

# An Empirical Study of Formative Assessment in College English Teaching

Qingshun He (Corresponding author)

Faculty of English Language and Culture, Guangdong University of Foreign Studies

North 2, Baiyun Avenue, Baiyun District, Guangzhou City, Guangdong, 510420, China

E-mail: markman1998@163.com

Received: August 1, 2014 Accepted: September 3, 2014 Published: September 13, 2014

doi:10.5296/ire.v2i2.6309

URL: <http://dx.doi.org/10.5296/ire.v2i2.6309>

## Abstract

It is generally accepted that formative assessment can better help students' autonomous learning and their cultivation of humane quality. This paper intends to examine whether formative assessment is universally applicable for students of different levels of language proficiency. For this purpose, an English teaching experiment and a questionnaire survey on learning strategies and motivations were conducted in a 211 Project university and the independent college of that university in China. Research shows that learning strategies and motivations have a threshold effect on the implementation of formative assessment, and formative assessment is applicable for the students who have better command of English and higher learning strategies and motivations, but it does not work for those poor in English and having lower learning strategies and motivations. Therefore, formative assessment cannot completely replace summative assessment in college English teaching.

**Keywords:** Formative assessment; Summative assessment; College English teaching; Learning strategies and motivations

## 1. Introduction

Education assessment can be divided into formative assessment and summative assessment. The former is "assessment for learning" (William, 2011). It supports teachers and students in decision-making during the educational and learning processes. The latter, however, "occurs at the end of a learning unit and determines if the content being taught was retained" (Ainsworth, 2006, p. 23). It "seeks to monitor educational outcomes, often for purposes of external accountability" (Shepard, 2005) and to examine whether the students have reached the preset standard of the curriculum. The subjects of the summative assessment are the teachers or the educational institutions, and it mainly functions to evaluate and identify the

students' achievements. In college English teaching in the background of China, the shortcomings of summative assessment include: (1) It cannot provide the feedback information timely for teaching and hence the teacher cannot improve their teaching in time to offer appropriate guidance to the students; (2) It evaluates the students by taking them as physical objects, focusing only on the students' language proficiency while ignoring the role of cognition, emotion and other factors in their learning processes; and (3) If there is only summative assessment in teaching, the teacher will be forced to conduct an exam-oriented education, and hence the students' value of subject will not be respected, which is not conducive to students' ability in autonomous learning.

Contrasted with summative assessment, formative assessment is "a range of formal and informal assessment procedures employed by teachers during the learning process in order to modify teaching and learning activities to improve student attainment" (Crooks, 2001). "It typically involves qualitative feedback (rather than scores) for both student and teacher that focuses on the details of content and performance" (Huhta, 2010) and the feedback can be used to "modify the teaching and learning activities in which they are engaged" (Black & William, 1998, p. 8). As for the impact of feedback on learning, there are two contrastive opinions. One is that the feedback will have a positive impact (e.g. Black & William, 1998; Hattie & Timperley, 2007; Shute, 2007). The other believes it will have a negative impact (e.g. Kluger & Denisi, 1996). According to Havnes et al. (2012), feedback becomes more useful for students as it is related to what they are doing at the moment by integrating feedback into instruction. In this research, we will investigate the impact of feedback through a questionnaire survey.

Formative assessment focuses on the practice in class, in which "evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited" (Black & William, 2009).

Formative assessment has the following characteristics. First, it follows the humanistic philosophy of education. Formative assessment came into being with the American humanism in the 1960s. It concerns about the efficacy of teaching and learning (Black & William, 1998); it is against the traditional goal-oriented evaluation model, emphasizing students' initiative and their comprehensive development in cognition, emotion and attitude, etc. Second, formative assessment emphasizes the process of students' learning, and hence the teacher and learners can both obtain timely the relevant feedback information which "gives chances to students to participate in modifying or planning the upcoming classes" (Bachman & Palmer, 1996). Third, both the teacher and students are the subjects of assessment. Under the guidance of the teacher, the students should actively summarize and reflect on their learning, and the teacher has also correspondingly changed their roles from the arbiter to the facilitator, encourager or helper. Fourth, formative assessment tools are diversified, such as self-assessment forms, learning portfolios, peer reviews, quizzes and questionnaires, etc. The final aim of formative assessment is to help students monitor their own learning process, enhance their self-confidence and cultivate a spirit of cooperation.

