

An Evaluation for an ESP Course for Mechanical Engineering Students

JONATHAN CALDERÓN BAILEY

Yorkin School, Costa Rica

MELISSA ROBLES GUZMÁN

Escuela de Lenguas Modernas,
Universidad de Costa Rica

LUCÍA UREÑA RODRÍGUEZ

Escuela de Lenguas Modernas,
Universidad de Costa Rica

Abstract

When designing a course, it is important to take into account the accomplishment of the goals set beforehand in order to evaluate the success of the process. This work presents the design, implementation, and evaluation of outcomes of an English for Specific Purposes course for Mechanical Engineering students. For this, the principles of Task-Based Language Teaching and English for Specific Purposes were followed in order to create a course that addressed the language needs of a group of students from the Mechanical Engineering major at the University of Costa Rica. A mixed methods approach was followed to analyze the data and present the results. The data was analyzed using the students' performance in certain tasks and their perception of how well they performed them. Based on the three goals established, the results showed that most participants achieved them to great or some extent.

Keywords: English for specific purposes, task-based language teaching, mechanical engineering, evaluation, language learning

Resumen

Al diseñar un curso, es importante tomar en cuenta el logro de los objetivos establecidos para evaluar el grado de éxito que el

proceso obtuvo. Este trabajo presenta el diseño, implementación y evaluación de los resultados de un curso de inglés con fines específicos para estudiantes de Ingeniería Mecánica. Para esto, los principios de enseñanza basada en tareas en inglés con fines específicos fueron la base para crear un curso que abordara las necesidades lingüísticas para un grupo de estudiantes de la carrera de Ingeniería Mecánica en la Universidad de Costa Rica. Se utilizó un enfoque de métodos mixtos para analizar los datos y presentar los resultados. Los datos fueron analizados usando el desempeño de los estudiantes en ciertas tareas y su percepción de este desempeño con respecto a los objetivos establecidos. Los resultados de esta investigación muestran que la mayoría de participantes alcanzó los objetivos establecidos al principio del curso con un nivel medio o alto.

Palabras clave: inglés con fines específicos, enseñanza de idiomas basada en tareas, ingeniería mecánica, evaluación, aprendizaje de idiomas

Introduction

Course evaluation is an important process in teaching a second language because it provides stakeholders, teachers, and students with a clear view of what was accomplished and what needs to be improved. In order to measure the effectiveness of a course, teachers can take into account course goals, objectives, and students' outcomes. For this research project, the achievement degree of the three goals stated was the guiding aspect to demonstrate the effectiveness of an English course taught to Mechanical Engineering students at the University of Costa Rica. Goal achievement was determined based on the students' outcomes in the test tasks and the students' perceptions of their own performance. The Mechanical English course was designed following the Task-Based Language Teaching (TBLT) and English for Specific Purposes (ESP) principles in order to meet the students' needs and wants in terms of their future job-related tasks. The evaluation implemented in this course was based

on different assessment types; some of them include criterion-referenced, task-based language, and self-assessment. All of these types of assessment practices helped us create appropriate tasks for the students to develop and for us to evaluate their performance. The following section is an organized account of what accredited scholars have published on course evaluation.

Literature Review

For the purpose of this course evaluation, we present different definitions of course evaluation and assessment aspects in order to evaluate the success of our course. Our literature review is divided into two main parts; definitions, characteristics, and purposes are presented in each of those sections. This part serves the purpose of clarifying how we assessed the effectiveness of our course through different types of assessment that we carried out during and at the end of the implementation of the course.

Evaluation

When considering the term evaluation, it is necessary to take into account the range of definitions that this concept has had over the years. Researchers and teachers have used this word with many different meanings that include theories, processes, and products that fall under a qualitative or quantitative approach to measurement. However, many of these definitions overlap and share many characteristics which make the term difficult to define. The following section presents several definitions and approaches to the term evaluation and tries to narrow down its characteristics.

