

# Comparison of Traditional and Portfolio Assessment Efficiency in English Language Teaching in High Schools

Ramazan GOÇTÜ\*

## Abstract

The article reviews traditional (oral/written, testing) and contemporary (authentic, performance, including portfolio) assessment. Definitions are given. A quantitative research (experiment held with 43 11th graders at Private Demirel College (Tbilisi, Georgia) in the academic year 2011/2012) is presented. The goal of the research was to test our hypothesis – that the application of portfolio assessment can really have a positive impact on teaching English (the level of students' skills). Though the study has its limitations, the conclusions made are: though traditional and portfolio assessment both have advantages and disadvantages, application of portfolio is more advantageous. The experiment has shown that the level of EFL skills increased faster in both experimental groups than in the control group, which proves portfolio's higher efficiency.

**Key words:** traditional assessment, oral and written assessment, testing, authentic assessment, portfolio.

## Definition of Terms

The traditional assessment in English Language Teaching involves:

-oral assessment (question/answer or interview of the examinee, retelling a text, speaking on a given topic, making up a dialogue on a given situation),

- essays,

- and testing (multiple choice, gap filling with choices in the box or without, close test, matching, true/false/no information, puzzle -arranging paragraphs, sentences, words in order, transformation - statement → question, Present → Past, etc.).

Though relatively new to Georgia, testing can be viewed as traditional way of assessment as worldwide it has been in use for decades.

While a desire to see authenticity in assessment is not new, the term "authentic" first appears in reference to educational tasks and achievements and not necessarily assessments (Archbald and Newmann, 1988). Wiggins (1989) suggests that authentic assessment should be associated with authentic achievement. In the 1980's educators looked for an alternative to standardized assessments. As with authentic assessment, researchers still seek to define performance assessment. Currently, the two terms, "authentic" and "performance assessment" are associated together, and sometimes even with a third term, "alternative assessment". Authenticity in assessment typically references those elements that most closely align with real-world tasks. Performance assessment emphasizes competency manifest through real-world actions; alternative assessment emphasizes an assessment other than old

or objective testing assessments (Miller, Linn & Gronlund, 2008).

According to Marzano, Pickering & MicTighe (1993) the response to educational reforms in the 1980's was emphasizing performance assessment and trying to find ways to more fully integrate it into the educational experience. The research team identified three reasons why educational assessment required reform through the use of performance assessments:

- the changing nature of educational goals;

- the relationship between assessment and teaching and learning;

- and the limitations of the current methods of recording performance and reporting credit.

The past decade has witnessed great change in the field of educational measurement. Standardized multiple-choice testing is no longer as popular as it used to be. The "new" assessment is offered which is termed the performance assessment, the authentic assessment or the "3 P's" (performance, portfolios, and products).

The idea of using portfolios as an instrument in performance-based assessment is not new. From past to present, especially painters, artists, writers, models and photographers have exhibited their vocational and acquired skills through portfolios (Zollman & Jones, 1994). The practice of using portfolios in job searches appears to be gaining popularity as well (Costantino & Lorenzo, 2002; Wolf & Dietz, 1998; Wyatt & Looper, 2004).

As Knapper and Wilcox (1995) have described, the recent origins of portfolio application in education, which is one kind of performance assessment, can be traced back to the work of a committee of the Canadian Association of

\* A Doctorate Student at International Black Sea University, Tbilisi, Georgia;

E-mail: ramazangoctu@yahoo.com

University Teachers (CAUT), which was concerned in the 1970's with the undue reliance on student ratings for the evaluation of teaching. This group called for a more broadly based approach to evaluation that would use multiple sources of information and place responsibility for compiling the documentation on the individual faculty member rather than a remote administrator. The committee chair, Bruce Shore, first articulated the idea that faculty members should build their own case for teaching effectiveness—a “portfolio of evidence” to demonstrate competence (Shore, 1975, p. 8). Shortly afterward, the committee set about preparing its Guide to the Teaching Dossier, which was first published in 1980; it appeared in a second edition in 1986 and has been widely emulated and excerpted since then in a variety of publications all over the world (Shore et al, 1980, 1986).

