

The Effect of Immediate Grammar Tests on the Improvement of Iranian Pre - university Students' Final Exam Achievement

Mahbanoo Alimoradi¹
Keivan Mahmoodi², Ph.D
Payman Rajabi³, Ph.D

^{1,2,3} Malayer Branch, Islamic Azad University, Malayer, Iran

Abstract: *The present study aims at investigating the effect of immediate grammar tests on the improvement of Iranian pre university students' final exam achievement. 60 pre university students studying at Shahed high school in Malayer, Iran were chosen for the study. They were assigned to one control and one experimental group with 31 students in the experimental group and 29 students in the control group. Four grammatical points were taught during three months to both groups with the same method predetermined at schools except for the fact that subjects in the experimental group were also given immediate grammar tests and were asked to answer them immediately after teaching. The instrument used in this study was multiple choice tests (as pre, immediate and post tests). Paired and independent sample t- tests were used to answer the research question. Results indicated a significant improvement in the scores of the subjects in the experimental group. In the end, some pedagogical implications were given.*

Keywords: *Immediate Tests, Grammar Tests, Pre- university students, Final Exam*

Background

Language testing as a methodology for investigating language ability comes from a long and honorable tradition of practical teaching and learning needs. Being central to language teaching, it provides goals for language teaching and it monitors for both teachers' and learners' success in reaching these goals. Language testing also provides a methodology for experiment and investigating both language teaching and language learning/acquisition.

Perhaps the most common use of language tests is to pinpoint strengths and weaknesses in the learned abilities of the students. Another important use of language tests is the decision of who should be allowed to participate in a particular program of instruction (Henning 1987). Based on Farhady et al. (1996), tests are applied to make decisions about people's lives. Therefore, fair decisions will be impossible if tests do not provide accurate information. On the other hand, specific samples of behavior can be obtained by tests which distinguish it from other types of measurement (Mousavi, 1999). Overall, any technique and procedures to assess and measure a factor or some ability is called a test.

In many circumstances, tests are given infrequently and are generally perceived as a bother by faculty and students alike. We shouldn't neglect the importance of testing. To state an

obvious point, if students know they will be tested regularly, they will study more and will space their studying throughout the semester rather than concentrating it just before exams (Banger-Downs, Kulik, 1991; leeming, 2002).

However, more important for present purposes, testing has a powerful position effect on future retention. If students are tested on material and successfully recall or organize it, they will remember it better in the future than if they had not been tested this phenomenon, called the testing effect, has been studied over a long period of time, (e.g., Gates, 1917), but is not well known outside cognition psychology. The importance of testing in education makes it an important topic of continuing research. As technology education evolves to emphasize more cognitive learning, the time devoted to testing and the effects of testing will become increasingly important. Most of the research on testing which has been reported in recent years has concerned standardized tests (Bridgeford, Conklin, and Stiggins, 1986). Most of the evaluation done in schools, however, is done with teacher – made tests (Haynie, 1983, 1991, 1992; Mehren, 1987; Mehren & Lehmann, 1987; Newman & Stallings, 1982). The available findings on the quality of teacher – made tests cast some doubt on the ability of teachers to perform evaluation effectively (Burdin, 1982; Carter, 1984, Fleming and Chambers, 1983; Gullickson & Ellwein, 1985; Haynie, 1992; Stiggins & Bridgeford, 1985; wiggins, 1993).

Despite these problems, Mehrens and Lehmann (1987) point out the importance of teacher – made tests in the classroom to evaluate attainment of specific instructional objectives. Evaluation by teacher –made tests in schools is an important part of the educational system and a crucial area for research (Haynie, 1990a, 1990b, 1991, 1992; Mehrens & Lehmann, 1987; Wiggins, 1993). Most previous research has used tests involving recognition (like multiple – choice test) or cued recall (like short answer tests).

Methods for improving long term learning, including the well – established use of testing, should be examined for various ages of learners to properly assess their usefulness. Testing effects may not generalize to learners beyond traditional academic settings. People not tested regularly in school may react adversely to the use of tests as learning events, as they are likely to be unaccustomed to taking tests, may be more anxious taking tests, or may have difficulty accessing relevant knowledge when tests are introduced apart from initial learning. If they underperform on these tests, they may benefit less from them due to a lack of processing that occurs with successful retrieval (Carpenter & Delosh, 2006).

