir. Araştırma sonucunda deney grubundaki öğrenciler kontrol grubundakilerden anlamlı derecede daha iyi bir performans sergilemişlerdir. Bu da test hazırlamanın öğrencilerin dilbilgisi öğrenimleri üzerinde olumlu bir etkisi olduğunu göstermektedir.

**Anahtar Kelimeler:** öğrencilerin ürettiği testler, alternatif değerlendirmeler, sınıf-içi değerlendirme teknikleri

---

<sup>1</sup> English Department, Shahid Beheshti University, G.C., Tehran, Iran.

**E-mail:** sasanbaleghizadeh@yahoo.com

<sup>2</sup> Allameh Tabataba'i University, Iran. **E-mail:** [zarghami\\_z@yahoo.com](mailto:zarghami_z@yahoo.com)

## INTRODUCTION

The last decade has witnessed a widespread change in language assessment concepts and methods. One of the main reasons for such a change is the growing interest of practitioners in the concept of assessment for learning, which means considering teaching, learning, and assessment as an integrated and interdependent chain of events (Lee, 2007). There has also been a great focus on learning improvement instead of evaluating students through tests. In many ESL and EFL contexts, the focus of assessment practices is changing from mastery testing of instructional syllabus content to ongoing assessment of learners' achievement during a course of study.

Since the 1990s, new ways of assessment have been introduced to the language testing field. Portfolios, journals, conferences, observations, self- and peer-assessment, and student-generated tests are some of the newly introduced techniques for language assessment which have been named as "alternatives in language assessment" (Brown & Hudson, 1998).

Student-generated testing as a sub-discipline of alternative assessment is regarded as a way of engaging learners in the process of test construction. This method can be a productive, intrinsically motivating, and autonomy building process (Brown, 2004). The main instructional goals of student-generated tests are (a) developing learners' ability to apply principles and generalizations already learned to new problems and situations, (b) developing appropriate study skills, strategies, and habits, (c) learning terms and facts of the subject matter, (d) developing a commitment to accurate work, and (e) developing learners' ability to perform skillfully (Angelo & Cross, 1993).

### **Student-generated tests**

As an alternative form of assessing language learners' abilities, student-generated tests engage learners in the process of test construction. One of the main purposes of administering tests is to encourage students to review a given course content and this is exactly what student-generated tests are expected to do. Most of the time, student-generated tests urge learners to review the materials "almost without awareness on the students' part that they are reviewing the material" (Brown, 2004, p. 276). This technique also helps instructors learn what students consider to be fair and valuable examination questions, shows them what students consider to be the most important parts of the course content, and indicates how well students model answers to the questions they themselves have created (Angelo & Cross, 1993).

Student-generated testing is an assessment technique which is in line with the constructivist approach to teaching and assessment. Based on this theory, the knowledge that learners construct on their own is more valuable than the one which is transmitted to them by the teacher. As Anderson (1998)

states, “In constructivist classrooms, students learn from active participation and having opportunities to explore their own ideas through debate and inquiry” (p. 7). In student-generated tests, this process is facilitated by 1) developing self-assessment and learning management skills in the students, 2) increasing their understanding and ability to think critically about the course content and 3) fostering an attitude in them that values understanding and long-term learning (Haugen, 1999).

