

مجله علوم اجتماعی و انسانی دانشگاه شیراز
دوره بیستم، شماره اول، تابستان ۱۳۸۲ (پیاپی ۳۹)
(ویژه نامه علوم تربیتی)
مقاله کوتاه

بررسی اثر ارزشیابی تکوینی بر پیشرفت دانشجویان

دکتر حسین سپاسی
دانشگاه شهید چمران اهواز

چکیده

استادان همیشه درگیر اثرات ارزشیابی تکوینی بر یادگیری هستند. بسیاری بر این عقیده اند که ارزشیابی تکوینی پیشرفت بیشتری به دانشجو می دهد و بر نگرش مثبت به یادگیری تأکید می کند. این پژوهش به بررسی این فرضیه می پردازد که استفاده از ارزشیابی تکوینی فراگیری را بهبود می بخشد. صد و دوازده نفر از دانشجویان تربیت دبیر دانشگاه شهید چمران که در درس آزمون و ارزشیابی حضور داشتند در این پژوهش شرکت کردند. این دانشجویان به طور تصادفی به دو گروه تقسیم شدند و در پیش آزمون شرکت نمودند. در ابتدای دوره هیچ تفاوت معناداری در زمینه اطلاعات، مهارت ها و درک بین آنها وجود نداشت ($P < 0.05$). به گروه آزمایشی پنج ارزشیابی تکوینی و یک ارزشیابی نهایی و به گروه دیگر فقط ارزشیابی نهایی داده شد. روش آموزش برای هر دو گروه یکسان بود و تنها گروه آزمایشی، ارزشیابی تکوینی دریافت نکرد. نتایج نشان داد که تفاوت معناداری در پس آزمون دو گروه وجود دارد. این نتایج نشان می دهد که استادان باید توجه بیشتری به مزایای ارزشیابی تکوینی در پیشرفت دانشجویان نشان دهند.

**THE ENGLISH
SHORT
ARTICLE**

A Study of the Effect of Formative Evaluation on Student Achievement

Dr. H. Sepasi*
Shahid Chamran University

Abstract

Teachers have always been concerned with formative evaluation's contribution to learning. Many believe that the use of formative evaluation produces greater student achievement and affects more positive attitudes toward learning. This study was designed to test the hypothesis that the use of formative evaluation improves learning outcomes. A total of 112 Shahid Chamran University teacher training students, enrolled in a measurement and evaluation course served as subjects of this study. They were randomly assigned to experimental and control groups and were given a pre-test. No significant difference was found between the two groups regarding their knowledge, skill and understanding ($p < .05$). The experimental group participated in five formative tests and a final examination, whereas the control group received only the final. Instructional strategies were the same for the two groups except for the experimental treatment (formative evaluation). The results showed a significant difference between the post-test of the two groups. This finding indicates that teachers should pay attention to the contribution of formative evaluation in student achievement.

Key words: 1. Formative evaluation 2. Summative evaluation 3. Student achievement 4. Testing 5. Educational measurement

1. Introduction

The history of educational measurement and evaluation is a long one. The initiation ceremonies by which primitive tribes tested the knowledge of tribal customs, endurance and readiness of the young for admission to the rank of adulthood can be linked to what we consider today as formative evaluation, and may be among the earliest examinations employed by human beings (Romberg, 1992). From historical references, we know that Chinese officials used test results, possibly as summative procedures, as the method by which to select their administrative officials 2200 years ago (Ebel and Frisbie, 1991).

Today the goals of educational evaluation are quite different from the past, and have been changed drastically to serve many different objectives. Nevertheless, even classroom tests teachers design to assess acquisition of subject matter knowledge throughout a course, an example of formative evaluation procedures, or the large scale standardized tests, example of summative evaluation methods, continue to be used in educational and vocational settings. With the expansion of western-style systems around the world, evaluation procedures for assessing students learning have shifted from individual or small heterogeneous group formative evaluation to large nationally standardized tests. Although educational professionals in western schools can be expected to continue to modify their assessment methods to account better for

*Associate professor of education

individual and grade level learning outcomes, these methods are appearing in schools in the Middle East, and offer the promise of more customized assessment of student learning. Teachers in Middle Eastern schools are shifting from individual oral assessment to various types of testing procedures.