The use of testing as an educational aid in these learners, however, can still be valuable, so it is useful to assess whether learners in pre university setting can benefit as much or at all from testing as a learning technique. If pre university students are tested on materials and successfully recall or recognize it, they will remember it better in the future not only for the final exam but also for university entrance exam. So students in our experiment are taught a grammar point of a special lesson of pre university book. Then they immediately take a test on the material

before taking a final exam. We predict that taking tests immediately after teaching will promote superior retention on final exam. This outcome may indicate that testing has positive effects on long term retention.

Purpose of the Study

The purpose of this research is to investigate the effect of immediate grammar test (Immediate testing “refers to the commonly employed evaluation by testing which occurs at the time of instruction or immediately thereafter) on final exam. (Dwyer, 1968, Dwyer, 1973; Duchastel, 1981; Nungester & Duchastel, 1982; Haynie, 1990a,

1990b, 1991, inpress). “Immediate testing “refers to the commonly employed evaluation by testing which occurs at the time of instruction or immediately thereafter.

The multiple – choice test is a form of assessment in which respondents are asked to select the best possible answer out of the choices from a list. Multiple choice testing is an efficient way to assess a wide range of knowledge, skills, attitudes and abilities (Haladyna, 1999). When done well, it allows broad and even deep coverage of content in a relatively efficient way. Though often maligned, and though it is true that no single format should be used exclusively for assessment (American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education, 1999), multiple choice testing still remains one of the most commonly used assessment formats (Haladyna, 1999; McDougall, 1997).

The multiple-choice test is a very flexible assessment format that can be used to measure knowledge, skills, abilities, values, thinking skills, etc. Such a test usually consists of a number of items that pose a question to which students must select an answer from among a number of choices.

The present study attempted to answer the question raised about the effect of immediate grammar test on the final exam in pre university students. The objective of the study could be expressed in the following question:

1. Do immediate grammar tests improve the learners’ grammatical knowledge in the final exams?

Methodology

Population and Sample

Sixty female students took part in this study. The age range of students was 17 to 18. They were studying in academic year 2013-2014 in the fourth grade of Shahed high school in Malayer. All participants are familiar with English language. They received their formal English

language education two days a week, 90 minute a day at high school. None of the participants had been in another English speaking country yet.

The participants' mother tongue was Persian. The English language was considered to be the students' foreign language. This school was chosen because the researcher was the teacher of this high school. It was expected that such choice would enable the procedures for doing this research.

At Shahed high school, fourth grade students were divided into three fields: science, math and literature. Two of them, science and math classes were selected to be experimental group and control group. The reason for this choice was that science students are usually more motivated than math students. Both groups consisted of a total number of 60 students: 31 experimental and 29 control. Additionally, they all had the same exposure to English through formal classes in high school and secondary school. Similarly, since they came from the same country, it was reasonable to assume that they shared a homogeneous EFL background. They also matched each other in grade (fourth), and school (Shahed pre- university). Moreover, they came mostly from the same neighborhood and were the same gender and age.

Design

The research employed quantitative data. The quantitative data is obtained with the help of the pre-test and post-test results. The study integrated two variables: independent and dependent. The independent variable of the study was the use of immediate grammar test. The dependent variable was the improvement of pre university students' final exam achievement measured by pre- and post-test scores, obtained from the differences between the pre-test and immediate test and final test scores. The independent variable is nominal; the dependent variable was numeric. The results of the pre- and post-tests and final tests were analyzed and tabulated with the help of the SPSS program to answer the research questions.

Instrumentation

The collection of the data was accomplished through the following instruments:

1. A pre- test of grammar points.
2. An immediate test of grammar points.
3. A Post test of grammar points.

The grammar tests targeted the following sub-skills:

(1)Conjunctions of time , reason and condition: when , as , because , since , whetheror ; (2) Verb + Object + Bare infinitive Expressing manner : By + Gerund (3) Reduced Adjective Clauses and (4) Modification of Adjectives : so / suchthat enough /too.

Fortunately, these fourth group of tests were accomplished in the pre-test and immediate test and could be done for the final -test for administrative reasons .