Student-generated tests are a sub-category of classroom assessment techniques (CAT). In fact, for considering the specific characteristics of such tests, we need to investigate them in the broad context of CAT's characteristics. Angelo and Cross (1993) describe seven characteristics of CATs as follows. The first characteristic is learner-centeredness. The primary focus of student-generated tests is on the students' needs to learn something. If we want our students to become life-long learners, then they must learn to take the responsibility of their own learning (Kumaravadivelu, 2001). Student-generated testing engages learners in the process of test construction and provides them with opportunities to become autonomous learners. Hammond and Collins (1991, cited in Wongsri, Cantwell, & Archer 2002) state that “If learners do not also develop the capability of directing their own learning and acting on the world around them, they will be only partially educated, and limited in what they can do” (p.1). Engaging learners in regular assessment can be used in productive ways to encourage students to take the responsibility of their own learning. A second characteristic is instructor-directedness. While student-generated testing is a learner-centered approach and students are involved in decision-making and test construction processes, it is the instructor who decides what to assess, how to respond to results, and with whom to share the obtained results (Morris, 2004). A third characteristic is that they are mutually beneficial. Involving language learners in the process of assessing their own knowledge and in decision making would help them better understand what the expected learning standards are. Language learners who deeply understand what they are expected to learn and how they will be assessed are more likely to make learning gains. One more characteristic is formativeness. Student-generated tests are a kind of formative assessment, with a focus on students' learning and their improvement during a course of study. One of the main goals of student-generated tests is to change the purpose of assessment from getting good grades to learning the content of the course (Haugen, 1999). Practitioners of assessment for learning believe that if we use assessment to motivate students to learn, the assessment should assist students to observe their success by helping them believe that success is within reach if they keep trying. Student-generated testing is a practical way to open the assessment process and invites learners in as partners, monitoring their own levels of achievement. Context-specificity is the fifth feature. Student-generated testing is designed to cater to the particular needs and characteristics

of instructors, students, and disciplines we are dealing with. Due to its flexibility, it can be applicable to every specific context and every group of language learners (Haugen, 1999). Another important characteristic is related to their ongoingness. Student-generated testing is obviously an ongoing process which caters to the students' leaning and improvement during a course of study, not just evaluating their ability at the end of a course (Stiggins & Chappuis, 2005). Last but not least, student-generated tests have an orientation toward best practice. By incorporating the best existing practices, student-generated testing is considered to be one of the most systematic, flexible, and regular types of language assessment. Assessment of students prior to instruction helps instructors take the learners' needs into consideration. Assessment during the course of instruction, however, helps instructors ensure that students are learning the content well. Assessment after instruction helps to reinforce materials taught and uncover any misunderstanding of the content before it becomes a significant problem (Enerson, Plank & Johnson, 2007).

### **Assessment for learning**

The term assessment for learning (AFL) has been used interchangeably with formative assessment (Lee, 2007). As Lam (2014) argues, "The formative function of assessment is to improve student learning .... Formative assessment is normally informal, continuous, interactive, small-scale, and classroom-based. Its task types could be any learning tasks such as self- and peer-assessment" (p. 2). The key promise of AFL is integrating teaching, learning, and assessment. In this view, "learning is a goal in its own right, and assessment is a means to achieving the goal" (Lee, 2007, p.1). William (2011) has investigated studies that have been made during a half a century ago to reach an acceptable definition for assessment for learning. The Assessment Reform Group (2002) has defined AFL as the process of finding evidence for use by learners as well as their teachers to decide where learners are in their learning, where they need to go, and how best to get there. Klenowski (2009) has introduced this new trend as "Assessment for learning is part of everyday practice by students, teachers and peers that seeks, reflects upon and responds to information from dialogue, demonstration and observation in ways that enhance ongoing learning" (p.264).

The following 10 principles are defined as the basic concepts underlying assessment for learning approach by The Assessment Reform Group (2002):

1. Being part of effective planning of teaching and learning.
2. Focusing on how students learn.
3. Being recognized as central to classroom practice.
4. Being regarded as a key professional skill for teachers.
5. Being sensitive and constructive because any assessment has an emotional impact.

6. Taking account of the importance of learner motivation.
7. Promoting commitment to learning goals and shared understanding of the criteria by which they are assessed.
8. Providing learners with constructive guidance about how to improve.
9. Developing learners' capacity for self-assessment so that they can become reflective and self-managing.
10. Recognizing the full range of achievement for all learners.

