

Formative Assessment for Promoting Intrinsic Motivation in an EAP Reading Comprehension Course

ANA LAURA ESCOBAR LEIVA

Centro de Idiomas
Universidad Estatal a Distancia, Costa Rica

LAUREEN ARIAS DURÁN

Siemens, Costa Rica

MÓNICA JIMÉNEZ MURILLO

Sede de Liberia
Universidad de Costa Rica

Abstract

This study aims to determine the extent to which formative assessment influences performance in summative assessment and student attitudes towards reading academic texts in English. In a sample of 9 psychology major students at a public university in San José, Costa Rica, the study gathered data pertaining to learner performance in both formative and summative assessment as well as self-reported attitudes and perceptions in questionnaires administered after assessment. The results have revealed that formative assessment has positively influenced student performance in summative assessment, and that in turn it has promoted a more enthusiastic attitude towards reading academic texts in English.

Key words: formative assessment, construct validity, content validity, intrinsic motivation, summative assessment

Resumen

Esta investigación busca determinar si la evaluación formativa influye en el rendimiento de los estudiantes durante la evaluación sumativa y en sus actitudes hacia la lectura de textos académicos en inglés. El estudio recopila datos del rendimiento de nueve estudiantes de psicología de una universidad pública en San José, Costa Rica, en ambos tipos de evaluación, formativa y sumativa. Los datos recabados en cuestionarios autoevaluativos sobre actitudes y percepción de rendimiento, administrados después de las valoraciones, revelan una influencia positiva de la evaluación formativa y en el desempeño de los estudiantes en la evaluación sumativa, lo que fomenta una actitud más entusiasta hacia la lectura de textos académicos en inglés.

Palabras claves: evaluación formativa, validez de construcción, validez de contenido, motivación intrínseca, evaluación sumativa

Introduction

Formative assessment has been proposed as a skill-building technique focused on appropriate and constructive guidance and monitoring of student performance. Extensive literature on this assessment practice has prompted research on its efficacy and contribution to the field of applied linguistics. However, the construct of formative assessment is still under development since some of the studies conducted raise questions towards the specificity of formative assessment practices and reliability of results regarding the diversity of populations available for sampling.

In the case of English for academic purposes, research on the efficacy of formative assessment requires creating instruments that produce valid and reliable results regarding performance closely related to students' target context. Therefore, the present course evaluation project intends to determine how formative assessment practices influence the accomplishment of

course goals and student motivation towards learning English.

Literature Review

Assessment has been and continues to be an intrinsic component in education, and language teaching is not the exception. Evaluation gives both teachers and students significant insight in terms of how successfully course goals are achieved in any educational program or institution. However, assessment is a complex and sensitive endeavor. Its complexity involves how closely assessment techniques reflect instruction and how well tests and graded assignments are designed. Its sensitive nature concerns how student performance and motivation may be affected by their administration. Validity and reliability are the pillars of effective assessment, which when disregarded may render a course pedagogically deficient and consequently demotivating for learners.

The Predominant Role of Testing

The “transmission model of instruction” (O’Malley & Valdez, 1996, p. 10), which has dominated the education realm throughout the years and consists of a vertical and teacher-centered approach of transferring knowledge to students, has used traditional testing as the sole instrument for learner assessment. Teachers trained under the transmission model use tests measure student achievement and usually make no conscious effort to reflect on prior instruction or learners’ target needs. Traditional summative language testing is mostly based on the discrete-point approach that emerged between 1970 and 1980, which claims that “language can be broken down into its component parts and that those parts can be tested successfully” (Brown, 2004, p. 8).

In addition, testing plays a very important affective role in educational programs, often overlooked yet responsible for leading students to success and motivation beyond the classroom. Its overlooking by educators has made testing a very unpleasant ordeal, creating anxiety and sometimes a predisposition to failure that for some students translates into poor performance.

Contrary to the general views by educational authorities in many countries that tests “raise standards” among student populations, research indicates that testing “not only inhibits the practice of formative assessment but has a negative impact on motivation for learning” (Harlen & Deakin, 2003, p. 170), markedly and progressively separating high from low achievers. Moreover, many teachers gear courses to prepare students for tests rather than to lead them towards forming their own

“learning identity” and explore their whole potential, leaving no room to implement formative assessment (Harlen & Deakin, 2003, p. 174).