Data Collection

As there was a homogeneous group of participants which were all at the same level of proficiency, so the researcher didn't have any problem in this field. The participants were divided into two groups, one of them is the experimental group given an immediate grammar test after teaching and the other group, control group, wasn't given any immediate test after teaching. The participants were informed of the study from the beginning. The goals and procedures of the study were presented to them. Two groups were involved in the experiment. Both groups used the same textbook - English for pre – university students and had English classes for the same amount of time. In both groups the same syllabus was used. The teacher of both groups was the researcher herself. The experiment lasted for one term (12 weeks); it started on the first day of Mehr and finished on the last day of Azar. At the beginning of the term both groups had a pre-test aiming to test their grammar proficiency level, and at the end of the term both groups were given a final -test aiming to test their grammar achievement. The latter intended to show whether the use of immediate assessment had had any impact on the learners' grammar enhancement.

As mentioned above, the experimental group received the treatment – the application of the ongoing four immediate tests, particularly in grammar learning. Thus, apart from their course materials, the participants in the experimental group were given immediate tests (multiple choice tests). Throughout the instructional process the teacher/ researcher monitored students' progress and provided feedback on their strengths and weaknesses. Feedback was the key element in this kind of assessment: this feedback allowed students to correct conceptual errors and encourages instructors to modify instructional activities in light of their effectiveness. The comparison group was taught the same materials with traditional book exercises and activities. These types of activities are not provided with feedback.

Data Analysis

The statistical tests and procedures used in this study for data analysis described here. Paired t-test was applied in order to analyze the differences between pre-test performance and post-test performance in each group. By computing mean, SD and SEM, the researcher compared the control and experimental groups post tests. The researcher use SPSS16 in order to analyze the data. Then a pre-test was given to the subjects in both groups before any treatment. Later grammatical points were taught to both groups and immediate grammar tests were given only to the experimental group. Finally, a standard post-test was given to the subjects in both groups to confirm or reject the Null hypothesis.

Findings

The means, standard deviation of the two groups in the pre test and post test are presented in table 1.

Table 1: Descriptive Statistics for Pre- Test and Post- Test of two groups

Group	Pre Test		Post Test	
	M	S.D	M	SD
Experimental	14.1129	2.20117	18.1129	2.20117
Control	14.0000	2.20389	15.6379	2.40497

The difference between the two means in pre test is very little. So it can be concluded that the two groups are homogeneous in terms of their grammar ability. The difference between the two means of performance in post test is large enough to show that the difference is actually related to the immediate grammar tests.

However, more statistical computations were done to show the differences between these two mean scores was not statistically significant.

Table 2: Independent t- test to Show the Difference between Two Groups in Pre test and post test

Test	T-Observed	T-Critical	D.F	Sig.(2 – tailed)
Pre Test	.198	1.676	58	.843
Post Test	4.162	1.676	58	0.001

By looking at the above table, one can find that the value of t- observed in pre test is .198 at 58 degrees of freedom which is lower than the value of t-critical at 0.05 level of significance. Meanwhile the two –tailed significance level shows the amount of .843 which is larger than 0.05. So, there is no statistically significant difference between the two groups in pre- test.

As table 2 shows, the value of t-observed in post test is 4.162. This amount of t at 58 degrees of freedom is much greater than the value of t- critical at 0.05 level of significance.

Meanwhile the two-tailed significance level shows the amount of .001 which is lower than 0.05. So, there is statistically significant difference between the performance of two groups in post –test. Therefore, it can be concluded that the difference is actually related to the immediate grammar tests.

The scores obtained by the students in experimental and control group were computed to compare the pre-test with the post-test results. Next, the scores were calculated to find the difference between the two mean scores.

The researcher used paired sample t –test to see whether there is any significant difference between the pre- test performance and post – test performance of the experimental and control group.

TABLE 3: Paired Sample Statistic to compare the pre-test with the post-test results in experimental and control group.

Group	Experimental			Control		
	M	N	SD	M	N	SD
Group before	14.1129	31	2.20117	14.0000	29	2.20389
Group after	18.1129	31	2.20117	15.6379	29	2.40497

TABLE 4: Paired t- test to show the difference between the pre-test and post-test performance of the experimental group and control group.