Then to what an extent can formative assessment be helpful for college English teaching in China? Relevant researches (e.g. Simpson, 1999; Henson & Eller, 1999) show that formative assessment can be helpful for cultivating students' autonomous learning ability, especially for those students of poor academic performance. The English learning ability and English proficiency of the students in Chinese colleges are at an overall lower level. If formative assessment is really universally applicable, then it is no doubt a best thing for the English teaching in Chinese colleges.

Therefore, this study attempts to answer the following three questions. (1) Is formative assessment better than summative assessment in improving the English proficiency of college students in China? (2) Can formative assessment better promote the college students' English learning strategies and motivations than summative assessment? (3) Is formative assessment popularly accepted by the students of different levels of English proficiency? To answer these questions, we will carry out an empirical study on formative assessment in college English teaching. In Section 2, the research design is offered, including the selection of subjects for research and the experimental teaching. Section 3 presents the results of the experimental teaching and the questionnaire survey. A discussion of the results will be conducted in Section 4.

## 2. Methodology

### 2.1 Subjects for Research

The subjects for research are sophomore students from two parallel classes of the economics department of a 211 Project university and two parallel classes of the business administration department of the independent college of the same university. Because of the large gap of the English proficiency between the two groups of students, it is comparative to choose the two groups of students to participate in the teaching experiment. We take class one in the economics department (or Economics 1) and class one in the business administration department (or Business 1) as the experimental classes, and class two in the economics department (or Economics 2) and class two in the business administration department (or Business 2) as the control classes. We test the students before the teaching experiment. The independent sample *t*-test shows that the average score of each of the experimental classes and that of each of the control classes has not reached significant difference ( $p = 0.411 > 0.05$  and  $p = 0.386 > 0.05$  respectively). See Table 1:

Table 1. Before-Experiment Test Scores

Class	N	Mean	Std. Deviation	Sig.(2-tailed)
Economics 1	50	65.0600	10.65011	P = 0.411 > 0.05
Economics 2	49	66.9184	11.75015	
Business 1	54	64.8704	11.79950	P = 0.386 > 0.05
Business 2	54	62.6759	14.26374	

## *2.2 Teaching Experiment*

The two parallel classes in the economics department are taught by one teacher and the two parallel classes in the business administration department by another. Formative assessment is used in the two experimental classes, and summative assessment, in the two control classes. After a semester's teaching experiment, the final examination scores of the students are used to examine the students' English proficiency and to check whether there is any significant difference between the experimental classes and the control classes.

In addition, to answer the research questions 2 and 3, the students are required to fill out a questionnaire at the end of the experimental teaching. All the questions in the questionnaire are open-ended, aiming to collect the information about the improvement in learning strategies and motivations in the control classes and to collect not only the information about the improvement in learning strategies and motivations but also the opinions of the subjects on the teaching experiment in the experimental classes. The data collected will be processed using statistical software SPSS 17.0.

The tools for formative assessment include learning portfolios, classroom observations, after class exercises and quizzes, etc. Learning portfolios are required to be handed in once a week, the content including a summary of a week's study, a plan for the next week and a suggestion to the teacher. After marking, the teacher should return the portfolios back to the students. Classroom observations will be conducted by the teacher. He needs to record timely the students' attendance, their participation in class activities, and their problems and difficulties in classroom and communicate with the students in time for the problems he observed in class. In addition to the regular homework, the teacher requires the students to recite a certain amount of reading materials or texts in the semester of experiment. The reciting materials can be selected by the students themselves or by the teacher. The reason why reciting is emphasized is that on the one hand reciting as an important English learning strategy has always been valued by researchers and, on the other hand, the teacher can directly observe the students' effort in learning from their reciting frequencies and quality. Quizzes are carried out after the teaching of each unit to examine students' mastery of the course content.