Since its inception, researchers have made a number of studies to investigate the impact of student-generated tests on learning in general and language learning in particular. As early as 1975, Frase and Schwartz investigated the recall consequences of generating questions about prose. In the first study, 48 high school students were supposed to ask questions about the course content, answer questions about it, or merely study the course content in a tutorial situation. In the second study, 64 college students were asked to construct 5 to 10 question regarding the content of the specific course individually. The results of the both experiments showed that engaging both college and senior high school students in question production, individually or in a face-to-face tutorial situation would improve their recall of prose material over merely studying. Pearson (1991) had general biology college students develop and answer their own reading comprehension questions, which had a favorable impact on their midrange (weekly quiz) performance, but was not more beneficial than teacher-made questions on enhancing their final exam performance. In the same vein, Brown (1991) argued that requiring students to develop an eight-question final exam as preparation for the teacher developed take-home final exam gave rise to the best final exam essay he had read for any course he had taught within ten years of college teaching. Likewise, Rash (1997) required college computer science students to construct a verbal problem and its solution in small groups during a whole semester. Although the researcher did not directly measure the learning effect of this experiment, the results obtained through course evaluation forms indicated that students found this assignment worthwhile and expended more efforts in the course than usual. More recently, Brink, Capps, and Sutko (2004), working with a group of engineering students, investigated the relationship between a) comprehensiveness of student-developed tests and final exam scores, b) the quality of student-developed tests and answer keys combined with final exam scores, and c) student-developed test similarity to instructor final exam and final exam scores. The result of this study showed that students' creation of exams and answer keys as a study method seems to be more effective for above average students than for below average students. In addition, the findings revealed that students who prepared a good comprehensive exam and answer key can expect to do better on the final exam than those who do not.

Bobak (2008) applied student-generated test questions as a classroom assessment technique to amend the traditional lecture format beyond the didactic confines so as to reach students with different learning preferences in a chiropractic technique course. The results of this study showed that the majority of the respondents reported that small group work was helpful. Many of the respondents perceived themselves as better prepared for the lecture exam when utilizing the study guide.

Within an English language teaching (ELT) context, Murphey (1995) encouraged students in creating their own word lists, grammatical concepts, and content that they thought were important. The lists from students were synthesized by the researcher and all the test items were based on that list. The results of the study showed that students involved in the process of test creation were intrinsically more motivated and achieved higher scores in the test. Finally, in a more recent ELT-related study, Ashtiani and Babaii (2007) probed the effect of EFL high school students' cooperative test construction on their grammatical knowledge as well as their attitudes toward such tests for a period of sixteen weeks. The students of the experimental group were asked to construct test questions. After discussing their constructed items in pairs and then in groups of four, checking and correcting these questions, the teacher read the students' constructed tests and modified them slightly. At the end of the term the students in the experimental group were asked to take part in the process of constructing the final test cooperatively. The findings revealed that the grammatical knowledge of the students involved in the experiment was significantly better than that of the students in the control group. Furthermore, administering a Likert-type scale questionnaire to investigate the students' views regarding this type of assessment revealed that the students' attitude to cooperative test construction was positive.

Considering the impact of student-generated tests on the learning of students in different fields of education and second language learning contexts, the effects of this assessment method on EFL learners is not so clear. In Iran, like many other EFL contexts, traditional methods of language testing are dominating language teaching and assessment fields. From elementary schools to universities, language learners are supposed to take different language tests at the end of the course of instruction to demonstrate their abilities and to be evaluated based on their test scores. In recent years, however, some innovative language teachers have tried to apply new ways of assessment in their language classes, but such practices are in their infancy. Moreover, it seems that the concept of assessment for learning has not gained its deserved reputation among Iranian researchers. There is little or no evidence of investigating the effects of alternative assessment methods, as assessment for learning tools, on students learning and their impact on learners' attitudes toward language learning in the context of foreign language learning in Iran. Given the lack of sufficient studies which explore the effect

of student-generated tests on language learning, particularly at university level in EFL settings, there is clearly a need for further research in this area. Therefore, the present study aims to investigate the effect of student-generated tests on grammar learning of Iranian EFL students by answering the following research questions:

1. Do student-generated tests have any impact on Iranian EFL students' mastery of their grammar course content?
2. Do student-generated tests have any impact on Iranian EFL students' performance on a standardized grammar test?

## METHODOLOGY

### Participants

The participants in this study were 68 Iranian intermediate EFL students (38 females and 30 males) majoring in English Language and Literature in one of the branches of Islamic Azad University in Iran in the first semester of the academic 2012-2013. They were all freshmen with an average age of 22. The participants, taking a course named *Grammar and Writing I*, were assigned to two classes by the specific rules and conventions of the respective university, but these two classes were randomly assigned to experimental and control conditions (each group containing 34 students).

