

Formative Assessment in an EFL Composition Course: Action Research

EFL 英作文クラスにおける形成的評価の効用 － 実地研究を通して －

Jeffrey DAVIS

要 約

本論文は、EFL 大学作文授業における形成的評価 (Formative Assessment) の効用について探求するものである。最近の文献の分析によって、形成的評価が理解の深度を助け、指導者・学習者間対話を改善する可能性が証明されている。4ヶ月にわたって実施した実地研究が理論と実践の橋渡しの役目を果たした。パフォーマンス測定指標、コメントのみによる評価と学習者同士の評価の利用効果の検証がこの論文の議論の焦点である。形成的評価の導入にともなう困難の調査も実施している。指導者・学習者双方にとっての学習効果の向上が EFL 大学作文授業における形成的評価法の採用の有効性を立証している。

Abstract

This paper explores the use of formative assessment in an EFL university composition course. An analysis of the current literature provides evidence that formative assessment promotes deep understanding and has the potential to improve instructor-learner dialogue. Action research conducted by the author over a four-month period contributes to a bridge between theory and practice. The use of performance standards, comment-only marking, and peer-assessment is the focus of this discussion. An examination of the difficulties surrounding the implementation of formative assessment is also included. The learning outcomes for both the instructor and the learners, renders insight into the utility of employing formative assessment strategies in an EFL university composition course.

Fullan (2003) noted that: “the leader’s job is to introduce new elements into the situation that are bound to influence behavior for the better” (p. 1). To improve learning within an organization, learning experiences must change from a receptacle orientation to one that embraces the most current learning practices; which makes a conscious effort to develop student voice and promotes meta-cognition to improve the depth of learning. Formative assessment may be a means to achieve this goal. This paper explores improved assessment practices in a university EFL composition class and the results for student learning. A brief examination of the two major forms of assessment, formative and summative, frames the discussion.

Summative assessment

Assessment of learning or summative assessment is based on formal testing at the end of a unit, term, or school year. Earl (2003) discusses this type of assessment and reminds us that students are compared with one another and that the teacher is the key assessor. This is the primary form of assessment that permeates Japanese education and many other education systems around the world. The Japanese university entrance exam is indicative of this heavy focus on testing. Rohlen's (1983) quote succinctly summarizes the importance of the university entrance exam: "Twelve years of schooling culminate in this moment, which is a crucial turning point in the life cycle of most Japanese" (p. 77). The university entrance exam is the ultimate summative assessment tool and throughout senior high school, Japanese learners are constantly being tested, compared, and sorted into perceived ability groups. As Bendict (1946) notes, hierarchical order is a tangible aspect of Japanese society. Summative testing serves as an effective method to place Japanese students into their prescribed position in life.

A summative assessment system does allow for extensive use of learning data to report on student achievement. Earl (2003) also notes that summative testing is a useful tool for stakeholders to examine the overall education program and the learning outcomes. However, students may lose a measure of control over their own learning and fail to develop deep understanding due to their focus on the all important summative assessments.

The most common rationale for the use of summative assessment is that students thrive in a competitive testing environment because exams serve as external motivation. Unfortunately, as Stiggins (2004) observed, learners who have a poor academic record often give up when confronted with high-pressure summative assessment tests. Black & Wiliam (1998) noted that the improved use of formative assessment raises the achievement of lower-performing students. Moreover, an educator is truly effective when they can narrow the gap between the low-performing and high-performing students (granted that the high-performers remain constant). This narrowing changes classroom dynamics and causes improvement in the overall learning community. Lower-performing students may feel empowered and can experience a philosophical sea-change in regard to their learning.

Stiggins' (2004) notion of balanced assessment procedures that utilize both assessment for learning and assessment as learning may be applicable to university EFL courses in Japan. The university system has more flexibility to implement balanced assessment procedures because the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) does not exert undue pressure on the tertiary system to rank and categorize students. Universities have an opportunity to take the lead in exploring alternative assessment options.