## **3. Results**

### *3.1 Result of Teaching Experiment*

The independent sample *t*-test to the final examination scores shows that there is significant difference between the experimental class and the control class in the economics department ( $p = 0.041 < 0.05$ ), but there is no significant difference between the experimental class and the control class in the business administration department ( $p = 0.514 > 0.05$ ), indicating that after a semester's teaching experiment, formative assessment can improve the English proficiency of the students from the 211 Project university more effectively than summative assessment, but fails to improve the English proficiency of the students from the independent college. See Table 2:

Table 2. After Experiment Test Statistics of the Four Classes

Class	N	Mean	Std. Deviation	Sig. (2-tailed)
Economics 1	50	67.8100	1070185	P = 0.041 < 0.05
Economics 2	49	63.0306	12.27014	
Business 1	54	61.1204	11.76818	P = 0.514 > 0.05
Business 2	54	62.6019	11.75351	

### 3.2 Questionnaire Survey

#### 3.2.1 For Learning Strategies and Motivations

The questionnaire survey aims to investigate if there is any improvement in learning strategies and motivations in the experimental and control classes. See Tables 3 and 4:

Table 3. Questionnaire Survey for Learning Strategies and Motivations before Experiment

	N	Mean	Std. Deviation	Minimum	Maximum
Economics 1	50	3.7858	.45313	3.00	4.65
Economics 2	49	3.7520	.50378	2.95	4.50
Business 1	54	2.9111	.26648	2.40	3.60
Business 2	54	2.8279	.31059	2.25	3.40
Total	207	3.3573	.60205	2.25	4.65

Table 4. Multiple Comparison Test before Experiment

Class	Class	Mean Difference (I-J)	Sig.	95% Confidence Interval	
				Lower Bound	Upper Bound
Economics 1	Economics 2	.0339	.980	-.1891	.2569
	Business 1	.8747*	.000	.6443	1.1052
	Business 2	.9579*	.000	.7246	1.1913
Economics 2	Economics 1	-.0339	.980	-.2569	.1891
	Business 1	.8408*	.000	.6083	1.0734
	Business 2	.0832*	.000	.6886	1.1595
Business 1	Economics 1	-.8747*	.000	-1.1052	-.6443
	Economics 2	-.8408*	.000	-1.0734	-.6083
	Business 2	.0832	.816	-.1593	.3257
Business 2	Economics 1	-.9579*	.000	-1.1913	-.7246
	Economics 2	-.9241*	.000	-1.1595	-.6886
	Business 1	-.0832	.816	-.3257	.1593

\* Significance at 0.05 level.

As can be seen from Table 4, there is no significant difference between the two classes in the economics department ( $p = 0.980 > 0.05$ ) or between the two classes in the business administration department ( $p = 0.816 > 0.05$ ). However, there is significance difference between the two classes in the economics department and the two classes in the business administration department ( $p = 0.000$ ). This indicates that the learning strategies and motivations of the students from the 211 Project universities are significantly different from those of the students from the independent colleges.

After a semester's teaching experiment, we use the same questionnaire to investigate the students' learning strategies and motivations. See Tables 5 and 6:

Table 5. Questionnaire Survey for Learning Strategies and Motivations after Experiment

	N	Mean	Std. Deviation	Minimum	Maximum
Economics 1	50	3.9896	.42283	3.25	4.70
Economics 2	49	3.7755	.41622	3.00	4.70
Business 1	54	3.0000	.25000	2.45	3.50
Business 2	54	2.9547	.29597	2.30	3.55
Total	207	3.4690	.58255	2.30	4.70

Table 6. Multiple Comparison Test after Experiment

	Class	Mean Difference (I-J)	Sig.	95% Confidence Interval	
				Lower Bound	Upper Bound
Economics 1	Economics 2	.2141*	.029	-.0148	.4134
	Business 1	.9896*	.000	.7837	1.1956
	Business 2	1.0350*	.000	.8264	1.2435
Economics 2	Economics 1	-.2141*	.029	-.4134	-.0148
	Business 1	.7755*	.000	.5677	.9833
	Business 2	.8208*	.000	.6105	1.0312
Business 1	Economics 1	-.9896*	.000	-1.1956	-.7837
	Economics 2	-.7755*	.000	-.9833	-.5677
	Business 2	.0453	.951	-.1713	.2620
Business 2	Economics 1	-1.0350*	.000	-1.2435	-.8264
	Economics 2	-.8208*	.000	-1.0312	-.6105
	Business 1	-.0453	.951	-.2620	.1713