To begin with, Weiss defines evaluation as “the systematic gathering of information for the purpose of making decisions.” (as cited in Bachman, 2003, p. 22) Similarly, Coombe, Folse and Hubley (2007) state that evaluation is “all inclusive and is the widest basis for collecting information in education.” (p. xv) Bachman and Palmer (2010) state that “evaluation involves making value judgments and decisions on the basis of information.” (p. 21) Based on the previous definitions, evaluation is a general term that involves making decisions about information that has been gathered. However, evaluation must not be confused with assessment. Lynch (1996) states that “evaluation tends to be used somewhat ambiguously in relation to other terms such as assessment and testing” (p. 2) but the author agrees with others when saying that evaluation is a process of gathering data in order to draw conclusions and make decisions about the learning process (p. 2). These decisions must be

well founded since they will directly or indirectly affect the individuals that participate in the learning process.

In regard to the characteristics of evaluation, Bachman (2003) states that a key aspect is to collect relevant and reliable information in order to make decisions about the best ways to evaluate an individual (p. 22). Moreover, this term “involves looking at all factors that influence the learning process, i.e., syllabus objectives, course design, and materials.” (Harris & McCann, 1994, as cited in Coombe *et al.*, 2007) In other words, evaluation should take into account many aspects that affect not only the learning process but also the results of students’ assessment. Gennessee expresses that evaluation goes beyond student achievement and language assessment to “consider all aspects of teaching and learning and to look at how educational decisions can be informed by the results of alternative forms of assessment.” (as cited in Coombe *et al.*, 2007, p. xv) In order to collect relevant information to make informed decisions, Bachman (2003) states not only that the information does not have to be exclusively quantitative but also that “verbal descriptions, ranging from performance profiles to letters of reference, as well as overall impressions, can provide important information for evaluating individuals, as can measures, such as ratings and tests scores.” (p. 22) Similarly, Lynch (1996) says that “evaluative information can be both qualitative and quantitative in form, and [it] can be gathered through different methods such as observation or the administration of pencil-and-paper tests.” (p. 2) Therefore, evaluation is concerned with collecting not

only numerical data about student and teacher performance but also qualitative data that will help teachers make decisions about the learning process and how to improve it.

As part of the evaluation developed for this course, language program evaluation and teacher-led evaluation were considered as follows.

Language Program Evaluation

Language program evaluation, as it is known today, is the result of what Jacobson recognized as the need of creating valid and convincing information related to language teaching programs (as cited in Lynch, 1996). The fact that program evaluations were being developed, but no clear, reliable information was available, led Jacobson to suggest the use of a needs assessment as a form of evaluation that intends to compare and/or contrast the original goals of the program and what is actually happening in the program at the moment of the evaluation (p. 32). The needs assessment became an instrument that would allow the evaluator to develop future studies based on an original starting point to be compared to the end of the process. Language program evaluation can serve different functions; its main objective is actually to evaluate a program in terms of methodologies, materials, teachers, and goal achievement, among others, in order to either make changes to improve the program or show results to a specific audience.

Teacher-led Evaluation

One possible way to develop program evaluation is conducting teacher-led evaluation or “projects which

teachers can carry out in their own teaching contexts.” (Kiely & Rea-Dickins, 2005, p. 246) These authors also discuss the idea that this type of evaluation is commonly known and encouraged among teachers who are allowed to try out new materials and activities in their daily classes. Most of the time, teacher-led language program evaluations focus on research questions that address concerns related to the effectiveness or the efficiency of specific aspects of language programs. Some of the factors that may encourage teachers to conduct teacher-led evaluations are the following: opportunities to change practice, need for change or improvement, availability of time, and teacher involvement in quality management (Kiely & Rea-Dickins, 2005, p. 248). This type of evaluation usually requires the use of qualitative data collection techniques like interviews, group discussions, recordings, and observations. By using these data collection methods, the teacher can involve students in the learning process and consider organizing workshops or meetings to report the results in a comforting environment (Kiely & Rea-Dickins, 2005, p. 249).

A useful approach to program evaluation in an ESP context is the one proposed by Lynch (1990); the author suggests a model for program evaluation which follows a set of established steps that can be adapted for different contexts. The first step refers to determining the audience and goals to be evaluated; secondly, based on the goals, the evaluator determines the context inventory which involves purposes of the course, language skills, and type of tests, among others. The third step is to establish a thematic framework that takes into consideration the students’ level of English proficiency and the use

of authentic materials as a means to provide real content. The next steps refer to the design of the data collection instruments that are going to be used and the actual collection of data. After this, the information is analyzed and displayed in a report for stakeholders and people directly involved in the program. Lynch's context-adaptive model represents the steps that we followed to develop our course evaluation.

