



Measuring Language Gains in a Foreign Language Context¹

Medición de las ganancias en el aprendizaje de una segunda lengua

Ana Muñoz-Restrepo²
Isabel Valderrama-Carvajal³
Alejandra Lopez Muñoz⁴
Ricardo Avendaño Franco⁵

Citation/ Para citar este Artículo: Muñoz-Restrepo, A., Valderrama-Carvajal, I., López-Muñoz, A. y Avendaño, R. (2020). Measuring Language Gains in a Foreign Language Context. *Colomb. Appl. Linguistic. J.*, 22(1), pp. 69-86.

Received: 29-Nov.-2018 / **Accepted:** 26-Feb.-2020

DOI: <https://doi.org/10.14483/22487085.14171>

Abstract

Learning gain can be defined as the difference between students' language competences demonstrated at two different points in time. In this article, we reported on a small-scale study aimed at measuring learning gains and piloting the methods chosen. Fourteen students of English as a foreign language participated in the study. We employed a student self-report survey, an oral performance assessment task and a standardised test for data gathering, and we applied them at three different moments during the eight-month study. Our data analysis includes the comparison of percentages rendered by test scores, the estimation of means for oral task scores and grouping and coding for qualitative data in the survey. Results show the gradual pattern of language improvement and the suitability of methods. These results can help stakeholders or policymakers in selecting relevant evidence to improve accountability in decision-making. We have provided conclusions and recommendations below for a larger scale study.


Keywords: adult students, EFL, learning gain, measuring learning, oral assessment, SALG

Resumen


La ganancia en el aprendizaje se puede definir como la diferencia entre las competencias en el idioma demostradas en diferentes momentos del aprendizaje. En este artículo, describimos un estudio a pequeña escala destinado a medir los logros de aprendizaje y realizar una prueba piloto de los métodos elegidos. Los participantes en el estudio fueron 14 estudiantes de inglés como lengua extranjera. Las herramientas de recopilación de datos incluyeron una encuesta de autoinforme de los estudiantes, una tarea de evaluación de desempeño oral y un examen estandarizado. Estas medidas se aplicaron en tres momentos diferentes durante el estudio que duró ocho meses. El análisis de datos incluyó la

1 This article corresponds to an investigation conducted between February and November of 2017 at the Language Center of EAFIT University.

2 Universidad EAFIT, Medellín (Colombia). ORCID  <https://orcid.org/0000-0002-6762-8762>. apmunoz@eafit.edu.co

3 Universidad EAFIT, Medellín (Colombia). ORCID  <https://orcid.org/0000-0001-6345-4374>. ivalder1@eafit.edu.co

4 Universidad EAFIT, Medellín (Colombia). alopezmu@eafit.edu.co

5 Universidad EAFIT, Medellín (Colombia). ORCID  <https://orcid.org/0000-0001-5884-0991>. ravenda1@eafit.edu.co

comparación de porcentajes generados por los puntajes de las pruebas, la estimación de promedios para tareas orales y la agrupación y codificación de datos cualitativos en la encuesta. Los resultados muestran un patrón gradual de mejora del lenguaje y la idoneidad de los métodos. Estos resultados pueden ayudar a las partes interesadas o formuladores de políticas a seleccionar evidencia relevante para mejorar su responsabilidad en la toma de decisiones. Se proporcionan conclusiones y recomendaciones para un estudio a mayor escala.

Palabras clave: estudiantes adultos, evaluación oral, ganancia en el aprendizaje, inglés como lengua extranjera, mediciones de aprendizaje, SALG

Introduction

Over the years, various assessment methods have been developed to effectively capture students' learning gains. These methods include grades, surveys, interviews, standardised tests, task performance, grammatical features, mixed methods and other qualitative methods (Briggs & Tang, 2011; Buckley, 2013; Buckley, 2015; McGrath, Guerin, Harte, Frearson & Manville, 2015; Maki, 2010). Every method has its advantages and disadvantages, presenting its own limitations and biases. Consequently, no method can be counted on to be completely error free. That is why the best practice in educational research is to triangulate the data, that is, to use a mix of measures (Breslow, 2007). In accordance to this, we will report on a study aimed at determining the appropriateness of three chosen methods to measure language-learning gains in an English programme for adults at a language centre in Medellin, Colombia. After a deep analysis of different ways to collect the data that included a revision of the literature, a summary of each method, consulting sessions with colleagues and discussions amongst the researchers, we decided to use student self-report surveys, a standardised test and oral performance tasks.

The study was conducted between March and November, 2017 with a group of students from the English programme for adults. The main objective of the research was twofold: to measure learning gains and to pilot the methods chosen.

The motivation to conduct the study came from an institutional requirement to determine if students in the English programme were reaching the proficiency levels established in the curriculum. Due to the magnitude and the necessary resources implied in such an endeavour, we conceived the current study as a small-scale research that could provide useful information for a future larger project from which generalisation could be made. Doubtless, the advantages that a more robust study can offer to a language-teaching institution are many: quality assurance, accountability, student retention, targeting of resources, curriculum development and pedagogical enhancement. Therefore, conducting such studies requires, first, the understanding that assessing learning gains is a complex process that cannot be carried out using just one single source such as a standardised test. Second, administrators and researchers must understand how research in this area can be conducted and what results may be obtained to make better evidence-based decision-making (Sands, Parker, Hedgeland, Jordan, & Galloway, 2018). Thirdly, they must understand that learning gain measures mainly aim to not only inform pedagogy but also provide the proof of the quality of a programme (Sands et al., 2018) or evidence for ranking purposes. As explained by Evans, Howson and Forsythe (2018, p. 30)

Learning gain as a concept has huge potential in being able to offer valuable insights into the learning process of all students if applied in a critical way as an integral part of curriculum design and delivery, and through utilising robust research design.