There are various kinds of portfolios described in literature. A *writing portfolio* is a purposeful collection of student texts that demonstrates students' effort, progress, and achievement in writing over a period of time (Weigle, 2007). A *showcase focuses* on the student's best and most representative work. It is similar to an artist's portfolio where a variety of work is selected to reflect the breadth of talent. Therefore, in this portfolio the student selects what he or she thinks is representative work. This folder is most often seen at open houses and parental visits (Columba & Dolgos, 1995: 174-175). The primary function of an *assessment portfolio* is to document what a student has learned. The content of the curriculum, then, will determine what students select for their portfolios. Their reflective comments will focus on the extent to which they believe the portfolio entries demonstrate their mastery of the curriculum objectives.

### Research Goal

I (Goctu, 2012) have analyzed the “traditional” and the alternative (authentic, performance) assessment and have come to the conclusion that both of them have advantages and disadvantages, thus, at least for now, a combination of them is an effective approach. However, my hypothesis is that contemporary (e.g., portfolio assessment) is more effective, as it is authentic, student-centered and thus motivating.

The goal of the research was to check our hypothesis – that the application of portfolio assessment can really have a positive impact on teaching English (the level of students' skills). The reasons are that portfolio assessment gives a more accurate measure of student's achievement and it contributes more to students' progress.

This is why I held an experiment (quantitative research), as measurements had to be applied to prove or disprove the hypothesis.

The aims of the experiment were the following:

- to increase student motivation in experimental groups through the application of portfolio,
- to provide students' active involvement in learning
- to stimulate students to take responsibility for their learning
- to develop a positive attitude toward all skills in English,
- to raise the students' general language competence
- to encourage students to study and work independently
- to give a chance to the students to reflect about their work and knowledge
- to encourage teacher-student conferencing

### Procedure

The experiment occurred with at the 11th graders (11-A, 11-B, 11-C.) at Private Demirel College Tbilisi, Georgia) in the academic year 2011/2012. The control group was taught without the application of portfolios, while in experimental groups two different kinds of portfolios were used (show-case portfolio in one group and teacher-student assessment portfolio – in the other). Otherwise teaching in all three groups was identical – the same number of hours (4 lessons per week, 40 minutes each), the same textbooks (Aim – High by S. Iannuzzi, P.Kelly, Oxford University Press, 2011), the same methods of material presentation, the same activities and assessment. The textbook contains portfolio assessment tasks, but very few. At the end of each unit there is a part where students self-assess their knowledge in order to check students' progress in different skills. Though the book involves portfolio assessment, it is not emphasized very much and for this reason, teachers were not using it before my experiment. A pre-test, 2 while-tests and a post-test were used to measure and compare the scores of the students. Both methods aimed at revealing the results before and after experimental instruction in all groups.

The students underwent a 15-week-long action research in the second semester of 2011-2012 academic year. The participants and the results are discussed separately in subsections.

In the educational process all groups periodically took the same tests, but the students of the control group were also assessed (for semester grades) through oral answers, individual written work like dictation, question-answer, and essays and testing (no self-assessment results were included), while the students of the experimental group were seldom assessed by oral answers, individual written work like dictation, question-answer, and essays and often assessed through portfolio:

- a) show-case portfolio in group A
- b) teacher-student portfolio in group B

In both experimental groups' semester grade 30% con-

stituted portfolio assessment results and 70% - traditional (mostly testing) assessment results. experimental groups are presented in the table:

The differences between the instruction in control and

Groups	Control group – traditional assessment	Experimental groups – portfolio assessment
Type of assessment	Traditional methods, testing, essay questions.	Portfolio assessment and testing
Activities	Individual, pair, group and whole-class work, with emphasis on teacher feedback	Individual, pair, group and whole-class work, with emphasis on student feedback and creativity
Centered	Mostly teacher centered	Student centered

Table 1. Differences between traditional and experimental instruction

Every week students of the experimental groups brought their portfolios (some – included in the book, others – given by the teacher) to share and show their own documents to each other. Once a week, teacher was checking the all portfolios.