\* Significance at 0.05 level

As can be seen from Table 6, after the experiment teaching, there is significant difference between the two classes in the economics department ( $p = 0.029 < 0.05$ ), but there is no significant difference between the two classes in the business administration department ( $p = 0.951 > 0.05$ ). There is also significance difference between the two classes in the economics

department and the two classes in the business administration department ( $p = 0.000$ ).

The questionnaire survey shows that formative assessment is helpful for the improvement of the learning strategies and motivations of those students in the 211 Project universities with good command of English but it does not work for those in the independent colleges with relatively poor command of English.

### 3.2.2 For Formative Assessment

Another questionnaire survey is to investigate the opinions of the two experimental classes on formative assessment. The independent sample  $t$ -test shows that the average scores of the two experimental classes have significant difference ( $p = 0.000$ ), indicating that most students in Economics 1 are in favor of formative assessment, while the majority of students in Business 1 do not appreciate formative assessment. See Table 7:

Table 7. Questionnaire Survey for Formative Assessment

	N	Mean	Std. Deviation	Sig.(2-tailed)
Economics 1	50	2.1951	.35035	
Business 1	54	1.7929	.35778	P = 0.000

It can be seen from Table 7 that the average score of Economics 1 is obviously higher than that of Business 1. Their opinions on formative assessment reflect their English learning strategies and motivations. The students with good command of English have good motivations and are good at using learning strategies. They have clear objectives and strong autonomous learning abilities. However, those students with poor command of English have not reached the level of autonomous learning. Most of them have not proper learning motivations or clear objectivities.

## 4. Discussion

Through reflections on the process of the teaching experiment and analyses of the questionnaires for the learning strategies and motivations and for the formative assessment, we find three factors affecting the implementation of formative assessment: the students' learning motivations, the assessment quality and the teacher's professional ability.

### 4.1 Learning Motivation

Research shows that the vast majority of students in the experimental class of the business administration department of the independent college are not in favor of formative assessment. This is largely because they lack relevant learning motivations. Although these students have passed their college English proficiency test, they are still at a basically lower level of English. Most of them are not interested in English learning at all, and they are obliged to study English for their qualification of graduation. Now that their graduation is now eligible, their motivations to learn English have disappeared. So formative assessment which functions to fully monitor their English learning process will only result in their

increasing resentment and resistance to English study. Taking reciting as an example, of the 54 students in the experimental class, nine have never completely recited one text. The reason lies in that they are not interested in English language at all or they believe that punishment will not go to the public.

The questionnaire survey shows that those students in favor of formative assessment share a common characteristic. Most of them prefer to learn English well, hoping to have greater success in English learning and hoping that English will be helpful for their future employment. Some students show a positive attitude towards the formative assessment because they want to go to study abroad. As for these students, they kept their learning portfolios specific and detailed with rich learning content, and they recited more reading materials than other students. It is these students who believe that they have improved their learning strategies and motivations.

We can hereby conclude that learning strategies and motivations have an obvious threshold effect on the implementation of formative assessment. Those having higher learning motivations tend to accept formative assessment which in turn helps improve their learning strategies and motivations. On the contrary, those having lower learning motivations tend to reject formative assessment, resulting finally in its ineffectiveness.

#### *4.2 Quality of Assessment*

Validity and reliability are two important parameters of quality evaluation. As is pointed out, the validity of formative assessment depends on the extent to which the teacher clearly understands the difficulties encountered by the students. The evaluation tools for the formative assessment used in this experiment can be helpful to accurately find students' problems in learning strategies and attitudes. Therefore, the validity of formative assessment is desirable.