2. Review of Literature

A great deal has been written about these two major and very different kinds of evaluation-formative and summative. It is widely accepted that these two methods of evaluation reflect different aspects of learning. Formative methods reflect daily or weekly acquisition of knowledge and processes. Summative evaluation samples the domain of knowledge under consideration. Formative evaluation refers to the use of data to make a process or operation effective as it goes along (Scriven, 1976). Summative evaluation occurs at the conclusion of an act or process; it is terminal assessment at the end of a course.

Generally, formative evaluation measures the status of educational programs while they are still in some stages of development. Mehrens and Lehman (1984), state that by formative evaluation, we mean that gathering information during the learning process in order to increase student learning experiences. Schutz (1970) has said that the product of formative evaluation activities is expected to be an improved instructional program, and that it is intended to provide data for instructional product validation. Experts in educational measurement (Biggs and Collis, 1982; Ebel and Frisbie, 1991 and Bloom, 1976) believe that special attention should be given to the importance of formative evaluation in assessing student achievement and learning processes. They also stated that formative evaluation should be used as a means to determine the strength and weakness of students' learning experiences.

Bloom (1976) contends that formative evaluation can contribute to a positive attitude toward what has been taught in the classroom. The significant relationship between attitudes and measures of school achievement makes it clear that effective characteristics are important in determining or influencing student achievement. Bloom estimated that positive attitudes could account for up to 25 percent of the variance in resultant academic measures.

Formative evaluation can also reinforce student expectations that what has been taught by the teacher will be the same as what will be evaluated, helping them focus more clearly on achievement objectives. Funch (1985) has found out that the student teachers in an experimental group whose students assessed by means of formative evaluation were more positive than those of control whose students did not receive any feedback regarding their performances.

Since these formative classroom assessments of student achievement can be very important to the teaching and learning process, many have criticized the way formative evaluation measures are constructed. Stiggins, Griswold and Wikelund (1989) mention that paper-pencil tests are dominated by recall questions across all grade levels and forms of evaluation. Comprehension and evaluation questions were rare. Although the effectiveness of formative evaluation is emphasized by many experts as a means to facilitate student learning experiences, it is, however, believed that more research is needed to investigate and compare the applications of different learning processes and different evaluation techniques in assessing students achievement (Feldhusen, 1964; Piage, 1966; Akin, 1969 and Chansarkar & Routary, 1981). It is also likely that while teachers from different disciplines hope to teach thinking skills to some extent, they are less often trained to assess such skills. Moreover, teachers in the Middle East have

different pedagogical traditions, and the extent to which formative evaluation can affect achievement outcomes deserves a test of its applicability in a non-Western setting.

Research on student learning indicates that students are not passive participants in constructing their own understanding of subject matter. Rather, they build existing knowledge and understanding (Wilson, 1992). These new views of learning have been followed by proposals for new testing theories and model to replace the traditional achievement testing approaches which place emphasis on precise defined student behaviors such as those in criterion-referenced testing (Bloom and Hastings, 1971). These new testing theories such as performance assessment (Vanhiele, 1986 and Funch, 2000) and the structure of learning outcome taxonomy (Biggs and Collis, 1982) require teachers to possess strategies to classify student assessment or evaluation assessment of the acquisition of such knowledge.

Although these are new and promising approaches to assess knowledge, skills and understanding of subject matter, the point that needs to be made clear is that teachers still need to know how these new approaches can teach the ways children's knowledge can be assessed, and the type of new assessment approaches that can be administered in the classroom. Those responsible for teachers' education should provide teachers the opportunity to find out how these new testing theories and models can be applied to measuring higher order thinking skills and what kind of classroom approaches would best serve the assessment of student achievement.