Bearing in mind all these considerations, we aimed to provide language programme administrators and researchers with an overview of good practice in learning gain measurements with the purpose of fostering future rigorous research.

Literature Review

Generally speaking, the concept of learning gain (or learning outcome) has been referred to as the effect of any specific educational intervention or as

the global distance travelled and learning acquired by students during an academic period of time with respect to skills and competences, content knowledge and personal development (McGrath et al., 2015; Rodgers, 2007). In the field of second language acquisition (SLA), learning gain has been defined as 'the loss of first language influences before a gradual gain of the second language system' (Ross, 1998, p. 4).

Most literature on the topic of factors influencing learning gains addresses language typology (the study of how languages differ regarding structure and function) and the extent of language distance between the native and the target language. Hence, the degree of the learnability of a second language is believed to depend mainly on the degree of relatedness between the languages (Ross, 1998). Thus, it seems that SLA by the speakers of typologically dissimilar languages entails substantial learning time and intensive exposure.

There are other factors intervening: the acquisition of a second language and the speed of the process. These factors are related to individual differences, such as beliefs toward learning a second language; intrinsic and extrinsic motivations and personal goal setting; previous learning experiences; foreign and native language literacy; non-linearity of the process; and physical impairment and personality traits (openness, adaptability, drive and determination, patience, confidence, meticulousness, optimism and persistence, amongst others).

On the basis of a review of models and approaches that have been used to measure learning gains in education, the following methods have been identified: grades, surveys, interviews, standardised tests, task performance, grammatical features, mixed methods and qualitative methods (Briggs & Tang, 2011; Buckley, 2013; Buckley, 2015; Maki, 2010).

Some methods like grades, standardised tests, grammatical features and mixed methods enable making comparisons in time (longitudinal analysis) and providing more general and easier way to measure data whilst achieving suitable response rate representativeness (Bonesronning, 1998; Brown,

1996; McGrath et al., 2015). Other approaches, including surveys, interviews and task performance, provide more in-depth information that could assist in the interpretation of gains. However, the latter measures are time consuming and costly and require extensive intervention and further validation (McGrath et al., 2015; McNaught & McGrath, 1997).

One of the most common and utilised approaches to directly measure learning consists of comparing the difference between students' grades at two points in time, using the grade point average as a systematic comparison, or using a set of grades (standardised or not) to make predictions on future grades. One of the difficulties of using this method is that teachers have different practices in grading and assessment, thereby only allowing broad comparisons (Allen 2005; Harlen, 2005).

Another approach is to ask students to self-report the extent to which they consider themselves to have learned using surveys that contain some questions related to student perceived skill development. Student surveys are one of the most powerful tools for understanding learning from students' points of view. They can never replace informal conversations between teachers and students, but properly contextualising the results can be highly revealing (Buckley, 2013).

Interviews also use standardised instruments, but they are conducted person-to-person (in person or over the telephone). Using this measuring method, respondents can elaborate on their answers, eliciting more in-depth information, but it is time consuming and expensive. Another standardised method is using tests that measure the acquisition of certain skills. These are considered to be more objective as the measures of learning gain than methods based on self-reports, such as surveys.

Task performances at two different times that allow for the comparisons of language gain (Briggs & Tang, 2011). The gain is inferred when a lack of competence on a performance assessment at an earlier date is replaced by a systematic review of successful task completion at a later date. According to McNaught and McGrath (1997), one of the disadvantages of tasks is that making inferences

about gains is often problematic because there must be extensive intervention.

Learning gains may also be observed over time as the development of particular grammatical features, which may be interpreted as reflecting an order of language acquisition (Gass & Selinker, 2008; Ellis 1984; Pienemann & Johnston, 1986). Some of the difficulties of this method are that specialised knowledge for analysing data is required and that it provides partial information about students' communicative competence.

The abovementioned methods can be combined to increase the robustness of the measurement through longitudinal data, i.e. data on the same group of students over at least two points in time (McGrath et al., 2015). These methods, given their mixed characters, include an element of objectivity. However, they are likely to require broad institutional capacity to implement in terms of systems and expertise to undertake data analysis.

Reference Studies

A search of the literature in language gains in SLA has proven unfruitful in the sense that there are not current and/or conclusive studies. The literature on the topic places emphasis on conceptual definitions rather than put-to-the-test samples applying these concepts on second language learning. Mainly, research has shown that despite the common use of standardised tests to assess gains, the legitimacy of the conclusions about observed changes in scores can be questioned. For instance, according to Beretta (1986), the use of standardised test results as the only criterion for evaluating language programmes can lead to inferential errors. Moreover, when different standardised tests are used, the evident question to be answered is the comparability of the tests. For instance, Geranpayeh (1994) conducted a correlational study on the IELTS and the TOEFL (N = 216 Iranian graduate students). The results showed that the scores of the most competent students on the two tests were less comparable than the scores of less proficient students. Geranpayeh's study concluded that language gain cannot be accurately deduced by the linear equating of scores on different instruments.