Formative assessment

English as a foreign language (EFL) composition courses at the tertiary level present an excellent opportunity for the implementation of formative assessment. EFL classes are not bound by a lecture-style approach - students are expected to perform and produce in each class. Effective diagnostic feedback, the use of performance standards, and a focused teacher-learner dialogue may then emerge as intrinsic parts of EFL language classes. Improved teacher learner dialogue will, as Black & Wiliam (1998) assert, help learners close the gap between what they know and what they need to know.

The foundational ideas of Earl, Black et al., and Stiggins were guiding forces in the planning and implementation of the action research. Earl's (2003) assessment paradigm: assessment of learning, assessment for learning, and assessment as learning provided a useful framework to examine formative assessment. The paradigm clearly explains the differences between formative and summative assessment and provides a base to evaluate one's own assessment practices.

Action research - the inquiry

My action research addressed two questions for inquiry. Firstly, to what extent will the implementation of improved formative assessment improve learner performance in an English composition class? Secondly, to what extent will the implementation of improved formative assessment develop focused instructor-learner dialogue and the development of student voice? The participants in my inquiry were two first-year EFL composition classes. Class *e2* contained twenty-six students, while *e3* contained twenty students. The inquiry was conducted during the second semester (fall) of the 2005-2006 year. The students received one ninety-minute lesson a week over 14 weeks from October through the end of January and a fifty-minute final examination during the first week in February.

The inquiry contained three assessment cycles: the first cycle was six weeks in duration, while the second and third cycles were both four weeks in duration. Assessment cycle one focused mainly on creating a performance standard (PS) and changing learner perceptions with regard to English composition. The first two weeks were devoted to introducing the learners to the concept of a performance standard (PS) (see Appendix A and B). In the following four weeks, peer-assessment, traffic-light marking, and the conventions aspect of the PS were introduced. Assessment cycle two was devoted to introducing the PS aspects of *form* and *style*, while further refining peer-assessment skills. Assessment cycle three was based upon introducing the *meaning* domain and conducting teacher-learner interviews where feedback was given and learners were expected to respond to a question about their learning.

Assessment cycle one was the most difficult component of the inquiry, as introducing the assessment focus areas presented significant philosophical challenges for the learners. The general focus of assessment cycle two was reinforcing class routines surrounding peer assessment and the use of a traffic light marking system. This period of the inquiry was an exciting one, as each class offered new insights into how to integrate formative assessment into everyday classroom activities. Assessment cycle three focused on concluding the inquiry and learning about student perceptions of their learning. One on one interviews, conducted between the students and the teacher, at the end of the term were an informal self-assessment designed to check for understanding of the course concepts and to deepen the instructor-learner dialogue.

Several significant discoveries occurred during the course of this inquiry. During assessment cycle one, I came to a realization that changing learner philosophy was the most challenging aspect of the inquiry. During the first three weeks of the inquiry, it was apparent that students hoped I would revert to a more traditional method. The class development of the performance standards and the initial experiment with peer-assessment reflected this comment as I experienced several collective groans and surprised comments. The accumulated educational experiences of the learners produced expectations regarding how a teacher should conduct

classroom activities.

The constant cognitive dissonance in the first three weeks was a strain on the learners. Once learners overcame their confusion and understood the motivations behind the method, the inquiry gathered steam. Assessment cycle two was perhaps the most rewarding time of the inquiry. Most learners had accepted formative assessment and an efficient pattern for peer-assessment was introduced. This utilized traffic-light marking in combination with the performance standard. In assessment cycle three, the value of instructor-learner dialogue was readily apparent. Informal in-class dialogue had been conducted throughout the term, but the final interview (in my office) was a tremendous learning experience for myself, if not the learners as well. Speaking to the students about their learning in English was fascinating. The class language fostered by the PS was instrumental in the achievement of this goal.