The course for which this study of formative evaluation was designed was a course of measurement and evaluation for teachers and the content of the course focused on measuring student achievement. This study was conducted in an effort to explicate the role of formative evaluation in acquiring knowledge of educational measurement. The primary purpose of this study was to determine the effect of formative evaluation on student achievement. The hypothesis that was tested for this study was "there are significant differences between the final scores of students two participated in the formative evaluation program and those who did not". Furthermore, the investigator planned to make some preliminary observations of student participation in class of which inferences about the effect of the course on student attitudes could be made. For this latter purpose, attendance records and gains in percentages of correct answers on the formative tests were determined.

3. Subjects

A total of 112 students majoring in Persian literature who had chosen to enter a teacher training program were the subjects of this study. These students enrolled for a two hour measurement and evaluation course. The students were randomly divided into experimental and control groups.

4. Procedures

At the beginning of classes, course content including objectives of the course, testing procedures, grading system, references (books and articles) were brought to the attention of both experimental and control groups. Then, in order to find out whether the two groups were similar in terms of their knowledge, skills, and understanding of the course, both groups participated in a pre-test on the subject-matter of the course, "Measurement and Evaluation".

During the course, subjects in the experimental group were given five short examinations (formative tests) at 3-week intervals. At the end of the term, the subjects in the experimental group were given a final examination (Summative test). In other words, the testing procedures for the experimental group consisted of a pre-test, five

short examinations (experimental treatment), and a final examination (post-test). For the control group, the testing procedures consisted only of the pre-test and the final test. Teachers in the experimental group were active in preparing for the five short formative exams, each of which received a value of 10 percent totaling 50 percent of each student final grade. The remaining 50 percent of the grade were based on the final examination.

The information that students should have about the testing procedures were all made clear to the subjects in the experimental group in advance. Each one of the short exams was given to students as the class started; a week later the test papers were given back to the students to show them how they performed in that particular examination. Moreover, basic statistical information regarding raw scores means, standardized scores and the ranks of each individual in class were made available on the test paper. Also, feedback regarding their performance on the tests, and their strengths and weaknesses were brought to the attention of the subjects in the experimental group.

In order to be able to study the possible effects of the experimental treatment, the control group participated only in a pre-test and in final examination (post-test). Their semester evaluation was based entirely on how they did on the final examination. In other words, the control group did not receive any short formative exams (experimental treatment) nor did they receive any feedback on strengths and weaknesses of their performances during the course beyond that normally given during instruction. On the other hand, instructional procedures remained the same for both groups during all the class periods to maintain a similar structure for both groups, except for the experimental treatment. In this case any differences observed at the end of the experiment could be attributed to the effect of the experimental treatment.

Records were kept of attendance in both experimental and control classes. These data as well as the percent of correct responses on the formative tests were used as a basis for making a preliminary assessment of student involvement and interest in the course.

5. Instruments

Two equivalent forms of a test were used for the experiment. Form-B and Form-A were randomly selected for pre-test and post-test respectively. The reliability coefficient of the two forms of the test on students who had previously enrolled for the same course was computed and reported as $r = .95$. Since the contents of test items were so chosen to measure the course objectives, the content validity of the two equivalent forms of a test was reported to be capable of achieving the content area of interest. Content validity of achievement test is by far the most important type of validity (Mehrens and Lehman, 1984; Ebel and Frisbie, 1991; Crocker and Algina, 1986).

6. Experimental Design

For the purpose of this study, the following randomized control-group pretest-posttest design was selected (see table 1). In order to test the hypothesis of the study, a correlated t-test was used to examine the significant differences between the pre-test scores of the experimental group. Also, the same procedure was used to test the significant differences between the pre-test and post-test scores of the control group. Finally, to study the effect of the experimental treatment, an uncorrelated t-test was conducted to compare the final exam scores of the experimental and control groups. Moreover, for the purpose of testing whether there were any differences between the two groups prior to the study, the uncorrelated t-test was used to study differences between the pre-test scores.

Table 1: Experimental Design

	Pretest	Experiment	Posttest
Experimental	T ₁	X	T ₂
Control	T ₁	-	T ₂

7. Results and Discussion

7. 1. Pretests of experimental and control groups

Although the subjects in the study were randomly assigned to experimental and control groups it was decided to find out whether there were any significant differences at the beginning of the semester on their prior knowledge of the subject matter of the course.