Assessment

Evaluation is an umbrella term that includes all factors that influence the learning process, and one of these factors is assessment. Bachman & Palmer (2010) define assessment as "the process of collecting information about something that we are interested in, according to procedures that are systematic and substantively grounded." (p. 20) This means that, in order to develop an assessment project, the data collection process should be based not only on methodological procedures already established but also on more tangible aspects such as the course program and the course objectives; this would provide an opportunity to replicate and meaningfully analyze the results.

In order to differentiate assessment from other terms, it is important to describe its essential characteristics as explained by Wiggins (1998). The first feature is that real assessment ensures authentic performance. In other words, an authentic assessment task has to be realistic and test learners' language abilities the way they would use them in real life. Moreover, a task that is authentic requires students to be innovative; certain language functions or structures are expected to be used,

but students have the opportunity to use them as needed based on the context. The second characteristic of assessment is that it provides ongoing quality feedback, which provides the students with opportunities to improve their performance and skills. The third characteristic of assessment is that it promotes learner understanding instead of mere memorization or controlled behavior. Assessment is characterized by encouraging students to move from the level of knowledge to understanding, analyzing, and evaluating; these levels would be determined based on the students' proficiency and the specification of course goals.

Assessing in general can serve different purposes depending on the aspects to be assessed in a specific program. Some of the purposes explained by Brown (2012) include the use of assessment to determine the following: aptitude, proficiency, placement, diagnostic information, progress, and achievement (p. 134). Based on these purposes, our assessment was done throughout the course with the aim of assessing goal achievement to determine the effectiveness of the course in terms of student performance. Wiggins (1998) explains that "the aim of assessment is primarily to educate and improve student performance, not merely to audit it" with the purpose of giving feedback and helping students achieve the stated goals (p. 7). Having these purposes in mind helps the teachers focus and choose the appropriate type of assessment for their specific population.

Assessment can be divided into many different types; however, for the purpose of this project, we focused on the following categories:

Summative Assessment

Summative assessment “measures students’ achievement at the end of instruction.” (Lepi, 2013) In this type of assessment, students are given a grade to measure the extent to which the goals of the course were achieved. As Coombe *et al.* (2007) state, “[t]ests or tasks administered at the end of the course to determine if students have achieved the objectives set out in the curriculum are called summative assessments.” (p. xix) Moreover, the MIT (n. d.) states the following:

Summative assessment is used for the purpose of documenting outcomes and judging value. It is used for providing feedback to instructors about the quality of a subject or program, reporting to stakeholders and granting agencies, producing reports for accreditation, and marketing the attributes of a subject or program. Most studies of this type are rarely exclusively summative in practice, and they usually contain some aspects of formative assessment (Formative Assessment, para. 1).

The MIT acknowledges that summative and formative assessments are usually combined in a course or program since teachers can provide useful feedback to students to improve their learning process, and at the same time, a grade is given with the purpose of determining the extent to which a student attained the expected performance.

Criterion-referenced Assessment

Another important type of assessment is criterion-referenced. Connoley

(2004) defines this type of assessment as follows:

Criterion-referenced assessment makes judgements about performance, rather than on people. It assess[es] the extent to which a student has achieved the intended learning objectives and performance outcomes of a subject. Thus, student performance is compared to a previously specified standard of achievement (criteria) and not relative to other students (p. 3).

This type of assessment matches ESP methodology in the sense that both focus on students’ performance in a specific real-life task.

Criterion-referenced tests are “usually produced to measure well-defined and fairly specific instructional objectives.” (Brown, 2005, p. 2) Similar to what Connoley (2004) states, Bailey (1998) explains that, in this type of tests, “a given score is interpreted relative to a pre-set goal or objective (the criterion), rather than to the performances of other test-takers.” (p. 35) In other words, criterion-referenced tests are based on the objectives established at the beginning of the course or program; therefore, the results obtained can be compared to the goals and objectives in order to determine how successfully students perform a specific task.