Bachman *et al.*, (1995) also examined the correlation and content co-variance between the FCE and the TOEFL. Their results suggested strong correlations between the test scores, but that different aspects of language knowledge were tested across the two batteries. This research suggests that language gain analysis cannot be effectively conducted if the pre-test is done with one set of measurement instruments and if the post-testing is done with another.

In a study on the effectiveness of an intensive English language programme at various levels of proficiency in the USA (Weissberg & Stuve, 1979), the Michigan Test was administered to a group of Latin American students (N = 63): once after entering the intensive programme and again after ten months. The sample was divided into three proficiency levels according to the students' initial scores. The improvement scores were obtained for each level, and the differences in gain rates were calculated within and between groups. Accordingly, they found a substantial improvement in proficiencies for each level, whilst no meaningful differences were found amongst the levels.

In the area of higher education, Roohr, Liu and Liu, (2017) evaluated learning gains of students' performance in reading, writing, critical thinking and mathematics using the ETS Proficiency Profile. The gains were estimated by calculating the score differences between the first and last test administrations. The results revealed that after spending one or two years in college, students did not show noteworthy learning gains and that after three or more years, students made small to moderate gains.

As can be seen by the description of these studies, conducting research in second language learning gains that not only uses standardised tests but also involves different data gathering methods is crucial. To our knowledge, there are no studies in foreign language-teaching contexts that have used different sources to estimate learning outcomes. We argue, therefore, the importance of an integrated approach that can account for what students have gained from their time in the language programmes.

Main Objectives

This study aims to determine English learning gains amongst a group of adult students and to pilot the selected methods to measure learning gains.

Method

Participants

We recruited fourteen students aged between 17 and 24 who were initiating course 1 in the English programme for adults. The participants were undergraduate students from different academic areas at the institution where the study was conducted and from different institutions in the city.

The English programme for adults covers distinct courses that have been aligned to the Common European Framework of Reference (CEFR; Council of Europe, 2001), which was run from a preparatory course (Pre A1), courses 1 to 17 (A1–B2) and a series of advanced courses. These are offered in different schedules and modalities, namely, the intensive modality (10 hours per week), the semi-intensive modality (6 hours per week) and the regular modality (4 hours per week). To recruit participants, we visited all the courses 1 that were being offered in the intensive modality, and on the basis of a written script, we explained the purpose, procedure and benefits of the study to the students. We did not include the preparatory course due to time constraints. The study began in March, and only courses 1 to 8 would allow us to cover the A2 level.

Students who agreed to participate in the study signed a consent letter where we ensured confidentiality and committed to provide them with three of the required course textbooks to be used during the length of the study as an incentive for their contribution.

Design and Procedure

We conducted a small-scale quantitative study that involved time 1 (T1), time 2 (T2) and time 3

(T3) measurements, including a standardised test, an oral performance task and a student self-report survey. Applying T1 measurement at the beginning of course 1 implies that students were not expected to know the answers to all of the questions; however, they were expected to utilise previous knowledge to predict rational answers. When applying the same measuring methods at the end of their studies, the students were expected to answer more questions correctly on the basis of an increase in knowledge and understanding. By applying these different measures at distinct times, we could obtain the depth of understanding and corroboration whilst balancing the weaknesses inherent in each approach. Additionally, this method allowed us to triangulate and validate the data (Breslow, 2007).

The data were gathered during eight months with students who started course 1 in March and finished course 8 in November, 2017 in the intensive modality. This modality and level of proficiency were chosen because they allowed us to follow the students for, at least, eight months and see how they evolved throughout that time.

According to the CEFR global scale (2001, p. 24), students at A2 can:

- understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography and employment),
- communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar or routine matters, and
- describe in simple terms the aspects of his/her background, immediate environment and matters in areas of immediate need.

During the second class of course 1 (March) and on the same day, we gathered T1 data from the fourteen students. In agreement with them, the appointments were scheduled for students to take an online test and Form A of an oral performance task in the Testing Centre of the institution. Following the same procedure, we gathered T2 from five students

in July using the same measurements. This time, we used Form B of the oral performance task to hinder the students from memorising the questions. Using an identical test can cause what is sometimes called the testing effect and can lead the assessment to measure familiarity instead of learning (Roedinger & Butler, 2011). Thereafter, in November, we collected T3 from one student again using the same instruments. This time, we used Form A of the oral performance task, pictures 2 and 3 (see Appendix 1) and the online survey.

The dropout rate was high. The Discussion section explains the reasons why the students abandoned the study and their English classes.

Data Gathering Tools

Standardised test: Track-Test to measure the students' performance at the A2 level. Track-Test is an online test aligned to the levels A1 to C2 of the CEFR that assesses listening, reading and writing. The test lasted 50 minutes and was administered in March (T1), July (T2) and November (T3).