Although complex and highly variable from person to person, motivation for learning is what leads students to success. However, in a world dominated by the summative fixed judgments on ability that grades place on students since elementary school, student motivation is mostly triggered by external factors such as approval from others based on numerical results instead of being sparked by the interest of exploring one’s potential. One of the findings related to summative assessment effects on motivation claims that testing affects students personally, increasing motivation in some and discouraging others on the basis of performance goals only, not learning goals (Harlen & Deakin, 2003). Performance goals have to do with developing test-taking skills with a specific grade in mind, not with critical thinking that would enable deep analysis and mastery of major skills and constructs.

Crooks (as cited in Harlen & Deakin, 2003) revealed evidence confirming the shallow learning phenomenon promoted by summative assessment and recommends that in order to counteract it teachers should invest their lessons in encouraging motivation and self-regulated learning, which consists of training students to use self-reflection, self-monitoring and organization skills to take control of their own learning processes. According to Ryan and Deci (2000), ideally, a balance between both intrinsic and extrinsic motivation would reflect the demands of a primarily summative mentality still prevalent in education more

realistically. However, despite its misguided and disproportioned prevalence in education, testing is still important as it reveals both quantitative and qualitative indicatives of learner performance necessary in any educational program.

The Importance of Formative Assessment

Constructivism, one of the theories in language instruction and acquisition, views learners as active participants in the construction of meaning by taking part in “dynamic” cognitive processes that relate input to schemata (O’Malley & Valdez, 1996, p. 10). Proponents of constructivism propose formative assessment as the “delivery (by the teacher) and internalization (by the student) of appropriate feedback on performance, with an eye toward the future continuation (or formation) of learning” (Brown, 2004, p. 6).

Formative assessment occurs in the classroom and is delivered informally in a number of ways such as verbal comments in error-correction, reminders of procedures or strategies learned in class and written comments on the board, homework or class material samples. It provides a safe place for students to make mistakes and focus on the process of learning by monitoring their progress without the pressure of grades or “fixed judgments” by the teacher on their performance (Brown, 2004, p. 5). Alvarez et al. consider formative assessment a “mirror” in which both students and instructors see their learning and instruction reflected.

The integration of formative and summative assessment based on providing continuous and appropriate

feedback promotes positive washback, which refers to how instruction and learning are affected during and after a course by the “information that ‘washes back’ to students [for diagnosing] strengths and weaknesses” (Brown, 2004, p. 29). Positive washback creates the ideal conditions for developing “intrinsic motivation, autonomy, self-confidence,” and other cognitive factors necessary for language acquisition.

As for research regarding the efficacy of formative assessment in EAP courses, important studies were found. One study addressed the positive results obtained by teachers when using virtual discussion boards to implement formative assessment with college freshmen in order to build the necessary academic writing skills for college courses (Horstmannshof & Brownie, 2013). Another study conducted in different university courses in the United States concluded that formative assessment is an essential tool to enhance teaching practices and learning opportunities, regardless of the subject matter (Stull, Varnum, Ducette & Schiller, 2011). In yet another study conducted in public and private Brazilian university physics courses, the implementation of formative assessment improved learning yet not grades (Cruz, Dias & Kortemeyer, 2011).

Research on the impact of formative assessment in education has been also geared toward elementary and high schooling in the form of meta-analysis reviews of recent relevant research. Some reviews conclude that different studies demonstrate improvement in performance when using formative assessment in courses. However, an experimental study conducted with middle school students

determined that formative assessment had no impact on learning outcomes or motivation (Yin, 2005). In addition, Bennet (2009) contends that the validity of much of the evidence is debatable because some of the reviews summarize results of very diverse formative assessment techniques in one meta-analysis, which compromises their validity as well as reliability. In addition, Bennet claims that since any type of feedback can be considered formative assessment, a much specific definition of the construct must be established, so that any results obtained from research on that particular technique can more clearly contribute to the literature on this field.

Formative assessment has emerged as the panacea of education, claiming to be the key to motivation for learning and internalization of input. Therefore, using results from both formative and summative assessment, this study intends to determine the extent to which course goals were achieved in a course of English for Academic Purposes for psychology students at a public university in San José, Costa Rica. Regarding motivation, the study focused on analyzing whether formative assessment influenced students' attitudes towards reading academic texts in English.

Method

The present research study was conducted using a qualitative method to gather data reflecting the language competence of learners when faced with specific language tasks as well as self-reported perceptions of competence and attitudes toward reading academic texts in English.