Task-based Language Assessment

Task-based language assessment (TBLA) is grounded on the same principles underlying TBLT; that is, the task is the central component of teaching and testing. A common task in

TBLT is “(1) goal-oriented, (2) content-focused, (3) has a real outcome, and (4) reflects real-life language use and language need.” (Shehadeh, 2012, p. 156) Therefore, a test task in TBLA should follow the same line and be very similar to what is used in a TBLT class. Furthermore, Shehadeh points out some defined characteristics of TBLA that should be taken into account when assessing students under this approach. The first characteristic is that TBLA is a type of formative assessment; there should not be a grade that determines students’ success or failure, but there should be enriching feedback that encourages students to improve their performance. A second characteristic is that TBLA is a performance-based assessment; hence, it is intended to provide information about the students’ abilities to perform a real-life task in a simulated context. Third, it is a direct assessment; instructors measure specific outcomes that are incorporated into the task, but they would still be making inferences based on the process that led to the specific outcome (Shehadeh, 2012, p. 157).

Language for Specific Purposes Assessment

Having purposeful, relevant tasks in tests is also related to assessment in ESP. Tasks that are evaluated in an ESP context are very specific to the students’ needs, wants, and lacks. The importance of ESP language assessment relies on the fact that competency is a key factor that will determine the success of a student and a course. Therefore, the test provided to students must be designed according to the specific features that characterize

the use of the target language in real life. Douglas (2000) states that “the material the test is based on must engage test takers in a task in which both language ability and knowledge of the field interact with the test content in a way which is similar to the target language use situation.” (p. 6) As a result of this characteristic, students’ motivation may increase significantly resulting in more committed test takers.

Self-Assessment

Self-assessment, as defined by Coombe *et al.* (2007), “refers to the student’s evaluation of his or her own performance at various points in a course.” (p. 141) It is a type of alternative assessment that allows students to take more responsibility since it provides an opportunity to reflect upon their own learning process. Brown (2004) explains that this type of assessment enhances learner autonomy because it encourages students “to set [their] own goals... to pursue them without the presence of an external prod, and to independently monitor that pursuit.” (p. 270) Moreover, students are most likely to develop a desire to succeed in real life because they develop a sense of autonomy (Brown, 2004, p. 270). Another advantage of this type of assessment is that both teachers and students become aware of the perceptions regarding the students’ language abilities (Coombe *et al.*, 2007, p. 141). Nonetheless, it is imperative that instructors make sure that they provide students with the opportunity to reflect about their own performance; this will make the self-assessment process the source for future change and improvement.

Some kinds of self-assessment instruments include checklists, guided journal entries, and teacher-student conferences (Brown, 2004, p. 296).

This review of what accredited people have written has helped us define the main theoretical aspects that we took into consideration to evaluate our course as accurately as possible. By defining course evaluation, we established a starting point for this research project. In addition, we presented the definition of teacher-led evaluation, which is how approached this investigation. Furthermore, by defining evaluation and presenting its characteristics and purposes, and by differentiating it from assessment, we tried to avoid confusion during the development of the project, especially in the data collection process. Moreover, by explaining the different types of assessment, we have certainty that our data collection instruments yielded useful information related to our investigation purpose. Having all this theory as support for our project allowed us to carry out an investigation that presents reliable results to answer the following question: To what extent were the goals of the Mechanical English course achieved according to students' performance and perceptions?

Methodology

A mixed methods approach was followed to carry out this study. Dörnyei (2007) states that "a mixed methods study involves the collection or analysis of both quantitative and qualitative data in a single study with some attempts to integrate the two approaches at one or more stages of the research process." (p. 163) In this research

project, these two methods were combined in the instruments used, the display of results, and the analysis of these results. Furthermore, the subjects of study, the instruments used, and the procedure followed are described in the following sections.

Participants

This research study was conducted in a group of 10 students from the B. A. in Mechanical Engineering at the University of Costa Rica. At the moment, these students were enrolled in second, third, fourth, and fifth-year courses. This group of students was heterogenous in terms of their English level since there were beginner, intermediate, and advanced students. We assigned a number to each student, and these numbers remained the same to analyze the student's performance in the three units.