Typical Track-Test methods include filling the gaps with the best word or sentence (multi-choice options), true or false questions related to an article and correct sequence ordering. Multi-choice questions assess the students' ability to infer links and connections between events and context that are implicit and to distinguish between literal and implied meanings and from factual and non-factual information. The test also includes different reading strategies: scanning, skimming, rapid reading and study reading.

In the A1/A2/B1 levels, questions are about short, simple texts, such as personal letters or advertisements or texts that consist mainly of high frequency everyday language. A threshold for the successful completion of the level is 65%. This test was chosen because it has accepted levels of validity and reliability. In addition, it renders score results automatically after its completion, and its price is reasonable.

Oral performance task: One of the researchers, who has knowledge and experience in test design, developed an oral performance task at the A2 level

(see Appendix 1 Form A and Form B). The task contained three parts, lasted 20 minutes and was taken individually. In Part 1, the students were asked factual and simple information: name, age, likes/dislikes and daily routines. Part 2 required students to describe a picture and then tell a story using the elements in the picture. In Part 3, the students were asked to describe a past event (birthday, special day, concert and trip—Form A) or make a comparison of how some everyday aspects of life, as shown in a picture, were in the past and how they are in the present (Form B). To validate the task, the researchers compared its content to the overall spoken descriptors of the CEFR for the A2 level: 'Can give a simple description or presentation of people, living or working conditions, daily routines and likes/dislikes as a short series of simple phrases and sentences linked into a list' (Council of Europe, 2018, p. 69).

The oral performance task was administered by one volunteer teacher, external to the study, who received written instructions and materials. The task was administered in March (T1), July (T2) and November (T3). All the tasks were recorded and then rated by two previously calibrated evaluators, who used an oral assessment rubric that had been designed and validated by previous research (Muñoz *et al.*, 2003). The aspects measured were communicative effectiveness, pronunciation, grammar, vocabulary and task completion.

Student self-report survey: This instrument was adapted and translated from an already validated survey called the Student Assessment of their Learning Gains (SALG; Carroll, Seymour & Weston, 2007). The survey contained 42 items that enquired into student self-reported learning gain within specific content areas, e.g. student understanding, skills, attitudes, teaching activities, evaluation, resources and integration of learning (see Appendix 2). The students indicated their perceived degree of gains on a three point-scale (no gains or help, limited gains or help and significant gains or help) and also provided answers to open-ended questions. The survey was administered online.

We decided to use this tool mainly because it had already been validated by different studies

(Canelas, Hillb, & Novickic, 2017; Lim, Hosack, & Vogt, 2012; Vishnumolakala, *et al.*, 2016; Wiese, Seymour, & Hunter, 1999) and because we considered that it would let us understand students' feelings and perceptions. The survey was applied in Spanish to facilitate the expression of feelings and reduce the possibility of misunderstandings, and for the purpose of this report, the researchers translated the comments into English. The survey was applied through Qualtrics (software for building, distributing and analysing surveys).

Data Analysis

Standardised test: The test provides online results for each of the skills measured in terms of percentages. Therefore, we compared percentages for T1, T2 and T3.

Oral performance task: Initially, we estimated the reliabilities between the two raters by counting the number of ratings in agreement for each language aspect of the rubric and the total number of ratings. Thereafter, we divided the total by the number in agreement and converted it to percentages; this procedure rendered an interrater reliability of 70%, which is considered to be high (Ruiz Bolivar, 2002; Palella, & Martins, 2003 in Corral, 2008). Finally, we calculated the score means for T1, T2 and T3.

Student self-report survey: The students' answers to each question were grouped under each corresponding scale (no gains or help, limited gains or help and significant gains or help). The answers to open-ended questions were analysed by identifying, grouping and coding the responses (Marshall & Rossman, 2016; Rubin & Rubin, 1995) pointed in similar directions in terms of gains or usefulness.

Results

Standardised Test

Table 1 below presents the results of the standardised test for the student (S1) who took the test in March, July and November and for the students who took the test in March and July. S1

increased scores progressively for all language skills at the three time measurements. Now, if we compare T1 and T3 for this student, the gains were important. The other students achieved minor gains in grammar, except for S5 who obtained lower scores at T2.

In reading comprehension, we can see that three of the students (S3, S4 and S5) increased their scores, particularly S5 whose scores were significantly higher at T2. In listening comprehension, S2 and S5 improved their scores, whereas S3 and S4 did not. The analysis of global score gains (Table 2) shows that all the students had some or important gains.

Table 1. Score Gains by Language Skill

	Grammar			Reading			Listening		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
S1	40	60	75	80	74	92	82	87	100
S2	50	60		80	72		78	83	
S3	45	50		20	21		23	22	
S4	40	50		36	63		80	53	
S5	40	35		36	72		80	92	

Table 2. Global Score Gains

	T1	T2	T3
S1	71	74.3	88.3
S2	70.3	70.5	
S3	68.4	71.6	
S4	52	57	
S5	45.5	65.8	

Oral Performance Task

The results of the oral performance task (Table 3) show that all the students improved their scores in all language skills from T1 to T2, except for S2 whose scores for communicative effectiveness and pronunciation lowered at T2. Of special interest is S1 who showed a progressive and substantial increase in all language aspects. In particular, we observed significant improvements in grammar and vocabulary from T1 to T3.