The population for this study consisted of nine psychology students from a public university in Costa Rica whose proficiency level ranged from true beginner to advanced, and they were coursed the first, second and third academic year in the psychology major while taking the English course. Participants' most immediate need was reading comprehension of psychology academic texts as well as vocabulary competence to cope with reading tasks. However, this population showed resistance towards reading in English since most of them reported negative attitudes towards the language.

Procedure

Three different types of data collection instruments were administered during this study. Two of them aimed at gathering information regarding students' performance in formative assessment through review tasks, and in summative assessment through quizzes. Both performance instruments consisted of three sections: section one pertaining to vocabulary, section two requiring text analysis at sentence or paragraph levels, and section three addressing text comprehension. The third instrument, a questionnaire, intended to collect data about students' perceptions regarding performance and any changes in attitudes towards English.

Course assessment was implemented after 4 to 5 weeks of strategy training at the end of Units 1, 2 and 3. Training consisted of guessing meaning from context for vocabulary recognition, several reading comprehension strategies and summarization in Spanish, which are all skills the study

subjects would need to master in their target context. A review task was administered during the last week of instruction for each course unit, allowing one week between the review task and the quiz for students to prepare for summative evaluation. Students were allotted 30 minutes to complete both formative and summative assessment instruments, followed by the self-reporting questionnaire, for which students had 10 minutes to complete.

Results and Discussion

To determine the extent of goal achievement during the ESP course, a comparison of student performance in formative and summative assessment for Units 1, 2 and 3 is analyzed. Then, the influence of formative assessment on students' attitudes towards English is appraised from the data collected in the self-report questionnaires.

Results of Performance in Formative and Summative Assessment

Unit 1 focused on comprehension and restatement of salient information from academic psychology texts. Results of the vocabulary section suggest the review task favored students' performance in the quiz due to familiarization with item design and task demands. Feedback provided in the review task clearly helped students to identify clues in the context of the vocabulary section of the quiz. Data collected for the second section, consisting of a timed skimming task, revealed that most students who completed both assessment instruments provided an

accurate and complete account of the most crucial information of the text in their answer, but not much difference was observed in performance from formative to summative assessment. In the third section, dealing with identifying and restating salient information in Spanish, the results show considerable improvement for most students in summative assessment after completing the review task. Five students who took the summative quiz obtained a good result and two of them had an excellent performance on this quiz. Only one failed the quiz, and another had a fair result.

Student performance during the diagnostic test prior to the course was only compared to the assessment implemented in Unit 1 because it focused on comprehension of salient information. However, the diagnostic test did not include a vocabulary or skimming strategy section, thus only the comprehension and restating section was triangulated with the results of Unit 1 assessment.

The diagnostic test results of the reading comprehension section show that only two students were able to effectively identify and restate salient information in Spanish and almost half the group of students had a poor reading ability before taking the course. In contrast, most students successfully completed the task of identifying and restating salient information in the quiz; this suggests that the five weeks of training in reading strategies such as skimming and guessing meaning from context, among others as well as students' exposure to formative assessment yielded a favorable improvement for the majority of students who had performed poorly in the

diagnostic test. In general, the students who maintained regular attendance to lessons and who completed the review task made important progress from the diagnostic test to the achievement evaluation of Unit 1, which in turn contributed to a satisfactory accomplishment of the goal for that unit.

Unit 2 aimed for the identification and restatement of supporting details from academic psychology texts. Students' performance in the first section, dealing with guessing word meaning from contextual clues showed a significant improvement from the review task to the quiz. In the review task, results ranged between poor and fair. In contrast, performance in the quiz ranged between good and excellent. The second section, which evaluated recognition of key content words, referent words, connectors, and text organization, yielded positive results since most of the students demonstrated full understanding of all aspects. However, only few learners identified text organization accurately, which may imply that students needed further practice or instruction to master that task. Furthermore, students' majorly low proficiency level may have also hindered the full understanding of text organization. As for the results of the third section, students' ability to restate one supporting detail in Spanish did not reveal great improvement in the quiz after administering the review task, and some inconsistencies were observed in two cases. For example, a student who was absent the day of the review task obtained excellent results in the quiz; this could be due to preparation during class instruction. Another student who performed well in the review task obtained a lower score in the quiz.

Nevertheless, the rest of the learners succeeded in extracting supporting details in both instruments, confirming a satisfactory accomplishment of the goal for Unit 2.