Successful strategies

The performance standard was an invaluable tool that enabled learners to develop a deeper understanding of the course concepts. Halbert & Kaser's (2003) pattern, where educators first share criteria with learners and then have the students construct new criteria in order to compare with the teachers' criteria was a useful starting point. In groups of four, learners were instructed to paraphrase the PS in an effort to develop deeper understanding and ownership of the assessment process. When this process was completed, I synthesized all the student work into a class-developed performance standard. This procedure was instrumental in helping the students understand the teacher language and in laying out the main course concepts for the semester (Halbert & Kaser, 2003). Learners could then refer to the PS to overcome their cognitive dissonance, and thus improve their English writing.

The use of traffic-light marking was an effective way to integrate the PS into daily classroom activities. Each performance category was given a color. *Not yet meeting expectations* was assigned a red light, *meets expectations* was assigned a yellow light, *fully meets expectations* a blue light, and *exceeds expectations* a green light. The course used this system (peer-assessment and instructor assessment) to identify where the learners fit on the performance standard. During peer-assessment, learners naturally began to deconstruct the criteria (PS) to identify varying skill levels within a given aspect. This was especially helpful for the *conventions* aspect, which contains several sub-aspects including: complete sentences, spelling, capitals, punctuation, grammar, and usage. Thus, during peer-assessment, specifically during the *convention* and *form* aspects, students were as able to help each construct meaning and use their accumulated knowledge – both vital to learning as Bruner (1999) asserts. Initially, learners were reluctant to share and engage in peer-assessment, but this changed as they began to recognize the importance of reciprocal learning.

Instructor-learner interviews provided a chance for students to see how they had progressed in their first year of English composition. This was an overt attempt to develop student voice and to determine whether the learners understood the course concepts as introduced in the PS. It was apparent that learners enjoyed this process. During the interview learners seemed quite nervous, however, most regained their composure when they realized the purpose was not to test their knowledge, but to discuss their learning. I received several positive comments from students regarding the interview, which signaled its importance within the inquiry context.

General challenges

Black, Harrison, Lee, Marshal, & Wiliam, (2003) assert that student proficiency and learner exposure to the target language are vital considerations when using formative assessment in modern foreign languages. In this inquiry, it was difficult to determine whether learner English proficiency constrained meaningful meta-cognition. Although class instruction was conducted in English, students were encouraged to discuss course concepts in their native language or English. The PS was useful; however, the *style* and *meaning* aspects were difficult concepts to introduce given the proficiency level of the students. *Meaning* was expressed as the level of detail and content and *style* was expressed as the level of sentence complexity.

Learner peer-assessment presented some initial challenges, as students understood the concept, but were often unable to act as a critical friend (Black et al., 2003). In most cases, students erred on the side of praise and gave their peers favorable comments. Initially, comment-only marking puzzled the students as they did not receive a score, but a traffic-light reference to the PS coupled with a focused comment. It took considerable time for students to understand the motivations behind comment-only marking, even though it was explained each class. Minimal teaching time was also a pressing issue. Meeting the students once a week for ninety minutes stifled the development of a class culture that understood and accepted formative assessment. Moreover, starting the inquiry in the middle of the academic year (September) made it difficult for learners to adjust to the new class format.

Assessment focus areas

Black et al. (2003) and their focused inquiry with formative assessment (KMAP project) was an invaluable resource in the development of this inquiry. As Black et al. (2003) noted, the assessment focus areas provided cogent examples of effective formative assessment in practice and theory. I followed Black et al.'s (2003) suggestion and implemented the focus areas incrementally by focusing on two areas at a time. I was therefore able to manage the implementation process more effectively. Initially, I focused on shared criteria and comment-only marking and as the inquiry progressed, peer assessment was integrated. The traffic-light marking concept was also borrowed from the KMAP project (Black et al., 2003). This powerful analogy was an effective meaning-maker for learners. In addition, the KMAP project allowed me to anticipate potential difficulties with my inquiry, including subject area concerns and learner resistance (Black et al., 2003). A brief analysis of the focus areas will explain their relevance to the inquiry.