Instruments

To collect data, seven instruments were used: four test tasks and three self-assessment forms. The four test tasks were evaluated with a different rubric each depending on the objectives stated for each unit. The first test that we applied was a job interview, which was carried out at the end of Unit 1: Job Interviews. The students played the role of an applicant, and the teachers played the role of a company representative. In this task, the students were assessed based on what was taught during the first unit. The second test task consisted of writing an e-mail; similarly to test task 1, this one was applied at the end of Unit 2: E-mails at work. In this task, the

students wrote two different e-mails: one asking for a quote for a machine and another one ordering a machine. The third test task was a telephone conversation to give a project update to a coworker. The fourth test task consisted of a project presentation; test tasks three and four were used to evaluate goal achievement in unit 3 in which the students had to present projects and project updates.

Three different results sheets, one for each unit, were designed to evaluate the students' performance based on the objectives established for each unit; this facilitated the comparison of students' performance and perceptions. The students completed a form at the end of each test task as self-assessment; these forms included a list of statements related to the tasks that the students performed during each unit.

Procedure

The basis of our analysis was the comparison between the data gathered regarding students' opinions about goal achievement and the results of the four test tasks carried out. The first phase of this research project was the application of the instruments, which took place at different points during the semester. In the second part, all the data collected was analyzed based on three different degrees of goal achievement: great, some, and limited. The analysis of the data in this project was based on Dörnyei's (2007) principles of qualitative and quantitative data analysis. He describes the analysis of qualitative data as language-based and iterative; on the other hand, he explains that tables and

figures are "reader-friendly data presentation methods" characteristic of quantitative reports (p. 285). For each unit, a holistic rubric was used to evaluate the students' performance in each test task. In order for the students to meet the criteria in each statement of the rubrics to a great extent, they were allowed to have only a few mistakes per statement. The students met the criteria in each statement to some extent if they accomplished the expected outcome with some limitations. Those students who were not able to meet the conditions for each statement had a limited performance. Grades of each test task were given after the students completed the self-assessment so that these grades would not influence their perception in these test tasks.

Results and Discussion

This study attempted to determine the achievement degree of the goals in the Mechanical English course based on the students' performance and their perceptions. The following section shows the most relevant findings that helped to answer our research question.

In the self-assessment in unit 1 (to participate in jobs interviews successfully), only one student considered that he/she achieved Goal 1 to some extent; however, during the evaluation task, it was demonstrated that four out of ten students achieved this goal to some extent. One important aspect to mention is that, from the students who achieved the goal to some extent, student 9 was the only one whose perception of goal achievement was accurate in relation to his/her performance in the

evaluated task. Most mistakes made in this evaluated task were related to the simple past structure, the pronunciation of the -ed inflectional ending in regular verbs, and word order in indirect questions. These mistakes highly influenced how the students conveyed the message when talking about their academic background and previous work experience. On the other hand, those students who achieved the goal to a great extent also made a few mistakes, but they were not taken into account for this project because those mistakes were not related to what was taught for unit 1. One possible reason why students three, five, nine, and ten partially achieved the goal is that they showed to be the shyest students in the class and the ones who showed to be really nervous when they were evaluated. Even though these four students did not achieve the goal to a great extent, they still showed improvement since they were the least proficient according to the language test applied before the course started. As a matter of fact, one of the four students mentioned before expressed in the self-assessment that, despite considering that he/she achieved the goal to a great extent, he/she was aware that he/she needed extra practice on his/her own to improve his/her performance. It is also relevant to point out that no student showed a limited achievement of the goal, which means that they had no major problems performing the task. In this unit, the students' performance and perceptions show that an area for improvement was the use of the simple past and present perfect structures in terms of pronunciation and conjugation of regular verbs.

The goal in unit 2 was to successfully write e-mails to ask for quotes and order machines and machinery parts. The students' general goal achievement in this unit was very high since 7 students achieved the goal to a great extent while only two students did it to some extent and one to a limited extent. An important aspect to mention is that the students' perceptions of goal achievement were more negative than those regarding goal number 1. This might have happened because the students received the grade from the first unit before they completed the self-assessment for unit number 2, which probably made them more aware of their performance and more careful while completing the second self-assessment; however, the results of test task 2 were given after the students completed self-assessment 2. In this case, five students considered that they had achieved the goal to some extent; four out of these five students were actually the ones who achieved goal number 1 to some extent, which might have influenced their perceptions of goal achievement in unit number 2. However, out of those five who considered that they had accomplished the goal in unit 2 to some extent, four actually did it to a great extent; the fact that the task was writing an e-mail message might have increased their confidence to perform the task since it did not require a direct interaction with the evaluators. Even though these students achieved the goal to a great extent, they made common mistakes like the use of contractions, which shows informality. The other student who considered that he/she achieved the goal to some extent actually did it to a limited extent.