The global scores indicate some gains for all the students from T1 to T2, except for S2, and important gains for S1 from T1 to T3.

Student Self-Report Survey

The following results are based on the students who took the survey at the end of the study: one at T3 and the others at T2.

The students were asked to describe what gains in the general understanding of the language they felt as a result of the instructional process. The students expressed the following⁶:

- It is easier for me to understand the language structures. The lessons have helped me a lot with oral fluency. [S5]
- I understand the concepts and structures with greater clarity. [S3]
- I feel that my comprehension of different texts related to daily life experiences has improved. [S2]

6. Translated comments from Spanish by the authors.

- I can apply to my daily life the different topics studied in class. [S4]
- Express my own ideas, buy different items, introduce myself, give personal information and speak in the past tense and in the future tense. [S1]

When asked about gains associated to specific skills or competences, the students marked on the scale that they had significant gains in the following:

- identifying patterns and grammar structures in oral and written texts,
- expressing own ideas orally and written, and
- writing different types of texts with vocabulary and style appropriate to the context.

They reported, however, that gains were limited in regard to effectively interacting with others (understanding and being understood) in delivering oral presentations and in participating in or listening to debates or discussions. More specifically, they commented that they still did not feel comfortable

Table 3. Score Gains by Language Skill

	CE			Pro			Gra			Voc			TC		
	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
S1	2.4	2.5	4.0	2.0	2.5	4.0	1.0	1.5	3.8	1.0	1.7	3.5	2.0	2.3	4.3
S2	3.4	3.0		3.3	3.0		1.5	1.5		2.0	2.3		3.0	3.2	
S3	2.5	2.8		2.2	2.3		1.3	2.0		1.5	2.3		3.0	3.2	
S4	2.5	2.8		2.7	3.2		1.0	2.3		1.3	2.3		2.8	3.0	
S5	2.0	2.6		2.7	2.7		1.0	1.5		1.2	2.2		3.2	3.2	

Table 4. Global Score Gains

	T1	T2	T3
S1	1.7	2.1	3.9
S2	2.7	2.6	
S3	2.1	2.5	
S4	2.1	2.7	
S5	2.0	2.4	

expressing themselves in English and that their confidence in comprehending and in learning the language was still limited. Furthermore, they indicated that the ability to connect key ideas and structures with their field of knowledge or study was poor.

In relation to a question about the contribution that the use of teaching activities and resources had in the learning process, the students expressed that teacher explanations, group work and practical activities were of significant help. Furthermore, the textbook, My English Lab platform, and links or information sent by the teachers were reported as being of significant help. The students considered, though, that the audio-visual materials used in class were not useful.

Students were also asked about the influence of the assessment practices to their learning gains. They reported that the following activities contributed significantly to learn more:

- Reviewing
- Quizzes
- Connection between tests and class content
- Use of rubrics and writing conventions
- Teachers' feedback after assessment activities

Students specifically affirmed that the assessment activities helped them identify and correct mistakes. However, they did not find that assessment conducted in group activities was useful.

To the question, 'how has the whole instructional process changed your attitude towards the language?', the students identified significant changes in their interest and enthusiasm in practising the language and in their disposition to look for help from peers and teachers.

Discussion

Learning Gains

There is a general pattern of improvement at the different time measurements. Even though the time lapse between T1 and T2 was short, improvements were noticeable. In comparing T1,

T2 and T3, in particular for S1 who obtained a much higher T3 score, we can say that the student achieved the expected learning outcomes for an A2 level of proficiency. For instance, as expressed by this student, she obtained gains in the ability to express own ideas, buy different items, introduce herself, give personal information and narrate events in the past and future tenses. To support this particular student's perception, one of the evaluators after rating T3 wrote: 'I listened to S1's audio pleasantly, the student has shown great discipline by completing all of the courses and staying in the study. It's also clear that her listening skills improved and that, although she still needs to continue her work on fluency and vocabulary, the grammar of the previous courses is well internalised.'

In addition, the students perceived gains in their ability to see connections and application of the language to daily life situations. This has important implications for motivation because when students see that the learning activities have a connection to their life experiences, they will be more naturally and autonomously inclined to participate in the tasks (Muñoz & Ramirez, 2015). Moreover, extensive research in learning motivation suggests that when students perceive that a lesson has personal value or relevance, they tend to engage more, make more efforts and achieve more (Miller & Brickman, 2004; Vansteenkiste, *et al.*, 2007; Wigfield & Eccles, 2000).

Students cannot yet see a clear connection between what they have learned with their fields of study. Asking about the connection between academic progress and English proficiency allowed us to confirm that students' perceptions are coherent with their current level of proficiency because A2 communication goals are more related to daily life situations and cover immediate needs, but do not extend to more specific academic knowledge. However, they also express that they are aware of the need to improve their knowledge of English to enable their professional growth.

Moreover, there is a match between students' own perceptions of improvement in grammar and the results from the test and the oral performance task. The triangulation of these sources gives us

a clear indication of improvements in this ability. Moreover, students' perceptions of improvement in reading and writing match results from the test.