During distribution of marked tests or assignments, learners often focus on the final letter grade or percentage given - this reduces the depth of learning. As Black & Wiliam, (1998) suggest, learners must be engaged in trying to understand how they can improve the gap between what they know and what they need to know based on their most recent performance. It is the educator's challenge to maintain intellectual tension in the classroom when providing feedback. Comment-only marking and shared criteria may achieve this goal by pushing the learners to carefully examine their work and engage in an ongoing learning dialogue with the instructor. In an EFL setting, this dialogue is vital to improving communication between the instructor, who speaks the target language, and the learner, who is struggling to develop proficiency.

Effective use of comment-only marking and shared criteria allows learners and their instructors to engage

in focused discussion about a particular aspect of learning. This discussion is based on a focused comment or aspect of a performance standard. In an EFL class, this discussion in the target language produces a synergistic learning experience as one specific learning issue may be addressed orally in the target language and meta-cognitively in the first language. Black et al. (2003) assert that in an EFL context, an aid to the implementation of comment-only making is a feedback lexicon, which denotes marking symbols and common language used during assessment.

My intention was to implement peer-assessment and self-assessment, yet due to time constraints, self-assessment was not dealt with explicitly. Black et al. (2003) and their suggestion to implement peer-assessment as a precursor to self-assessment was helpful as learners need peer-assessment to help them become proficient in self-assessment. As Black et al. (2003) observed, peer-assessment motivates students to work carefully because a peer will be assessing their work (Black et al., 2003). During the action research, the PS used in conjunction with peer-assessment, allowed for a common class language to emerge.

Learner & instructor impact

Claxton (2003) explains that “when learners are encouraged to talk more about the process of learning, their achievements improve within a term” (p. 2). The data in Appendix A demonstrates that students were able to improve their achievement in one semester. The results were congruent with my expectation that Japanese learners would respond to improved assessment practice. More specifically, learners were able to develop proficiency in the use of English *conventions* and basic *form*. The use of peer-assessment produced significant changes in overall classroom dynamics and the development of student voice. Students became accustomed to a pattern where a composition would be assigned and the following week a student in the class would assess the work based on one aspect of the performance standard. The color-coding of this procedure using the analogy of traffic lights became firmly entrenched. Our in-class discussions on peer-assessment and how to effectively assess someone were ongoing and some student feedback was extremely valuable to the learning process. The impact on my teaching style was profound. I took more time assessing assignments, but checked fewer. I regularly engaged in focused dialogue regarding specific aspects of English writing as we had a common base to work from – the performance standards.

My initial inquiry has now served as a base for a continued school improvement plan, which commenced April 1st, 2006. This academic year, I will implement improved formative assessment in all of my composition classes using the framework developed in October of 2005. In addition, instructional issues surrounding the meaning and style aspects contained in the PS will be explored. The full-year version of this inquiry will allow ample time for self-assessment to be used extensively.

References

- Black, P. & Wiliam, D. (1998). Inside the Black Box: Raising Standards Through Classroom Assessment. *Phi Delta Kappa International*. Retrieved August 1, 2005, from <http://www.pdkintl.org/kappan/kblac9810.htm>
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D., (2003). *Assessment for Learning: Putting it into practice*. New York, NY: Open University Press.
- Benedict, R. (1946). *The Chrysanthemum and the Sword*. New York: Meridian Books.
- Bruner, J. (1999). Culture, Mind, and Education. In B. Moon & K. Coates (Eds.), *Curriculum in Context* (pp. 148-173). Thousand Oaks, CAL: Paul Chapman Publishing.
- Claxton, G. (2003). *Learning to Learn: a key goal in a twenty-first century curriculum*. Qualification and Curriculum Authority. Retrieved April 27th, 2006, from http://www.qca.org.uk/downloads/11469_claxton_learning_to_learn.pdf
- Earl, L. (2003). *Assessment as Learning: Using Classroom Assessment To Maximize Student Learning*. Thousand Oaks, CL: Corwin Press.
- Fullan, M. (2003). *The Moral Imperative of School Leadership*. Thousand Oaks, CL: Corwin Press.
- Fullan, M. (2005). *Leadership and Sustainability*. Thousand Oaks, CL: Corwin Press.
- Government of British Columbia. (2002). *BC Performance Standards*. Victoria: Ministry of Education.
- Halbert, J., & Kaser, L. (2003). Improving Schools from the Classroom Out. *The Network of Performance Based Schools*. Retrieved September 10, 2005, from www.npbs.ca/Articles.htm
- Rohlen, T. (1983). *Japan's High Schools*. Berkeley: University of California Press prepared for presentation at the annual meeting of the American Educational Research Association, San Diego, CA, April, 2004. Retrieved September 1, 2005, from
- Stiggins, R. (2004). New Assessment Beliefs for a New School Mission. *Phi Delta Kappan*. September 2004, p. 22-27