The t-test ($t=1.17$, $n=56$) was not significant at $p<.05$. The lack of significant difference between the two groups prior to the experiment led this researcher to investigate the effects, if any of the experimental treatment. At the end of the study, any significant differences between the scores of the two groups could be attributed to the experimental treatment.

7. 2. Pretest-Posttest of experimental group

Comparison between the scores of the pre-test and post-test for the experimental group was not the main purpose of this study. The computed t-test, however, ($t =23.3$, $n =56$) was significantly different from the critical value at the level of $p<.05$. This means that the experimental group advanced their learning experiences during the term.

7. 3. Student attendance and gains in performance during the course

It was also noted from formal school records and from attendance records that none of the subjects in the experimental group missed any classes and rarely came late.

The gradual observed improvement in the percent of correct responses on the tests of each subject in the experimental group from an average of 80% on the first test to 95% on the last one provided feedback on their efforts in learning the materials. These results may be interpreted to mean that the continuous feedback that the experimental group received about their performances from each formative test score enabled them to make intra-individual and within group comparisons. In addition, they were able to assess their performance on each of the formative tests and compare their achievement with their efforts to learn the materials. This improvement in learning as well as their perfect attendance records can be taken as evidence of their attitudes toward the class as well as their progress in meeting the course objectives.

Table 2: Mean Scores and t-test of Pre and Post-Tests for Experimental and Control Group

	Pretest		Posttest		Sig.of t
	n	X	X	T	
Experimental	56	14.66	40.46	23.3	<.05
Control	56	15.12	36.14	20.41	<.05

7. 4. Pretest-Posttest of control group

Comparing the scores of the pre-test and post-test of the control group shows that the control group also performed at a significantly higher level at the end of the course. The t-test ($t=20.41$, $n=56$) was significantly different from the critical value of $p<.05$. Table 2 shows the mean scores of pre-test and post-test for both groups.

The performance of control group was lower and significantly different from that of the experimental group. Although the instructional procedures remained the same for the two groups, the results show that the control group performed at a lower level than the experimental group at the end of the course.

7. 5. Student attendance

Since no test was scheduled to be given during the semester and student evaluation was based only on a final test by examining the attendance records for the control group, it was noticed that there were some absences and more tardiness in contrast with the perfect attendance of the experimental group. This was interpreted as tentative evidence of the fact that the formative tests had a positive effect on student attitudes toward the test and measurement course.

7. 6. Posttests of the two groups

The major purpose of this study was to investigate the effect of the experimental formative test program on end-of-course achievement. There was a significant difference between the post-test scores of the experimental and control groups. The computed t-test ($t= 2.07$, $n= 56$) was statistically different from its critical value at $p<.05$.

The kind of information the students in the experimental group received about their performances from each test score may have enabled them to improve their learning experiences. On the other hand, the students in the control group who were not provided with feedback about their progress in meeting the course objective performed at a lower level, even though they did significantly better on the post-test. It means likely that formative evaluation, which takes place during the stage of learning new material, provides teacher and students with the basis for instructional changes and learning progress.

8. Conclusion

Although teachers cannot be held responsible for all factors that affect student achievement, they can be aware of the effectiveness of different evaluation techniques. The results of this study may indicate that students should be informed about what they have gained as soon as possible. This would help them to learn more about their strengths and weaknesses to adjust their learning experiences to attain course objectives.

Immediate feedback in formative tests can reinforce the learning process. Kulik and Kulik (1990) reported that teachers who want their quizzes to help students learn should try to arrange conditions so that students receive feedback as quickly as possible after they answer quiz question-to delay feedback is to hinder learning. Careful and sensitive planning, preparation, and processing of exams can make them vital parts of teaching and leaning (McMullen & Weimer, 1987).

Although it seems to be widely accepted in the literature that formative evaluation increases learning (Stuffelbeam and Webster, 1980), it still remains for teacher education programs to communicate this to teachers. Teachers should know that increased achievement is due to such factors as knowledge of progress (the manner in which the results of formative evaluation are treated), extrinsic motivation (students will work harder throughout the course because they want to get good grades), knowledge of

results (with a greater opportunity to see their areas of strength and weakness in the subject matter) and responses to class (evidenced by attendance records).