Unit 3 dealt with identifying the author's purpose while appraising the usefulness of psychology academic texts for specific reading purposes. Students' performance in the vocabulary section of both assessment instruments demonstrated that most students improved results from good in the review task to excellent in the quiz regarding referent words of subordinators, which may have been influenced by formative assessment. For instance, two students performed poorly and one fairly in the review task due to having difficulty identifying the referent words of the subordinators whose, that, and which, but they obtained better results in the quiz. Results obtained in the second section, regarding identification of the purpose of hedging in the text ranged between good and excellent. The four students who did not do well on this section of the review task succeeded in the quiz. Regarding the third section, the results of students' ability to select and restate in Spanish an example of factual information and a hedging statement from the text showed that most answers in both assessment instruments provided an accurate and complete account of one conclusive finding and one cautious statement as required in the task. Data confirmed that students' performance in the quiz surpassed the review task. Interestingly, Unit 3 was expected to be the most difficult to teach and evaluate due to its cognitive load, yet students' performance was the best overall in the course, possibly

attributed to the closeness of the unit contents to the target context tasks often encountered by these students, which deal with finding the usefulness of academic texts for specific reading purposes. Therefore, the latter confirms the accomplishment beyond expectation of the goal for Unit 3.

Results of Self-Report Questionnaire

The self-report questionnaires collected qualitative data about student's perceived performance and attitudes after administering the review tasks and the quizzes, which allowed a deeper understanding of their needs and feelings while promoting self-reflection.

Regarding students' perceived performance, the instruments revealed that the perceptions of their own performance gradually improved throughout the course. For instance, after completing the first review task, which assessed achievement of course goals for Unit 1, only two students out of seven felt that they were able to carry out the tasks without difficulty. For the same criterion, that number increased from two to five students after completing the first quiz. Students attributed this improvement to further instruction after the review task and before the quiz, to the feedback obtained from the review task, to the adjustments made to quiz-item design and task difficulty after implementing formative assessment, and to a resulting increase in their self-confidence. The data collected for the subsequent units of the course shows even further improvement in this criterion. Indeed, after completing the review task for Unit 2, seven out

of eight students reported being able to perform tasks without difficulty. Similarly, after the review task for Unit 3, the majority of students expressed that they were able to carry out the tasks comfortably; and after completing the quiz for Unit 3, even more students expressed no issues regarding task completion. In general, students' perceptions of their performance improved after the quizzes. They found that the lessons after the review task and before the quiz and the feedback obtained from the review task had a very positive impact in their performance in the quiz and a lower affective filter towards summative assessment. Also they mentioned that adjustments made to quiz-item design and task difficulty after review tasks was also beneficial. Students added that strategy training, but more importantly, familiarity with test design and instructions had aided their performance. These findings reflect Black and Williams' (as cited by Harlen & Deakin Crick, 2003) assertion that formative assessment seems to improve "standards of attainment" in education (p. 170). Formative assessment during the study seems to have provided students with opportunities for getting acquainted with their expected performance, for making mistakes without getting penalized or judged by scores, and for reflecting on their performance; all these factors may explain their improvement during summative assessment.

A subsequent section of the questionnaire inquired about students' perceptions of previous preparation on their performance in both assessment instruments. Namely, the item referred to how instruction helped them carry out the review tasks, and if the

combination of instruction, practice with and feedback from the review tasks had increased their competence for the quizzes. The answers collected were mostly positive. Most learners considered instruction a useful and mirroring tool for the assessment phase, which indicated high levels of content and face validity among the participants. In addition, all learners felt that the review task provided a clear picture of the expected performance in the quiz, which they considered very advantageous. As suggested by Brown (2004), formative assessment seems to have promoted positive wash-back because it gave students the opportunity to identify their strengths and weaknesses before carrying out summative assessment. Formative assessment appears to have also encouraged students to view learning as a continuous process that should cohesively integrate instruction, practice, feedback and formal, traditional assessment.

In terms of perceived performance, students were asked to appraise their accomplishment of course goals. The results obtained are reassuring because the majority of learners felt capable of carrying out the tasks for the three units, both during instruction and assessment. However, two students stated that even though they were able to identify the information requested, that did not necessarily lead to comprehension. Further research would be needed in order to examine the relationship between text analysis and in-depth comprehension. Overall, these findings indicate an increased level of self-confidence and a sense of achievement among students. The combination of instruction and formative assessment appeared to have had

a positive impact on learning outcomes and motivation.