Groups	T- observed	T- critical	D.F	Sig. (two tailed)
Experimental	-24.644	1.697	30	0.001
Control	-8.679	1.701	28	0.001

As table 4 shows, the very low amount of t (-24.644 shows that not only the mean scores of experimental group are not equal in pre – test and post-test but also the difference between these two means of performance is large enough to show that the difference is actually related to the immediate grammar tests treatment.

Meanwhile the two – tailed significance level shows the amount of .001 which is lower than 0.05. So there is statistically significant difference between pre-test and post test performance in experimental group.

The low amount of t (-8.675) shows that not only the mean score of control group are not equal in pre –test and post - test but also the amount of the mean in the post – test is large enough.

Meanwhile the two -tailed significance level shows the amount of .001 which is lower than 0.05. This shows that there is statistically significant difference between pre –test performance and post-test performance in control group which is related to traditional exercises and activities not immediate grammar tests.

The results of the participants' performance indicated that the scores of the participants in both groups improved in the post test due to the effectiveness of the treatment, but the improvement of the control group was not statistically significant.

In sum, as the results of this study showed, using immediate grammar tests can advance the grammatical knowledge of pre university students in final exam.

Discussion

In education, tests are considered devices of assessment. Students take tests in class to assess what they have learned. The assessment of students' learning in the classroom (both by teachers and by students themselves) is an integral component of the teaching-learning process. Much of this kind of assessment is subjective, informal, immediate, on-going, and intuitive, as it interacts with learning as it occurs, monitoring student behavior, scholastic performance, and responsiveness to instruction. Its role is to determine students' current level of knowledge, skill, or understanding, to diagnose problems they may be encountering, to make decisions about the next instructional steps to take (to revise or to move on), and to evaluate the learning that has taken place in a lesson. As teachers gather information/data about student learning, several categories may be included.

Obtaining the needed data in this study, the researcher ran the necessary statistical procedures. It was proved that the experimental group given immediate grammar tests outperformed the control group on post tests. Therefore, the null hypothesis, the immediate grammar test has no effect on the improvement of pre- university students 'final exam achievement was rejected. Within the constraints of this study, immediate grammar test did promote retention learning.

The main finding of the analysis indicated a positive answer to the question of the study. It was found that the immediate grammar tests had a positive effect on the performance of Iranian pre-university student's final exam. This was proved through the higher mean scores that the experimental group obtained in the post-test. Specifically, the experimental group's performance was more differentiated than that of the control group in the post-test. Furthermore, the pre-test results for both groups did not reveal any statistically significant difference between the two groups. This means that before the application of the experiment they both had nearly similar performance. That is to say, they had the same background knowledge.

Recommendations

This study can be replicated to find out whether the same result is taken. The research covered a limit number of students because this research was covered with the researchers'

students in ordinary classes. It was impossible for the researcher in a large area, conducting a similar research with more participants is suggested. Interpretations of the findings of this study also leads to several recommendations for further research: (a) The same research can be run with students of the other levels such as guidance school and university to find out whether the same results will be taken. (b) The other researchers can use more than one experimental group in order to investigate the different result and see the effects of each strategy separately. (c) The researcher in this study has the chance of using just the immediate grammar tests, so conducting a similar research in the other field such as vocabulary and pronunciation is also suggested. (d) This study was concerned with the female subjects. Conducting a similar research with male participants is also suggested.

References

- Bangert - Drowns, R.L., Kulik, J.A. & Kulik, C.L.C. (1990). Effects of frequent classroom testing. *Journal of Educational Research*, 85.89-99.
- Burdin, J. L. (1982). Teacher Certification. In H. E. Mitzel (Ed.), *Encyclopedia of education research (5th ed.)* New York: Free Press.
- Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: *Support for the elaborative retrieval explanation of the testing effect. Memory & Cognition*, 34,
- Carter, K. (1984). Do teachers understand the principles for writing tests? *Journal of Teacher Education*, 35(6), 57-60.
- Duchastel, P. C. (1981). Illustrations in text: a retentional role. *Programmed Learning and Educational Technology*, 18 (1), 11-15. *about assessment?* NCREL, Oak
- Dwyer, F.M. (1968a). When visuals are not the message. *Educ.Brct. Rev.*, 2, 38-43
- Dwyer, C. A. (1968z). Achievement Testing. In H. E. Mitzel (Ed.), *Encyclopedia of Educational Research (4th ed., Vol. 1, pp. 13-22)*. New York: The Free Press.
- Fleming, M., & Chambers, B.(1983). Teacher-made tests: *Windows on the classroom. In W. E. Hathaway (Ed.), Testing in the schools: New directions for testing and measurement, No. 19 (pp.29-38)*. San Francisco: Jossey-Bass
- Farhady, H., & Keramati, M. N.(1996). A text-driven method for the deletion procedures in cloze passages. *Language Testing*, 13(2), 191-207.
- Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, 6. No. 40.
- Gibb, B.G. (1964). Test- wiseness as a secondary cue response (University Microfilms Document No. 64 -7643). *Unpublished doctoral dissertation*. Stanford University.
- Gullickson, A. R., & Ellwein, M. C. (1985). Post-hoc analysis of teacher-made tests: The goodness of fit between prescription and practice *Educational Measurement: Issues and Practice*, Spring, 15-18.
- Haynie, W. J. (1990a). Effects of tests and anticipation of tests on learning via