### Instruments

The main instruments used in this study were two grammar tests: a standardized grammar test developed by the researchers and another grammar test based on student-generated items.

a. **The standardized grammar test.** This was a 40-item multiple-choice grammar test administered to both groups in the first session of the treatment period. The questions were based on the topics students were supposed to study during *Grammar and Writing I*, namely (a) verb tenses (including simple present, present continuous, present perfect, simple past, past continuous, past perfect, and simple future), (b) active and passive voice, (c) coordinators (including coordinating conjunctions and conjunctive adverbs), and (d) subordinators (including relative pronouns and adverbial subordinating conjunctions). There were ten items for each topic, most of which were adapted from previous actual samples of the TOEFL officially released by ETS. Given that the items were selected and adapted from various sources, there was a need to check the reliability as well as the content validity of the whole test. The reliability of the test, measured through Kuder-Richardson 21 formula, was 0.91 and its content in terms of item relatedness and content coverage was approved by the course instructor. This test served three purposes: it was used as the pretest as well as one of the posttests of the study. Moreover, it functioned as an instrument to determine the homogeneity of the

two groups at the beginning of the study in terms of their grammatical knowledge.

b. **The student-generated grammar test.** This was a 40-item multiple-choice grammar test administered at the end of the treatment period. The content of this test was similar to the standardized grammar test. This means that there were ten items for each of the four grammatical topics covered during the given grammar course; however, the items of this second test were developed by the participants of the experimental group (A number of sample items from this test appear in Appendix A).

## **Procedure**

The instructional treatment was part of a university English course offered in one of the branches of the Islamic Azad University in Iran. An experienced male instructor taught both groups of the participants who took part in two 2hr sessions per week (10 sessions totally). In the first session of the treatment, the standardized grammar test was administered to both groups by the respective teacher. After the instructional period, the participants took a 40-item multiple-choice test. The specific procedure for each of the two groups of this study during their 10-session instructional treatment was as follows.

### **A. Experimental group**

After teaching each of the four topics mentioned earlier (verb tenses, active/passive, coordination, and subordination) the instructor asked each of the 34 students of this group to develop 10 related grammar items (5 fill in the blank and 5 multiple-choice). To do this, the students were asked to select 10 grammatical words/ terms as a core to construct their test items. Then, the participants were supposed to find appropriate sentences for each word to form the stem (the initial part of multiple-choice questions) of each item. After checking the sentences to be both grammatically and semantically correct by the instructor, the students were asked to delete the target word from the sentence. In the case of multiple-choice items, the students were asked to put the deleted word under the sentence as an option. Then, the students were supposed to write three similar words as distracters. The instructor checked all the items and provided each student with appropriate corrective written feedback. The participants were asked to follow a similar procedure for each of the other three grammatical topics selected for the purpose of this study and submitted their constructed items to the instructor to receive appropriate teacher feedback on their work. Overall, each student constructed 40 test items. The students were advised to read units 20-40 from *Practice and Progress* by L. G. Alexander to select and adapt the stems of their items from. In order to ensure that they would do their best in doing this assignment, they were told that 15% of their final exam score would be allocated to the task of item construction. Ten selected items related to each of the four given



grammatical topics were written on the board and discussed with the students once each of the grammatical topics was about to finish. Thus, at the end of the treatment period there were 40 selected items, from among all the items generated by the students that were written on the board and were analyzed and modified slightly with active participation on the part of the students. These 40 items were used to form the student-generated grammar test, serving as the second instrument of this study.

## **B. Control Group**

In the control group the routine syllabus based on the presentation, practice, and production model was followed without any resort to alternative assessment techniques. In this group the students were not involved in class procedures, and they were not required to develop any grammar items. The students were most of the time passive except the time they were doing different grammatical exercises.

To give the students in the control group more or less the same learning opportunities, they were also asked to read the same units from *Practice and Progress* and answer the comprehension questions related to each passage. Moreover, the same forty selected items generated in the experimental group were written on the board and discussed with the students of the control group too, once each grammatical topic was finished. The rationale for providing the participants in the control group with the specific items constructed by their peers in the experimental group was to make them acquainted with those items and put these in a similar situation to the participants of the experimental group.