*In group A* the students prepared some pieces of work in order to describe their interest in related areas and held presentations (showcase) once a week. During the presentations everybody could ask presenters questions concerning his/her portfolio. Assessment (based on criteria that students knew beforehand) was done by teacher and students in cooperation. For example, one of my students

whose portfolio was assessed as the best in the class at the end of the year, likes playing the guitar very much, therefore, drew a picture of guitar, in his paper there were some sayings and quotes by famous guitarists. There was also his article written for the students' olympiad. It included several different samples of creative writing (a poem and a short story, an essay, a written assessment for a novel study/unit test, and a presentation/project).

Criteria of assessment of students' portfolios were:

(Text) volume of one piece should not be less than a page and should not exceed 2 pages (so that they are commensurable).

To pass, it is necessary to obtain 5 points, so all aspects should not be satisfactory.

	Satisfactory	Good	Excellent
Contents	0.5	1	1.5
Structure (logical)	0.25	0.5	1
Language: vocabulary	0.5	1	1.5
grammar	0.25	0.5	1
style	0.25	0.5	1
Visual aspect	0.25	0.5	1
Answering questions on the file	0.5	1	1.5
Overall impression	0.5	1	1.5
<b>Total</b>	<b>3</b>	<b>6</b>	<b>10</b>

Table 2. Rubric for portfolio assessment

**In group B (teacher-student portfolio)** after each unit I gave students an assessment paper of a studied unit which was divided in two columns. The first was assessed by student himself/herself, while the second - by the teacher. Teacher and student discussed those parts of portfolio where their assessments differed. Both gave reason why they thought so and they had to come to a consensus. By the end of the semester students' and teachers' assessments basically coincided, which means that students learned to do a more objective self-assessment, to look at their knowledge and skills critically and realistically. This (hopefully) also should lead to better learning, as students knew about their strengths and weaknesses and understood the need to take them into consideration in the process of study. This kind of portfolio seems less creative and exciting, but it certainly is student-centered.

**Participants**

At the beginning of the second semester of 2011-2012 academic year seventeen-year-old students of the 11th grades at Private Demirel College were selected with their consent for the experimental teaching. I was informed by the

school administration of these students that they were weak and their grades in language courses were not very low but average. But this was not a problem, as the main reason was to select students with lack of motivation in order to increase it. I observed the classes and spoke to students, so I understood all 11th graders were bored having traditional assessment all the time. They were really fed up with testing, essay questions, multiple choice questions, retelling texts by heart, and others. The students filled in the questionnaire forms, which contained questions about students' age, nationality, duration of English language learning, and social conditions of the family.

**Features of class A (experimental)**

The students had low motivation in learning. Their interest was deficient in English language. The students' language scores revealed their weakness. The students were divided into the groups with approximately equal language levels according to their language learning aptitude: at pre-intermediate level – 5 students, intermediate- 4 students, upper-intermediate – 5 students.

<b>Total number of students</b>	13	
<b>Nationality</b>	12 Georgian	1 Azerbaijani
<b>Age</b>	17	
<b>Language Level</b>	9 intermediate	4 upper intermediate
<b>Social Condition</b>	Average	
<b>Sex</b>	9 male	4 female

*Table 3. Statistical data of class A*

The students were 17 years old by the time the experimental teaching started. Four of them were girls and the rest of them are boys. All except one were Georgian. There was one Azerbaijani student. The question about social condition of the students' families has revealed that only in 5 cases both parents work, and one of the parents works in 8 students' families. The duration of experience of learning English (at school, at language courses, and with tutors - altogether) varied from 3 to 10 years.

**Features of class B (experimental)**

The students had low motivation in learning. This group does not have enough competence learning English language. The students were divided into the groups with approximately equal language levels according to their language learning aptitude: at intermediate- 11 students, upper-intermediate – 6 students.

<b>Total number of students</b>	17	
<b>Nationality</b>	14 Georgian	3 Azerbaijani
<b>Age</b>	17	Two students are 16
<b>Language Level</b>	11 Intermediate	6 upper intermediate
<b>Social Condition</b>	Average	
<b>Sex</b>	7 male	10 female

Table 4. Statistical data of class B

The majority of students were 17 years old by the time the experimental teaching started. In this group there were 10 girls and 7 boys. Majority were Georgian, except three students. There were three Azerbaijani students. The question about social condition of the students' families has revealed that only in 5 cases both parents work, one of the parents works in 9 students' families, 3 students' parents are unemployed. The duration of experience of learning English (at school, at language courses, and with tutors - altogether) varied from 3 to 10 years.