Even though there were some improvements in speaking abilities, students felt that they still could not communicate appropriately. Most importantly, they felt that they lack the confidence to speak and interact with others. It may be the case that the students' expectations regarding oral production are high for their level of proficiency. An A2 student is expected to produce simple phrases related to people, daily living or working conditions. However, students referred to higher-level skills such as debating or discussing topics for which they are not yet prepared. Therefore, students may need to be guided to be more tolerant with their own learning and set out more realistic goals. In so doing, teachers can help foster students' confidence and a sense of accomplishment. As Crow (2007) argued, it is possible that students may not completely understand how to perform certain tasks, but they must have the confidence that their teacher will provide them with individual support as they develop those tasks. Teachers can also foster a sense of confidence by giving students positive feedback that emphasises success and feelings of efficacy (Ryan & Deci, 2000).

Gains in task completion seem to indicate that students can better comply with the learning objectives and provide a certain degree of elaboration and detail.

Finally, although we do not have T3 data from most of the students, we could hypothesise, on the basis of the students' improvements at T2, that there could be significant language gains for all the students in a T3 measurement.

Suitability of Methods

The results indicate that the triangulation of the three methods used proved coherent as one method supported the results provided by the others. Nonetheless, individually, each technique presented its advantages and disadvantages. For instance, the standardised test captured unbiased learning gains rather than the perceptions of gains

and, therefore, depended less on the students' abilities to properly self-appraise their own learning progress as proposed by the SALG. Moreover, the test provided efficient and quick results aligned to the CEFR, which allowed for immediate comparison with our curriculum descriptors per proficiency level. However, the test had its limitations because some students expressed being nervous at the moment of taking the exam.

Considering that the oral performance task was directly designed on the basis of the A2 indicators, we were sure that the task had content validity. We also observed that the activities were not threatening; students seemed confident to try to answer them from the beginning. However, using tasks was time consuming and costly. Firstly, calibration sessions needed to be undertaken between the two raters prior to scoring the tasks. Secondly, it was necessary to organise the schedules and rooms and pay the task administrators. Despite the difficulties, we can conclude that the [oral] performance task is one of the most suitable methods to measure gains because it was directly aligned to the A2 performance indicators.

The SALG was chosen because of its proven validity, and it provided us with motivational variables that we initially considered important. However, these data were excluded from the results because many students did not answer some items at T2. Another difficulty with this instrument was that the software used (Qualtrics) combined the data for T1, T2 and T3 in a single file. It also did not discriminate the statistics for each time. To do the analyses, one of the researchers had to split the information for each time and perform the statistical calculations manually. In light of these difficulties, we recommend to upload the SALG in Google forms or other more practical survey tools and to accompany it by a list of performance indicators of the programme, that is, a checking list on a Likert scale with indicators by proficiency levels.

Desertion

To determine the causes for desertion, one of the researchers contacted the students by telephone. The reasons were as follows. Five students decided

to quit the programme due to the lack of time to study English. They commented that the academic load at the university interfered with their availability to enrol in the programme. Two of these students specified that, in addition to the academic load, they had to work. Therefore, they did not have time to dedicate to their English studies. Other students (3) told the researcher that they did not have the money to continue. They also considered the courses to be expensive; notwithstanding, they said that they liked the classes and were planning to continue attending them in the future. Two of these students added that they could not accommodate the schedules offered by the programme. Further reasons provided by the students were the following: one student decided to take English lessons at the college he was attending, and another student commented that she had family problems and could not continue. Finally, the other three students could not be contacted.

The reasons for dropping out are mostly due to external motives not dealing with, for instance, motivation, teacher interpersonal styles, teaching methodology or curricular matters. In fact, students perceived gains in their attitude or disposition to learn and mentioned that the resources and teaching and assessment practices helped them improve. However, reflecting on ways to maintain students in the programme is crucial. Thinking about more flexible schedules that accommodate diverse students' needs and about scholarships or aids for those students who have financial difficulties is possible.

Limitations

Firstly, the sample size was very small; therefore, caution must be exerted when making generalisations. The results provide preliminary information about the students' learning gains and the methods to consider undertaking a larger study. Secondly, there was a high dropout rate, which did not allow us to obtain conclusive data. It would be important for a larger research to consider strategies to maintain the participants in the study. This implies more human, financial, technical and logistical resources for the research. Taking these limitations into consideration, further research must

build upon the methods, limitations and results discussed in the current study.

Conclusions and Implications

We found very preliminary support for the notions that students indeed achieve gains and that some students are better than others. Even though it is satisfactory to find several gains, it is hard to discern whether the positive changes are due to learning in the classroom or simply due to natural maturation or other factors, such as students' abilities and motivation, teaching methodology and students' exposure to the language, amongst others, to provide a more comprehensive picture of the gains. Therefore, these variables must be considered in further projects with a bigger sample and following the students more consistently. Conducting a larger study including not only the variables mentioned above but also all the different language programmes at the language centre (Adults, Children, Young learners, Schools and other languages) and the different schedules is essential. Considering the variety of schedules and findings from the literature, following students for at least two years to capture their learning gains and providing them with more possibilities to accommodate different schedules are recommended to inhibit them from discontinuing their studies. This longer period would also allow the researchers to determine gains beyond the A2 proficiency level.