At the end of the treatment period, which lasted for about two months, students in both groups were given two posttests. The first one was the standardized grammar test they had already taken as the pretest and the second was the student-generated test. Both tests were given to the learners two weeks after the last session of the treatment period. This is the reason why the same pretest was used as the posttest inasmuch as there was a 10-week interval between the two administrations, long enough for the participants not to remember the items from the first administration.

## **Data Analysis**

The researchers analyzed the results of the participants' scores on the pre and posttests by means of 3 independent *t*-tests. The first independent *t*-test was run to compare the pretest scores of both groups on the standardized grammar test. The second independent *t*-test was run to compare the scores of the experimental and control groups on the first posttest, i.e. the student-generated grammar test. And, at last, the third independent *t*-test was run to compare the second posttest scores of the experimental and control groups on

the standardized grammar test. All statistical analyses were carried out using Statistical Package for Social Sciences (SPSS) with alpha set at .05.

## Results

Table 1 displays the descriptive statistics related to the pretest. The results indicate that there is not a statistically significant difference between the mean scores of both groups  $t(66) = .28, p = .59$ . This suggests that the two groups were homogeneous in their grammatical knowledge at the start of the treatment period.

**Table 1. Descriptive statistics for the pretest scores of the experimental and comparison groups.**

Groups	n	M	SD
Experimental	34	16.74	1.310
Control	34	16.65	1.203

In the posttests session, three students from the experimental group, and two students from the comparison group were absent. Therefore, the data analysis for the posttests is carried out with 63 students.

To answer the first research question, regarding the impact of student-generated tests on grammar learning of Iranian EFL students an independent  $t$ -test was run to compare the pretest scores and first posttest scores of the experimental and comparison groups. Table 2 shows the related descriptive statistics, which suggests that there is a significant difference between the mean scores of both groups  $t(61) = 6.83, p = .001$ , Cohen's  $d = .433$ . This suggests that students in the experimental group had a much better performance on the student-generated test than their peers in the control group. Therefore, the first research question was answered in the positive.

**Table 2. Descriptive statistics for the first posttest (student-generated test)**

Groups	n	M	SD
Experimental	31	17.16	2.77
Control	32	12.25	2.92

In order to answer the second research question, another independent  $t$ -test was used to compare the results of the second posttest, i.e. the standardized grammar test, which had acted as the pretest, too.

**Table 3. Descriptive statistics for the second posttest (standardized grammar test)**

Groups	n	M	SD
Experimental	31	33.39	2.604
Comparison	32	30.47	2.155

Table 3 shows the mean scores of both groups on the second posttest. Unlike the result of the pretest, which showed no significant difference between the two groups, this time students in the experimental group significantly outperformed their peers in the control group  $t(61)= 4.85, p=.001$ , Cohen's  $d=.278$ . Thus, the answer to the second research question is another positive.

### Discussion

This study was an attempt to explore the impact of student-generated tests on Iranian EFL learners' grammar learning. The results indicated that involving students in the process of test construction through generation of grammar items significantly improves their grammar learning. This was evidenced by the outperformance of the students in the experimental group compared to their peers in the control group on two tests: a student-generated one and a standardized one. This suggests that using student-generated tests as a way of encouraging learners to focus more on learning rather than on assessment seems to improve grammar learning of foreign language learners.

The results of this study indicate that student-generated tests can improve foreign language learners' mastery of what is taught in grammar classes and enhance their performance on standardized tests of grammar addressing the target content.

The findings of the present study corroborate the findings of studies conducted by Ashtiani and Babaii (2007), Brink et al. (2004), and Murphey (1995) in that involving learners and their judgments in constructing test items would have a positive effect on their learning. While the above studies show that cooperative test construction in an EFL context, encouraging learners to generate their own list of words, grammatical concepts, and content that they consider as the most important part of their course content, and involving engineering students in the process of generating their own final exam and answer sheets in a non language learning context resulted in students' improved learning, the results of the present study pointed to the significant effect of student-generated tests on the grammar learning of university students in an EFL context. The findings of this study suggest that continuous grammar item construction as an ongoing process of assessing learners' ability can change their nature from being an assessment tool to a learning tool.

Student-generated tests can affect grammar learning of students in a number of ways.