#### Features of class C (control group)

6 students had the pre-intermediate level, 2 students - intermediate, while 5 students - upper-intermediate level.

<b>Total number of students</b>	13	
<b>Nationality</b>	11 Georgian	2 Azerbaijani
<b>Age</b>	17	Two students are 16
<b>Language Level</b>	6 pre intermediate 2 Intermediate	5 upper intermediate
<b>Social Condition</b>	Average	
<b>Sex</b>	8 male	5 female

Table 5. Statistical data of class C

The students were 17 years old by the time the experimental teaching started. In this group there were 5 girls and 8 boys. The majority were Georgian except two students. There were two Azerbaijani students. The duration of learning English (at school, at language courses, and with tutors - altogether) varied from 3 to 10 years. Only in 4 cases both parents work, one of the parents works in 7 students' families, and 2 students' both parents are unemployed.

#### Description of the tests used for language skills assessment in all groups

As we have mentioned students were assessed before experiment, 2 times during the experiment and once after the experiment. All these tests were in the same format to provide comparability of results. We applied them in all three groups. We avoided using portfolio in assessment. The tests involved the following tasks: vocabulary (naming a synonym, choose the correct word), reading compre-

hension (true, false, no evidence; answer the questions), grammar (gap-filling, making up sentences with the given word), checking reading/writing/grammar and vocabulary skills simultaneously (completing the dialogues). It is easy to see that tasks include closed and open-ended, mechanical and creative, receptive, reproductive and productive tasks. This structure of tests permits to measure students' skills both holistically and according to skills, objectively and communicatively. The assessment is done out of 100 points and then transformed into traditional school assessment (10 points in Georgia)

#### Results and discussion

Test results are shown in tables below.

Student	Pre-test	While-test-1	While-test-2	Post-test
Student 1	8	9	9	10
Student 2	10	10	10	10
Student 3	7	9	10	9
Student 4	9	10	10	10
Student 5	8	8	10	10
Student 6	10	10	10	9
Student 7	8	8	10	10
Student 8	9	9	9	10
Student 9	8	10	10	10
Student 10	9	8	9	10
Student 11	7	9	9	10
Student 12	9	9	9	10
Student 13	9	8	8	9
Mode	9	9	10	10
Median	8.5	9	8.5	9.5
Mean	8.54	9.00	9.58	9.83
Standard deviation	0.97	0.82	0,51	0.39

*Table 6. Test results for Experimental Group A (Showcase Portfolio)*

The results in the table 5 show that the grades in this group were constantly improving. Not only the mean, but also the mode (most often received grade) was growing. Median (the middle point between the top and the lowest result), which often demonstrates the trends better, fell to some degree in the second while-test, but finally demonstrated growth. What is interesting is that the standard deviation is decreasing from time to time, which shows that the level of skills in the group does not simply increase on average, but also from student to student, which means that the improvement of the average is not reached due to the effort of a few students, but due to the growth of skill level of the majority of students.

Student	Pre-test	While-test-1	While-test-2	Post-test
Student 1	8	10	9	10
Student 2	10	10	10	10
Student 3	8	9	8	9
Student 4	9	10	10	10
Student 5	8	8	9	10
Student 6	10	10	10	9
Student 7	8	8	7	9
Student 8	9	9	8	9
Student 9	8	10	10	10
Student 10	7	8	9	8
Student 11	7	8	8	8
Student 12	10	9	9	10
Student 13	9	8	8	8
Student 14	8	9	9	10
Student 15	8	8	9	9
Student 16	7	7	10	10
Student 17	7	8	8	9
Mode	8	8	9	10
Median	8.5	9	9.5	9.5
Mean	8.2	8.7	8.8	9.2
Standard deviation	1.04	0.97	0.92	0.70

*Table 7. Test results for Experimental Group B (Teacher-Student Portfolio)*

The results in the table 6 show that the grades in this group were constantly improving. Not only the mean, but also the mode and the median were growing. What is interesting is that the standard deviation is decreasing from time to time, which shows that the level of skills in the group does not simply increase on average, but also from student to student, which means that the improvement of the average is not reached due to the effort of a few students, but due to the increase of skill level of the majority of students.