The problem in formative assessment is the difficulty to ensure its reliability. Sometimes, the teacher will give extra scores to those students who have made greater progress in recitation in order to encourage them, other students may complain that the scores offered by the teacher may be too exaggerated for their actual performance. This kind of balance between relative progress and absolute performance is often difficult to control for the teacher. In addition, many items of the formative assessment themselves are not easy to be quantified. If they are really needed to be quantified for the management of teaching, it is often based on the subjective judgment of the teacher. The non-assessed factors such as the personal preferences of the teacher to the students may affect the reliability of evaluation.

In fact, because of the objects and functions of the assessment, formative assessment focuses mainly on the qualitative information about the students, rather than the quantitative assessment. Therefore, the question for the future researchers is how to better adapt formative assessment to the existing evaluation system.

#### *4.3 Teacher's Professional Ability*

During the experiment process, the teacher did find many problems in teaching and in the



students from the assessment tools, but he himself is often confused with the problems. For example, some students may have dyslexia, some students may have no interest in learning, some students may even have conduct disorders, and some students cannot find effective learning strategies. These kinds of problems are related not only to language teaching but also to psychology and education, etc. To solve these problems is by no means an easy task. If we say the teacher is playing a judge in the summative assessment, then he is playing a doctor in formative assessment. A doctor has to be able to not only diagnose the disease but also prescribe the right medicine. Obviously, formative assessment requires the teacher to have more professional knowledge and higher ability of teaching.

## 5. Conclusion

It is popularly accepted that formative assessment plays an important role in college English teaching in China and it has proved to function well in the practice of teaching. However, to what an extent that it works better than summative assessment is worth reflection. Based on the empirical research on the two parallel classes of students in a 211 Project university and two parallel classes of students in the independent college of the same university, we can draw the following conclusions. (1) Formative assessment is helpful for the teacher to collect feedback information timely, to monitor their students' progress and to modify the instruction accordingly, and it is suitable for the students with higher learning strategies and motivations. (2) Formative assessment requires the teacher to have more professional knowledge of the English language and higher ability of teaching. (3) Formative assessment is difficult to guarantee the reliability in terms of quality, so it is not universally applicable and will not completely replace summative assessment. The questionnaire survey shows that learning strategies and motivations have an obvious threshold effect on the implementation of formative assessment.

## References

- Ainsworth, L., & Viegut, D. (2006). *Common Formative Assessments*. Thousand Oaks, CA: Corwin Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Black, p., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, evaluation and Accountability*, 21, 5-31. <http://dx.doi.org/10.1007/s11092-008-9068-5>
- Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74. <http://dx.doi.org/10.1080/0969595980050102>
- Crooks, T. (2001). The validity of formative assessments. *British Educational Research Association Annual Conference*, University of Leeds, September 13–15, 2001.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 90-91. <http://dx.doi.org/10.3102/003465430298487>

Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation*, 38, 21-27. <http://dx.doi.org/10.1016/j.stueduc.2012.04.001>

Henson, K. T., & Eller, B. F. (1999). *Educational Psychology for Effective Teaching*. Cambridge: Wadsworth Publishing.

Huhta, A. (2010). Diagnostic and formative assessment. In B. Spolsky and F. M. Hult (Eds), *The Handbook of Educational Linguistics* (pp. 469-482). Oxford: Blackwell.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: Historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254-284. <http://dx.doi.org/10.1037/0033-2909.119.2.254>

Shepard, L. A. (2005). Formative assessment: Caveat emptor. ETS Invitational Conference *The Future of Assessment: Shaping Teaching and Learning*, New York, October 10-11, 2005. Retrieved 25 August, 2011.

Shute, V. J. (2007). *Focus on formative feedback* (Rep. No. RR-07-11). Princeton, NJ: Educational Testing Service.

Simpson, M. (1999). Diagnostic and formative assessment in the Scottish classroom. In T. G. K. Bryce & W. M. Humes (Eds), *Scottish Education*. Edinburgh: Edinburgh University Press.

William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37, 3-14. <http://dx.doi.org/10.1016/j.stueduc.2011.03.001>

### **Copyright Disclaimer**

Copyright reserved by the authors.

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).