First and foremost, involving language learners in the process of test construction and decision making process can help them better master the subject matter content and understand what the expected learning standards are. Language learners who deeply realize what they are expected to learn and how they will be assessed are more likely to make learning gains.

Second, such frequent assessments provide the continuous feedback that teachers need to determine what should come next in the grammar teaching process and that language learners need to learn how to do better the next time. As Stiggins (2004) states “When everyone is trying to learn, feedback about the effort has three elements: redefinition of the desired goal, evidence about present position, and some understanding of a way to close the gap between the two” (p. 8). In this study, the instructor provided learners with descriptive feedback, namely not merely grades or scores, but a kind of focused guidance which is specific to each learner and also specific to each grammar target. This kind of feedback can lead learners to learn to self-assess and set their own goals.

As opposed to traditional, end-of-the-term summative assessments, which were incapable of providing both the teacher and the learners with systematic information on students' progress and achievements, student-generated tests can translate classroom assessment results into day-to-day feedback for students, and provide them with helpful insights as to how to improve. Such formative assessment can help students learn how to study the course content and guide them in a self-analysis of their own processes.

Third, engaging language learners in regular assessments can be used in productive ways to encourage them to be responsible for their own learning. Such ongoing formative assessment enables students to watch themselves grow over time and be in charge of their own learning.

Fourth, one of the most valuable aspects of this form of assessment is its consciousness-raising role. Engaging students in generating grammar test items may lead to increased learner awareness of special linguistic forms. By creating grammar test items, which involves some kind of grammatical consciousness-raising, learners can better realize a particular grammar feature, how it works, and what it consists of. Such activities can help language learners construct their own learning.

Finally, the fact that the participants in the experimental group had been faced with some extra challenge through constructing grammar items might have enhanced their motivation for grammar learning. According to Harmer (1987), encouraging students to discover grammar rules for themselves is a valuable way of helping them to get to grips with the target language. Thus, involving learners in the process of test construction may be a highly motivating assignment, extremely beneficial to students' understanding of English grammar rules. Practitioners of assessment for learning assert that for learning to be effective, knowledge and understanding of what is to be

achieved is not sufficient. Students must want to make the effort and be willing to keep on engaging, even when they find the learning task difficult. Assessment that encourages learning promotes motivation by emphasizing progress and achievement rather than failure (Stiggins, 2004).

In this study, the participants of the experimental group were in a position to decide whether success is within or beyond reach, whether learning is worth the required effort, and so whether to try or not. Involvement in these decision making processes can improve learners' motivation to learn and may affect their perceptions of their own capabilities as was reflected in assessment results.

## **CONCLUSION**

The findings of this study suggest that having EFL learners create test items while learning grammar in language classes is likely to enhance their achievement and improves their performance on standardized grammar tests. Based on the results of the present study, it seems that involving learners in generating their own test items encourages them to study more deeply and focuses their attention more on learning rather than just achieving good grades. Providing learners with continuous feedback can help them easily bridge the gap between their current level of knowledge and their target goals. By making assessment an ongoing, motivating, and a learner-centered process, assessment can be changed into a useful learning tool. Involving learners in constructing grammar test items may provide them with good opportunities to discover grammar rules for themselves. Recent learning theories suggest that the things we discover for ourselves are more firmly fixed in our minds than those we are simply told. It also seems that students' item construction, as a consciousness-raising process, may lead to increased grammar awareness of the language learners and have positive effects on their learning. Focusing on the learning dimension of assessment and trying to integrate teaching, learning, and assessment within a grammar learning setting is likely to improve students' learning.

The findings of the present study could be applied for other modern foreign languages as well as other language skills and components. It might be an exciting experience for teachers of other modern foreign languages to let their learners have a hand in developing test items of various sorts when it comes to other dimensions of language learning, say vocabulary or reading comprehension. These experiments are likely to provide learners with unique opportunities for further language learning and give them a deeper insight into how languages work.