Regarding students' attitudes, the questionnaires inquired about students' dispositions before and after the quizzes. Results revealed that before administering the first quiz, none of the students had a negative attitude. However, for two students the situation changed after they took the quiz, since they reported not feeling positive about their performance. Fortunately, this was different when quiz two was administered because none of the students reported a negative attitude after the quiz, and the same happened with quiz three. Actually, one student who had reported a negative attitude before the third quiz felt better after taking it.

In general, the majority of students maintained a positive attitude throughout the course; this fact was unexpected because most of them had expressed reluctance to learn the language during the needs analysis. In fact, several students stated that by increasing their competence and self-confidence, the course influenced their attitudes because they felt no longer fearful, reluctant or unable to comprehend texts. Others mentioned that they now were more motivated to pick up a text and start reading it because it has become a much less daunting and more pleasant task. A few students stated that they still did not like reading in English, but that at least readings were more accessible. In general, these findings revealed that there was indeed a gradual change in attitudes among the majority of the students and that instruction as well as assessment had a positive impact on students' attitudes.

Overall, the data collected from the questionnaires demonstrated that by the end of the course, most of the students perceived an improvement in their reading skills, felt they had accomplished the course goals, and reported a more positive attitude towards the language and reading English academic texts.

Conclusions

The findings of the present study support the premise that formative assessment favors summative assessment results and accomplishment of course goals. Participants had poor reading ability before the course, but they had considerably improved it when the course concluded. The latter correlates to course success in terms of goal achievement since it proves that students accomplished the main objectives of the course satisfactorily.

Moreover, the implementation of formative review tasks before summative quizzes proved to be a very powerful tool when aiming at improving students' performance as well as self-confidence. Students felt guided, reassured and motivated to take quizzes because they were familiar with their expected performance and type of evaluation. The data from the self-reported questionnaires provided evidence of the gradual increase in motivation, enthusiasm and willingness to approach texts in English, as compared to the same data collected during the needs analysis done prior to the course.

One of the limitations encountered in the administration of formative assessment was absenteeism; only Unit 3 yielded complete sampling results.

In addition, the effect that different texts had on student performance during the review tasks and quizzes remains unknown, a note-worthy limitation because even though texts in both review tasks and quizzes contained similar carrier content, students pointed out that some texts were less demanding than others. Moreover, previous or concurrent exposure outside the English course to the topics or similar texts covered in assessment may have also affected sampling results.

A recommendation for future EAP courses in the psychology major is to promote a more comprehensive use of formative assessment, in which students regularly engage in self and peer evaluation exercises as well as group feedback sessions that promote learning reflection.

Bibliography

- Alvarez, L., Ananda, S., Walqui, A., Sato, E. & Rabinowitz, S. (2014). *Focusing formative assessment on the needs of English language learners*. San Francisco: WestEd.
- Bennett, R. E. (2009). *A critical look at the meaning and basis of formative assessment (RM-09-06)*. Princeton, NJ: Educational Testing Service.
- Brown, H. D. (2004). *Language Assessment: Principles and classroom practices*. White Plains, N.Y.: Pearson Education.
- Cruz, E., Dias, H. & Kortemeyer, G. (2011). The effect of formative assessment in Brazilian university physics courses. *Revista Brasileira de Ensino de Física*, 33(4), 4315. Retrieved November 22, 2014, from <http://www.>

- scielo.br/scielo.php?script=sci_arttext&pid=S1806-11172011000400016&lng=en&tlng=en. 10.1590/S1806-11172011000400016
- Harlen, W. & Deakin, R. (2003) Testing and motivation for learning. *Assessment in Education*, 10 (2), pp. 170–207.
- Horstmanshof, L. & Brownie, S. (2013). A Scaffolded Approach to Discussion Board Use for Formative Assessment of Academic Writing Skills. *Assessment & Evaluation In Higher Education*, 38(1), 61-73.
- O'Malley, J. M. & Valdez-Pierce, L. (1996). *Authentic Assessment for English learners*. Addison Wesley Publishing Company.
- Ryan, R. & Deci, E. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25, pp. 54-67.
- Stull, J., Varnum, S. J., Ducette, J. & Schiller, J. (2011). The Many Faces of Formative Assessment. *International Journal of Teaching and Learning in Higher Education*, 23(1), 30-39.
- Yin, Y. (2005). The influence of formative assessments on student motivation, achievement, and conceptual change. Doctoral dissertation. Retrieved from <http://gradworks.umi.com/31/86/3186430.html>