In addition, it is essential to consider that the main purpose of measuring learning outcomes is to improve teaching and learning. Therefore, teachers must be informed of the strengths and weaknesses found in learning measurement studies so that they can make informed decisions about instructional practices, e.g. emphasise the development of students' interaction skills, establish clear learning goals, promote meaningful learning tasks, foster students self- and peer assessments and give individualised feedback. These actions can increase students' self-regulation and motivation that can support the quality of their learning by helping them address the areas of relative weaknesses (Evans, 2016, 2018; Forsythe & Jellicoe, 2018).

This research has important limitations that deserve cautious interpretation of the data. Nonetheless, considering how the current study can contribute new knowledge to an understanding of the development of a larger project that can allow for generalisations programme wise is necessary. We believe that the methodology used was appropriate and rendered matches between data collection methods and results. In addition, considering quantitative and qualitative data can enable institutions to make decisions to better instruction, stronger curricula and more effective and efficient policies about the assessment of learning gains with the overall goal of improving teaching and learning.

References

- Allen, J. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 78(5), 218-223.
- Bachman, L. F., Davidson, F., & Ryan, K. (1995). *An investigation into the comparability of two tests of English as a foreign language* (Vol. 1). Cambridge University Press.
- Beretta, A. (1986). Program-fair language teaching program evaluation. *TESOL Quarterly*, 20(3), 431-444.
- Bonesronning, H. (1999). The variation in teachers' grading practices: Causes and consequences. *Economics of Education Review*, 18(1), 89-105.
- Breslow, L. (2007). *Methods of measuring learning outcomes and value added*. Working paper, Teaching and Learning Laboratory: Massachusetts Institute of Technology. Retrieved from <https://tll.mit.edu/sites/default/files/guidelines/a-e-tools-methods-of-measuring-learning-outcomes-grid-2.pdf>
- Briggs, J. & Tang, C. (2011). *Teaching for quality learning at university*. Midenhead, UK: McGraw-Hill and Open University Press.
- Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River NJ: Prentice Hall.
- Buckley, A. (2015). *Students' perceptions of skills development*. Higher Education Academy: England. Retrieved from https://www.heacademy.ac.uk/system/files/ukes_2015.pdf
- Buckley, A. (2013). *Engagement for enhancement: Report of a UK survey pilot*. York: Higher Education Academy. Retrieved from https://www.heacademy.ac.uk/sites/default/files/engagement_for_enhancement_final_0.pdf
- Canelas, D. A., Hillb, J. L., & Novickic, A. (2017). Cooperative learning in organic chemistry increases student assessment of learning gains in key transferable skills. *Chemistry Education Research and Practice*, 18(3), 441-456.
- de Franco, C., & Josefina, Y. (2009). Validez y confiabilidad de los instrumentos de investigación para la recolección de datos. *Revista Ciencias de la Educación*, 19(33), 228-247.
- Council of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division. (2001). *Common European framework of reference for languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Sanz, E. G. (2018). Common European framework of reference for languages: learning, teaching, assessment. Companion volume with new descriptors. marcoELE. *Revista de Didáctica Español Lengua Extranjera*, (26). Retrieved from <https://www.coe.int/en/web/common-european-framework-reference-languages>
- Ellis, R. (1984). *Classroom second language development*. Oxford: Pergamon.
- Evans, C. (2016). *Enhancing assessment feedback practice in higher education: The EAT Framework*. Southampton: University of Southampton.
- Evans, C., Kandiko Howson, C. & Forsythe, A. (2018). Making sense of learning gain in higher education. *Higher Education Pedagogies*, 3(1), 1-45, <https://doi.org/10.1080/23752696.2018.1508360>
- Forsythe, A., & Jellicoe, M. (2018). Predicting gainful learning in higher education; a goal orientation approach. *Higher Education Pedagogies*, 3(1), 103-117, <https://doi.org/10.1080/23752696.2018.1435298>
- Gass, S. M. & Selinker, L. (2008). *Second language acquisition: An introductory course* (Third edition). London: Routledge.
- Geranpayeh, A. (1994). Are score comparisons across language proficiency test batteries justified?. An IELTS-TOEFL Comparability Study. *Edinburgh Working Papers in Applied Linguistics*, 5, 50-65.
- Harlen, W. (2005). Teachers' summative practices and assessment for learning tensions and synergies. *Curriculum Journal*, 16(2), 207-223.
- Lim, B., Hosack, B., & Vogt, P. (2012). A framework for measuring student learning gains and engagement in an introductory computing course: A preliminary report of findings. *Electronic Journal of e-Learning*, 10(4), 416-428.
- Marshall, C., & Rossman, G. (2016). *Designing qualitative research*. Sage Publications: USA.