Student	Pre-test	While-test-1	While-test-2	Post-test
Student 1	8	8	8	7
Student 2	7	8	9	9
Student 3	6	9	9	9
Student 4	5	6	6	6
Student 5	8	7	7	8
Student 6	10	8	9	9
Student 7	6	6	8	7
Student 8	9	7	8	8
Student 9	8	8	6	8
Student 10	7	8	9	9
Student 11	8	9	9	8
Student 12	9	9	9	8
Student 13	7	8	7	8
Mode	8	8	9	9
Median	7.5	7.5	8	7.5
Mean	7.54	7.69	7.92	7.85
Standard deviation	1.39	1.03	1.86	0.90

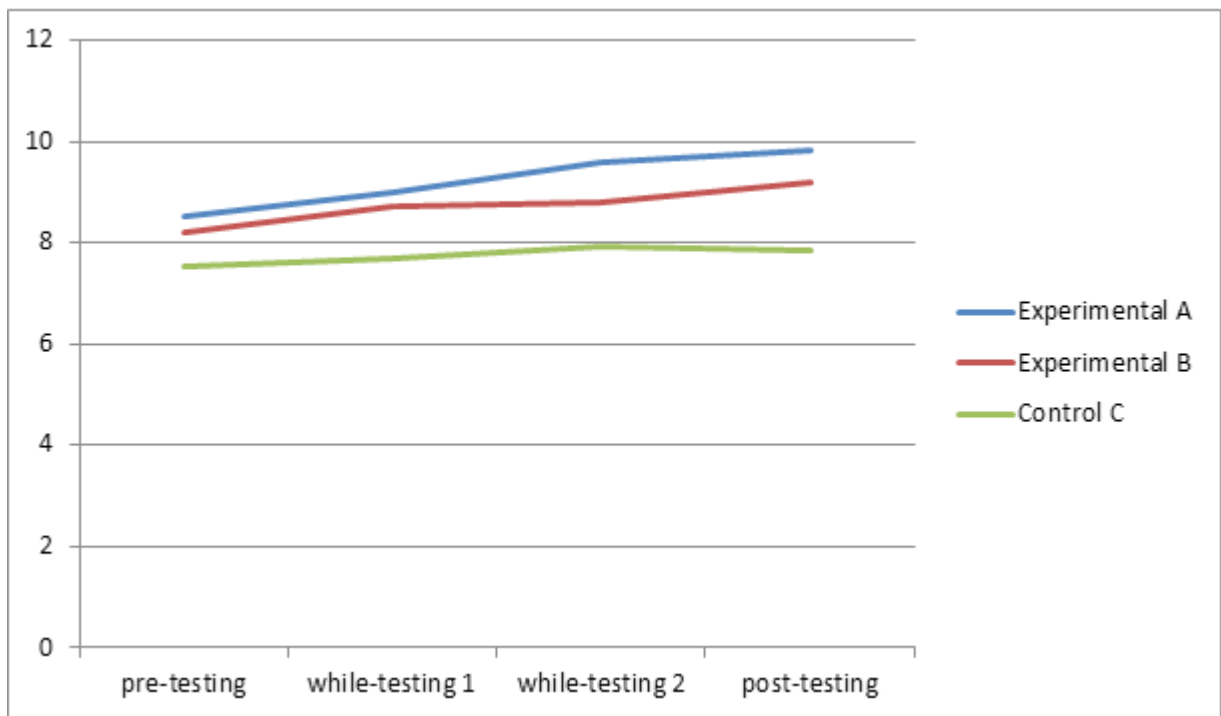
*Table 8. Test results for the Control Group C*

It is important how much the students' skills improve and if they improve at all. Group C was somehow weaker (pre-test result 7.54) than group A (pre-test mean result 8.54) and B (8.2). The problem is not that their final result is lower, but the fact that the language skill level increases slower, in fact, practically does not increase (mean 7.54 → 7.85).