## REFERENCES

- Angelo, T. A., & Cross, P. A. (1993). *Classroom assessment techniques: A handbook for college teachers*. San Francisco: Jossey-Bass.
- Ashtiani, N. S., & Babaii, E. (2007). Cooperative test construction: The last temptation Educational reform? *Studies in Educational Evaluation*, 33, 213,228.
- Assessment Reform Group. (2002). *Assessment for Learning: 10 Principles*. Port Melbourne: Cambridge University Press.
- Brink, J., Capps, E., & Sutko, A. (2004). Student exam creation as a learning tool. *College Student Journal*, 38, 262-272.
- Bobak, K. A. (2008). Use of student-generated test questions as a classroom assessment technique within the chiropractic technique classroom. *Journal of Chiropractic Education*, 22, 49-50.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practice*. White Plains, NY: Pearson Education.
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32, 653- 675.
- Brown, L. W. (1991). To learn is to teach is to create the final exam. *College Teaching*, 39, 150-153.
- Enerson, D. M., Plank, K. M., & Johnson, R. N. (2007). An introduction to Classroom Techniques. Retrieved 20, 3, 2010, from [http:// www. Schreyerinstitute.edu/](http://www.Schreyerinstitute.edu/).
- Frase, L.T., & Schwartz, B.J. (1975). Effect of question production and answering on Prose recall. *Journal of Educational Psychology*, 67, 628-635.
- Harmer, J. (1987). *Teaching and learning grammar*. London: Longman.
- Haugen, L. (1999). Classroom assessment techniques. Retrieved 3, 9, 2009, from [http:// www.Celt.iastste.edu/teaching/cat.html](http://www.Celt.iastste.edu/teaching/cat.html).
- Klenowski, V. (2009). Editorial: Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy, and Practice*, 16, 263–268.
- Kumaravadivelu, B. (2001). Toward a postmethod pedagogy. *TESOL Quarterly*, 35, 537-560.
- Lam, R. (2014). Can student-generated test materials support learning?. *Studies in Educational Evaluation* (in press). Retrieved 30, 5, 2014 from <http://www.sciencedirect.com/science/article/pii/S0191491X1400008X>.
- Lee, I. (2007). Assessment for learning: Integrating assessment, teaching, and learning in the ESL/EFL writing classroom. *The Canadian Modern Language Review*, 64, 199-214.
- Leung, C., & Mohen, B. (2004). Teacher Formative Assessment and Talk in Classroom Contexts: Assessment as Discourse and Assessment of Discourse. *Language Testing*, 21, 335-359.

- Morris, T. (2004). CAT characteristics. Retrieved 23, 9, 2009, from <http://IAR Assess Teaching.html>.
- Murphey, T. (1995). Tests: Learning Through Negotiated Interaction. *TESOL Journal*, 4, 12-16.
- Pearson, J.A. (1991). Testing the ecological validity of teacher-provided versus student-generated postquestions in reading college science text. *Journal of Research in Science Teaching*, 28, 485-504.
- Rash, A. M. (1997). An alternative method of assessment: using student-created problems. *PRIMUS*, 7, 89-95.
- Stiggins, R. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan*, 86, 22- 27.
- Stiggins, R., & Chappuis, J. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory and Practice*, 44, 11- 18.
- William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*. 37, 3-14.
- Wongsri, N., Cantwell, R. H., & Archer, J. (2002). The validation of measures of self-efficacy, motivation, and self-regulated learning among Thai tertiary students. Paper presented at the Annual Conference of the Australian Association for Research in Education, Brisbane. Retrieved November, 2009, from: <http://www.aare.edu.au/02pup/won02088.htm>.

## **Appendix A**

### **Samples from the student-generated items**

1. The house ..... five months ago.  
A. completed      B. is completed      C. was completed      D. was completing
2. The man ..... you met yesterday is an actor.  
A. whom      B. what      C. which      D. whose
3. He usually ..... at 6 o'clock, but today he's working late.  
A. knock off      B. knocked off      C. had knocked off      D. knocks off
4. I ran to the station, ..... missed the train.  
A. but      B. and      C. so      D. therefore
5. While he ..... the letter, he heard a knock at the door.  
A. read      B. reading      C. was reading      D. is reading
6. Every time he ..... to our house he eats as much food as he can.  
A. came      B. comes      C. has came      D. is coming
7. Most animals normally ..... on other animals or plants.  
A. feed      B. fed      C. feeds      D. feeding
8. If I had studied harder, I ..... have passed all my exams.  
A. can      B. would      C. may      D. will