Even though his/her performance declined compared to unit 1, he/she was the only student who presented problems such as fragments, run-on sentences, and inappropriate greetings and farewells, and these mistakes really affected communication in the task. Furthermore, this student missed two classes from unit number 2; therefore, we can infer that this might have affected his/her performance in this task. Another salient aspect is that students six and eight, who accomplished goal number 1 to a great extent, had a poorer performance in unit 2; however, they still had very positive perceptions about goal achievement. The areas in which these students presented weaknesses were punctuation and the use of connectors when writing e-mail messages; nevertheless, their overall performance was sufficient to ensure effective communication. In general, areas for improvement after evaluating unit 2 are the use of connectors and punctuation marks; these two areas definitely represent aspects of the English language that require more than a unit for students to learn, practice, and finally acquire.

Unit 3 was a real success since nine students fully accomplished its goal: to participate in simulated videoconferences and telephone conversations to present projects and project updates appropriately, which matched most of the students' opinions about their performance. In terms of perceptions, students one and six believed that their performance was less successful than it actually was while student five perceived that he/she performed better than what his/her actual performance showed. Nonetheless, student five showed great improvement in relation

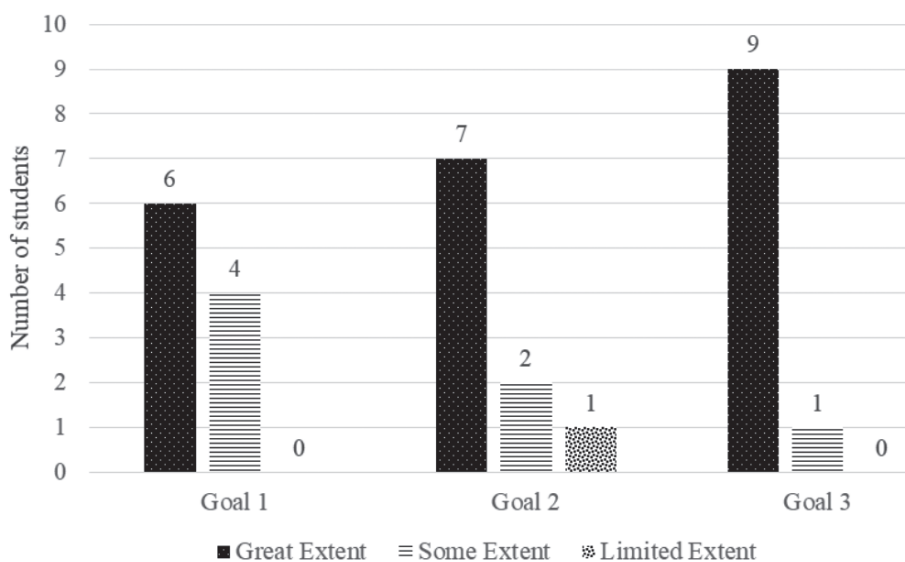
to unit 2. According to the students' comments, some aspects that were difficult for them were the use of technical vocabulary and the use of past tenses. A relevant aspect to highlight is that the students commented on the specific cases in which they had made the mistakes; this fact shows that the students were aware of their performance, and therefore, had a more realistic perception of their performance. Not only were these students' perceptions more accurate, but also their performance showed that they achieved the goal to a great extent and without any major communication problems. These same verb tenses represent the areas that needed improvement in unit 3; actually, we can say that the use of the simple past is the structure that challenged the most both teachers and students since it was present in two units of the course, and there were still some problems when the students used it in the last test task.