What we can see is that the median is not increasing and the mode and the mean are increasing slowly, and the mean even drops a bit in the end, so the results in this class are not as good as in the experimental classes. Besides, the standard deviation in three cases out of four is above one (in a 10-point assessment system), which means that the class was rather heterogeneous in the beginning of the experiment and remained so till its end.

To make the results more visual and comparable, let us present them in graphs.





Graph 1. Mean results during the experiment

One can notice fast and more or less even growth in experimental group A, uneven, but finally stable growth in group B and slow growth and then even a small drop in group C. This means that both experimental groups showed better results than the control group, besides, the experimental group A (show-case) showed a higher result than the experimental group B (teacher-student assessment portfolio).

### Limitations

The limitations of the research are connected with its scale (43 students at one school during one semester) and limited geographic (cultural) area (one country). To come to more decisive conclusions, larger scale research is necessary in various cultural settings.

The research deals with foreign (English) language teaching, and results may be rather different concerning teaching other subjects. Even the results for English as a second language may differ.

Besides, the difference between the results obtained from the three groups exists, but is not crucial. One reason may be that I didn't use only portfolio assessment in groups A and P, but a mixture of portfolio (30%) and traditional (70%) assessment. Another reason may be that (exactly as I have hypothesized) traditional assessment has its advantages and doesn't have to be totally abandoned.

### Conclusions

Comparison of traditional and portfolio assessment held by me both on theoretical and experimental level is in favor of portfolio method. The experiment has shown that the students' involved in the experiment level of English as a foreign language skills increased faster in both experimental groups than in the control group, which proves portfolio's higher efficiency.

### References:

- Archbald, D. A. & Newmann, F. M. (1988). Assessing authentic academic achievement in the secondary school (Reston, Virginia, National Association of Secondary School Principals).
- Columba, L., and Dolgos, K.A. (1995). Portfolio assessment in mathematics. *Reading Improvement*, 32 (3), 174-176
- Costantino, P., & De Lorenzo, M. (2002). *Electronic portfolios: An option*. In *Developing a Professional Teaching Portfolio*. Boston: Allyn and Bacon: 47-55
- Göçtü, R. (2012). Portfolio assessment in English language teaching. *IBSU Journal of Education*, 1 (1), p.23-27
- Knapper, C. K., and Wilcox, S. (1998). *Preparing a Teaching Dossier*. Kingston, Canada: Queen's University Instructional Development Centre
- Marzano, R.J., Pickering, D., & MicTighe, J. (1993). *Assessing Student Outcomes: Performance Assessment Using the Dimensions of Learning Model*. Virginia, U.S.: Association for Supervision and Curriculum Development

- Miller, M.D., Linn, R.L., & Gronlund, N.E. (2008). *Measurement and Assessment in Teaching*. New Jersey, U.S.: Merrill
- Shore, B. M. (1975). Moving beyond the Course Evaluation Questionnaire in Evaluating University Teaching. *CAUT Bulletin*, 23(4), 7-10
- Shore, B. M., and Associates. (1980). *The Teaching Dossier: A Guide to its Preparation and Use*. Ottawa: Canadian Association of University Teachers
- Shore, B. M., and Associates. (1986). *The Teaching Dossier: A Guide to its Preparation and Use*. (rev. ed.) Ottawa: Canadian Association of University Teachers
- Weigle, S. (2007). Teaching writing teachers about assessment. *Journal of Second Language Writing*, 16/3: 194-209
- Wiggins, G. (1997). Practicing What We Preach in Designing Authentic Assessments. *Educational Leadership*. December 1996-January 1997: 18-25
- Wolf, K., & Dietz, M. (1998). Teaching portfolios: purposes and possibilities, *Teacher Education Quarterly*, Vol. 25(1), pp. 9-22
- Zollman, A., & Jones, D.L. (1994). Accommodating student assessment and learning: Utilising portfolios in teacher education with pre-service teachers. Paper presented at the 21st Annual Conference of the Research Council on Diagnostic and Prescriptive Mathematics, Fort Worth, Texas