Some teachers believe that the formative evaluation is a waste of time and that it takes too much time to be implemented (Ross, 1974). Still, others say that frequent testing is necessary only for children in lower grades of students of low ability (Hopkings, 1987). The results of this study suggest that use of formative evaluation had a positive impact on teacher training students in a measurement and evaluation course.

Because the formative exams given during the course appear to account for significant differences on summative evaluation or the final exam at the end of the course, the findings of this study support Romberg, Wilson and Khaketa's (1990) claims that progress through the mastery of simple steps, the development of learning hierarchies, explicit directions, daily lesson plans, frequent quizzes, objective testing of smallest steps in learning facilitate students learning.

This study also supports the need for future teachers to be aware of different testing and evaluation procedures and the value of formative assessment as an effective instruction method to promote student achievement (Biggs and Collis, 1982; Porter, 2000; and Stecher, 2000).

References

- Akin, M. G. (1969). **Evaluation Theory Development**, Belmont, CA: Wordsworth Publishing Co.
- Biggs, J. B. & Collis, K. F. (1982). **Evaluation: The Quality of Learning, SOLO Taxonomy, Structure of the Observed Learning Outcome**, San Diego, CA: Academic Press.
- Bloom, S. B. (1956). **Taxonomy of Educational Objective Handbook, the Cognitive Domain**, New York: David McKey Co.
- Bloom, S. B. (1968). **Learning for Mastery, Evaluation Comment, 1**, UCLA: Center for the Study of Evaluation.
- Bloom, S. B. & Hastings, M. (1971). **Handbook on Formative and Summative Evaluation of Student Learning**, New York: McGraw Hill Book Company.
- Bloom, S. B. (1976). **Human Characteristics and School Learning**, New York: McGraw Hill Book Company.
- Chansarkar, B. A. & Routary, U. (1981). *How Relevant is Continuous Assessment in Higher Education*, **Educational Research Journal**, 6, (1), 46-56.
- Crocker, L. & Agina, J. (1986). **Introduction to Classical and Modern Test Theory**, New York: Holt, Rinehart And Winston.
- Ebel, L. E. & Frisbie, D.A. (1991). **Essentials of Educational Measurement**, Engelwood Cliffs NJ: Prentice Hall.
- Feldhusen, J. E. (1964). *Student Perceptions of Frequent Quizzes and Post-Marten Discussion of the Tests*, **Journal of Educational Psychology**, 51(1), 51-54.
- Funch, L. R. (1985). *Effects of Curriculum-based Measurement and Consultation on Teacher Planning and Student Achievement in Mathematics Operations*, **Educational Research Journal**, 28, 617-641.