Overall, some important aspects must be pointed out after analyzing the data obtained during the three units of the course. Firstly, students two and seven were able not only to accurately perceive their goal achievement for each unit but also to actually achieve the three goals to a great extent. In addition, students one and four also achieved the goals to a great extent, but their perceptions of goal achievement were not precise probably because of lack of confidence. Another relevant aspect is that students three, six, eight, nine, and ten achieved two goals to a great extent and one goal to some extent, which shows that their overall course performance was positive. Furthermore, even though student five was not able to perform any

task to a great extent, his/her overall performance was acceptable because he/she showed great improvement in his/her speaking skills from unit 1 to unit 3. Finally, one of the most remarkable aspects is that students three, five, nine, and ten showed great improvement in terms of goal achievement from unit 1 to unit 3. We consider that some aspects that contributed to this improvement include practice, exposure to the language, the students' responsibility, and focused instruction.

As displayed in Figure 1, the extent to which the goals were achieved increased significantly from goal to goal. Moreover, in goal one, none of the students achieved the goal to a limited extent, and four students did it to some extent whereas in goal three, nine students achieved the goal to a great extent and only one to some extent. Based on this information, we still cannot assume that the students are better at speaking than at writing; nonetheless, we can state that the students' speaking skills significantly increased.

Figure 1. Achievement of course goals



n=10

In general terms, the majority of the students achieved the three course goals to a great extent; therefore, after this course, the students are now able to successfully participate in job interviews, write e-mails at work to ask for quotes and order machines or machinery parts, present project updates to coworkers, and present projects to superiors.

Conclusions

The following statements summarize the final remarks of this research project:

- Feedback was very important in our course since the students monitored themselves more often after receiving feedback not only

from the instructors but also from their classmates.

- Using test tasks was an effective technique to evaluate goal achievement since they resembled real-life actions. In addition, these test tasks encompassed all the performance aspects needed to determine goal achievement.
- Having the students complete self-assessment forms was important to obtain information from the students; however, the results obtained in the test tasks might not be consistent with the information gathered from the self-assessments due to the fact that self-assessment is very subjective.
- Course evaluation is necessary to determine if a course was successful or not and if certain areas need improvement.
- In general terms, we conclude that the Mechanical English course was successful because most of the students attained the three goals to a great extent.

Bibliography

- Bachman, L. & Palmer, A. (2010). *Language Assessment in Practice*. Oxford: Oxford University Press.
- Bachman, L. (2003). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bailey, K. M. (1998). *Learning about language assessment: Dilemmas, decisions, and directions*. United States of America: Heinle & Heinle Publishers.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. New York: Pearson Education Inc.
- Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment*. New York: McGraw-Hill.
- Brown, J. D. (2012). Choosing the right type of assessment. In Coombe, C., Davidson, P., O'Sullivan, B. & Stoyanoff, S. (eds.). *The Cambridge guide to second language assessment* (pp. 133-139). New York: Cambridge University Press.
- Connoley, R. (2004). *Criterion-referenced assessment*. Retrieved from: <https://www.deakin.edu.au/itl/assets/resources/pd/tl-modules/scholarly/studyguides/criterion-ref-assessment.pdf>
- Coombe, C., Folse, K. & Hubley, N. (2007). *A practical guide to assessing English language learners*. United States of America: The University of Michigan Press.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.
- Kiely, R. & Rea-Dickins, P. (2005). *Program evaluation in language education*. New York: Palgrave MacMillan.
- Lepi, K. (2013). *The 6 types of assessments (and how they're changing)*. Retrieved from: <http://www.edudemic.com/the-6-types-of-assessments-and-how-theyre-changing/>
- Lynch, B. K. (1990). A context-adaptive model for program evaluation. *TESOL Quarterly*, 24(1), pp. 23-42.
- Lynch, B. K. (1996). *Language program evaluation: Theory and practice*. Cambridge: Cambridge University Press.
- Massachusetts Institute of Technology. (n. d.). *Types of assessment and*

- evaluation*. Retrieved from: <http://tll.mit.edu/help/types-assessment-and-evaluation>
- Shehadeh, A. (2012). Task-based language assessment: Components, development, and implementation. In Coombe, C., Davidson, P., O'Sullivan, B. & Stoyhoff, S. (eds.). *The Cambridge guide to second language assessment* (pp. 140-146). New York: Cambridge University Press.
- Wiggins, G. P. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco, CA: Jossey-Bass, Inc., Publishers.