- videotaped materials. *Journal of Industrial Teacher Education*, 27(4), 18-30.
- Haynie, W. J. (1990b). Anticipation of tests and open space laboratories as learning variables in technology education. In J. M. Smink (Ed.), *Proceedings of the 1990 North Carolina Council on Technology Teacher Education Winter Conference*. Camp Caraway, NC: NCCCTTE.
- Haynie, W. J. (1991). Effects of take-home and in-class tests on delayed retention learning acquired via individualized, self-paced instructional texts. *Journal of Industrial Teacher Education*, 28(4), 52-63.
- Haynie, W. J. (1992). Post hoc analysis of test items written by technology education teachers. *Journal of Technology Education*, 4(1), 27-40.
- Haynie, W. J. (1983). Student evaluation: The teacher's most difficult job. Monograph Series of the Virginia Industrial Arts Teacher Education Council Monograph Number 11.
- Haynie, W. J. (1994). Effects of multiple-choice and short answer tests on delayed retention learning. *Journal of Technology Education*, 6(1), 32-44.
- Haynie, W. J. (1995a). In-class tests and posttest reviews: Effects on delayed-retention learning. *North Carolina Journal of Teacher Education*, 8(1), 78-93. *Journal of Technology Education* Vol. 9 No. 1, Fall 1997
- Haynie, W. J. (1995b). An analysis of tests developed by technology teachers. *Unpublished manuscript*.
- Haladyna, T. M. (1992a). Context-dependent item sets. *Educational Measurement: Issues and Practice*, 11 (1), 21-25.
- Haladyna, T. M. (1992b). The effectiveness of several multiple choice formats. *Applied Measurement in Education*, 5, 73 - 88.
- Haladyna, T. M. (1999). Developing and validating multiple-choice test items. Mahwah, NJ: *Lawrence Erlbaum Associates, Inc.*
- Henning, G. (1987). A guide to language testing: *Development, evaluation, research* Rowley, MA: *Newbury House*.
- Leeming, F. C. (2002). The exam-a-day procedure improves performance in psychology Classes. *Teaching of Psychology*, 29, 210 - 212.
- McDougall, D. (1997). College faculty's use of objective tests: *State-of-the-practice versus state-of-the-art*. *Journal of Research and Development in Education*, 30, 183-193.
- Osterlind, S. J. (1989). Constructing test item.
- Mehrens, W. A. (1987). "Educational Tests: Blessing or Curse?" *Unpublished manuscript, 1987*.
- Mehrens, W. A. & Lehmann, I. J. (1987). Using teacher-made measurement devices. *NASSP Bulletin*, 71(496), 36-44.
- Newman, D. C., & Stallings, W. M. (1982). *Teacher Competency in Classroom Testing, Measurement Preparation, and Classroom Testing Practices* Nilsson; I. and Wedman, I. On test- wiseness and some 'related

constructs, Educational Reports, UMEA, No. 7, 1974.

Nungester, R. J., & Duchastel, P. C.(1982). Testing versus review: Effects on retention. *Journal of Educational Psychology*, 74(1), 18-22.

Stiggins, R. J., & Bridgeford, N. J.,(1985). The ecology of classroom assessment. *Journal of Educational Measurement*, 22(4), 271-286.

Wiggins, G. P. (1993). Assessing student performance: *Exploring the purpose and limit of testing*. Jossey - Bass: San Francisco