- Funch, L. S. (2000). *The Importance of Providing Background Information on the Structure and Scoring of Performance Assessment*, **Applied Measurement in Education**, 13(1), 201-209.
- Hopkins, C. D. (1987). **Classroom Measurement and Evaluation**, Itasca, Ill: F. E. Peacock Publisher.
- Kulik, J. A. & Kulik, C. C. (1990). *Timing of Feedback and Verbal Learning*, **Review of Educational Research**, 58, 79-97.
- McMullen, P. M. & Weimer, M. G. (1987). *Student Examination Accentuating the Positive*, **Education Digest**, 52, 14-17.
- Mehrens, M. A. W. & Lehman, J. I. (1984). **Measurement and Evaluation in Educational Research**, New York: Holt Rinehart and Winston Inc.
- Piaget, D. D. (1966). *Learning While Testing*, **The Journal of Educational Research**, 59, 276-277.
- Porter, P. R. (2000). *Testing the Academic Achievement of Limited English Proficient Students*, **Applied Measurement in Education**, 13 (4), 35-41.
- Romberg, R. A., Wilson, L. & Khaketa, M. (1990). **An Examination of State and Foreign Tests**, Madison, WI: Center for Educational Research.
- Romberg, R. A. (1992). **Mathematics Assessment and Evaluation, Imperative for Mathematics**, Albany: State University of New York Press.
- Ross, C. C. (1974). **Measurement in Today's Schools**, Englewood Cliffs NJ: Prentice Hall.
- Schutz, R. E. (1970). *The Nature of Educational Development*, **Journal of Research Development in Education**, 3, 39-46.
- Scriven, M. (1976). **The Mythology of Evaluation**, AERA Monograph Series on Curriculum Evaluation, (1), Chicago: Rand McNally.
- Stecher, B. M. (2000). *The Effect of Content, Format, and Inquiry*, **Applied Measurement in Education**, 13 (3), 34-41.
- Stiggins, R. J., Griswold, M. M. & Wikelund, K. R. (1989). *Measuring Thinking Skills through Classroom Assessment*, **Journal of Educational Measurement**, 26, 233-246.
- Stuffelbeam, D. L. & Webster, W. J. (1980). *An Analysis of Alternative Approaches to Education*, **Educational Valuation and Policy Analysis**, 2, 5-10.
- Vanhiele, P. M. (1986). **Structure and Insight. A Theory of Mathematics Education**, Orlando: Academic Press.
- Wilson, M. (1992). *Measuring Levels of Mathematical Understanding in R.A.* (Ed.), **Mathematics Assessment and Evaluation of State and Foreign Tests**, Albany: State of University of New York Press.

- Funch, L. S. (2000). *The Importance of Providing Background Information on the Structure and Scoring of Performance Assessment*, **Applied Measurement in Education**, 13(1), 201-209.
- Hopkins, C. D. (1987). **Classroom Measurement and Evaluation**, Itasca, Ill: F. E. Peacock Publisher.
- Kulik, J. A. & Kulik, C. C. (1990). *Timing of Feedback and Verbal Learning*, **Review of Educational Research**, 58, 79-97.
- McMullen, P. M. & Weimer, M. G. (1987). *Student Examination Accentuating the Positive*, **Education Digest**, 52, 14-17.
- Mehrens, M. A. W. & Lehman, J. I. (1984). **Measurement and Evaluation in Educational Research**, New York: Holt Rinehart and Winston Inc.
- Piaget, D. D. (1966). *Learning While Testing*, **The Journal of Educational Research**, 59, 276-277.
- Porter, P. R. (2000). *Testing the Academic Achievement of Limited English Proficient Students*, **Applied Measurement in Education**, 13 (4), 35-41.
- Romberg, R. A., Wilson, L. & Khaketa, M. (1990). **An Examination of State and Foreign Tests**, Madison, WI: Center for Educational Research.
- Romberg, R. A. (1992). **Mathematics Assessment and Evaluation, Imperative for Mathematics**, Albany: State University of New York Press.
- Ross, C. C. (1974). **Measurement in Today's Schools**, Englewood Cliffs NJ: Prentice Hall.
- Schutz, R. E. (1970). *The Nature of Educational Development*, **Journal of Research Development in Education**, 3, 39-46.
- Scriven, M. (1976). *The Mythology of Evaluation*, **AERA Monograph Series on Curriculum Evaluation**, (1), Chicago: Rand McNally.
- Stecher, B. M. (2000). *The Effect of Content, Format, and Inquiry*, **Applied Measurement in Education**, 13 (3), 34-41.
- Stiggins, R. J., Griswold, M. M. & Wiklund, K. R. (1989). *Measuring Thinking Skills through Classroom Assessment*, **Journal of Educational Measurement**, 26, 233-246.
- Stufflebeam, D. L. & Webster, W. J. (1980). *An Analysis of Alternative Approaches to Education*, **Educational Valuation and Policy Analysis**, 2, 5-10.
- Vanhiele, P. M. (1986). **Structure and Insight. A Theory of Mathematics Education**, Orlando: Academic Press.
- Wilson, M. (1992). *Measuring Levels of Mathematical Understanding in R.A. (Ed.), Mathematics Assessment and Evaluation of State and Foreign Tests*, Albany: State of University of New York Press.