Appendix A

2005-2006

Nagasaki University of Foreign Studies Learning Inquiry

Inquiry focus

1. To what extent will the implementation of improved formative assessment improve learner performance in an English composition class?
2. To what extent will the implementation of improved formative assessment develop focused instructor-learner dialogue and the development of student voice?

Table 1 summarizes the findings of the inquiry.

Table 1 Composition II learning data (e2 and e3).

| | Not Yet | Meets | Fully meets | Exceeds |
|-----------|---------|-------|-------------|---------|
| Jul. 2005 | 2% | 38% | 52% | 7% |
| Jan. 2006 | 7% | 23% | 49% | 21% |

*Note these percentages were based on all course work completed.

The percentages were delineated as follows:

NY=less than 60%

M = 60 to 79

FM= 80 to 90

EE= 91 to 100

Strategies (focus areas)

- Collaborative development of class-based performance standards
- Integration of a traffic-light marking system with the performance standards
- Extensive peer-assessment using traffic-light marking system and performance standards
- Comment-only marking
- One on one interviews

Appendix B

Performance standard – student version

Performance Standards for Writing Assignments: *Composition II*

| Aspect | Not yet within expectations (Red) | Meets expectations (Yellow- basic level) | Fully meets expectations (Blue) | Exceeds Expectations (Green) |
|--|--|--|---|--|
| Snapshot | The composition is short and has many mistakes. The student needs help. | The composition has many simple ideas. Parts may be confusing with many mistakes | The composition has easy to follow ideas with detail and explanation; few errors. | The composition is clear and focused. It is thoughtful and has a strong identity. |
| Meaning ➤ ideas ➤ details | - very short - few details - topic is not clear | - may give different views - may add some opinions - few details and explanations | -some supporting details to make people imagine - more reasons and explanations of how the student felt | - it's written with a sense of individuality and interesting details - clear viewpoint |
| Style ➤ clarity ➤ variety | - makes mistakes choosing words - basic sentences - no variety in the sentences | - writing is very basic - different types of sentences, but little variety | - clear direct language - variety in sentence structure | - uses new words many times - has sentence variety |
| Form ➤ beginning ➤ organization ➤ transitions ➤ ending | - confusing beginning - ideas unrelated to topic - no topic sentence - no closing sentence | - can generally understand the topic - the body is often a list of events that are not related - often has no ending | - introduces the topic effectively - uses a variety of connecting words - (conjunctions and transitions) - ending comes suddenly | - excellent structure - ideas match the topic - good use of connecting words - has a great conclusion |
| Conventions ➤ complete sentences ➤ spelling ➤ capitals ➤ punctuation ➤ grammar/ usage | - uses and, so on, but many times - it is hard to understand - grammar and syntax problems - sentences are too long or not complete | - some mistakes - only simple sentences are correct - some incomplete or run-on sentences | - some mistakes, but these do not affect meaning - most sentences have good structure | - few mistakes - takes risks - complete sentences |