- Maki, P. L. (2010). *Assessing for learning: Building a sustainable commitment across the institution*. Sterling, VA: Stylus Publishing, LLC.
- McGrath, C. H., Guerin, B., Harte, E., Frearson, M., & Manville, C., (2015). *Learning gain in higher education*. Santa Monica, CA: RAND Corporation. Retrieved from http://www.rand.org/content/dam/rand/pubs/research_reports/RR900/RR996/RAND_RR996.pdf
- McNaught, C., & McGrath, J. (1997). *Review of AMEP program outcomes for 1994*. Sydney: National Centre for English Language Teaching and Research.
- Miller, R. B., & Brickman, S. J. (2004). A model of future-oriented motivation and self-regulation. *Educational Psychology Review*, 16(1), 9-33.
- Muñoz, A., & Ramirez, M. (2015). Teachers' conceptions of motivation and motivating practices in second language learning: A Self-determination theory perspective. *Theory and Research in Education*, 13(2), 198-220.
- Muñoz, T., Alvarez, M., Casals, S., Gaviria, S., & Palacio, M. (2003). Validation of an oral assessment tool for classroom use. *Colombian Applied Linguistics Journal*, (5), 139-157.
- Pienemann, M., & Johnston, M. (1986). An acquisition based procedure for second language assessment (ESL). *Australian Review of Applied Linguistics*, 9(1), 92-122.
- Rodgers, T. (2007). Measuring value added in higher education: A proposed methodology for developing a performance indicator based on economic value added to graduates. *Education Economics*, 15, 55-74.
- Roediger III, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15, 20-27.
- Roohr, K. C., Liu, H., & Liu, O. L. (2017). Investigating student learning gains in college: a longitudinal study. *Studies in Higher Education*, 42(12), 2284-2300, <https://doi.org/10.1080/03075079.2016.1143925>
- Ross, S. J. (1998). *Measuring gain in language programs: theory and research*. National Centre for English Language Teaching and Research, Macquarie University: Sydney.
- Rubin, H. J., & Rubin, I. S. (1995). *Qualitative Interviewing, the Art of Hearing Data*. Sage Publications: Thousand Oaks California.
- Ruiz Bolívar, C. (2002). *Instrumentos de Investigación Educativa*. Venezuela: Fedupel.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68-78, <https://doi.org/10.1037/110003-066X.55.1.68>
- Sands, D., Parker, M., Hedgeland, H., Jordan, S., & Galloway, R. (2018). Using concept inventories to measure understanding. *Higher Education Pedagogies*, 3(1), 60-69.
- Seymour, E., Carroll, S., & Weston, T. (2007). Student assessment of their learning gains (SALG). <http://www.salgsite.org/>
- Vansteenkiste, M., Matos, L., Lens, W., & Soenens, B. (2007). Understanding the impact of intrinsic versus extrinsic goal framing on exercise performance: The conflicting role of task and ego involvement. *Psychology of Sport and Exercise*, 8, 771-794.
- Vishnumolakala, V. R., Southam, D. C., Treagust, D. F., & Mocerino, M. (2016). Latent constructs of the students' assessment of their learning gains instrument following instruction in stereochemistry. *Chemistry Education Research and Practice*, 17(2), 309-319.
- Weisberg, R. C., & Stuve, M. (1979). Differential Gain Rates in Intensive ESL Programs: Who Gains the Most?. *System*, 7(1), 61-65. Retrieved from <https://eric.ed.gov/?id=EJ199687>
- Wiese, D., Seymour, E., & Hunter, A. B. (1999). Report on a panel testing of the student assessment of their learning gains instrument by instructors using modular methods to teach undergraduate chemistry. *Report to the Exxon Education Foundation*. Bureau of Sociological Research, University of Colorado, Boulder.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68-81, <https://doi.org/10.1006/ceps.1999.1015>



Appendix 1. Oral Performance Task

Form A

Greet the students. This part can be done in Spanish. Given that the task is at the A2 level, the students probably will not understand and/or produce much. Tell the students not to worry if they do not understand, and just reassure them.

Part I

Ask the students some factual information

- What is your name?
- What do you do?
- What do you usually do on weekends?
- Do you like.....? How often do you do it?
- Do you have siblings? What does your brother/sister do?

Part II

Give the students Picture #1, and ask them to describe it. Then, ask them to tell a story about it. Ask them to mention what happened at the time of the picture and before that. Give them approximately one minute to complete the task.

Provide the students with enough time to complete the task without intervention. Prompt only if necessary.

Part III

Say to the students:

- Now look at this picture (Picture #3). Make a comparison of how some everyday aspects were in the past and how they are now.

Hand over the picture to the participants. Provide them with enough time to complete the task without intervention. Prompt only if necessary.

Form B

Part I

Ask the students some factual information

- What is your name?
- What do you do?
- What do you usually do on weekends?
- Do you like.....? How often do you do it?
- Do you have siblings? What does your brother/sister do?

Part II

Describe the following situation to the participants:

You are living and studying in the US. Your classmates do not know about Colombia. Tell them about your country and your city. Include information about the culture, the people, the food and the interesting places to visit.

Provide them with enough time to complete the task without intervention. Prompt only if necessary.

Part III

Say to both participants:

- Student A: You just arrived from your vacations, and you are meeting your best friend. Describe him/her that experience.
- Student B: Your friend is telling you about his/her last vacations, and you want to know more details about them. Ask him/her questions.

The students should then change roles. Student A asks the questions, and Student B answers and describes a past event (e.g. birthday party, special day, concert and trip).

Provide them with enough time to complete the task without intervention. Prompt only if necessary.

Appendix 2. Student Self-Report Survey

Percepción sobre ganancias en el aprendizaje

El siguiente cuestionario busca conocer sus percepciones en torno a cómo ha aprendido en sus cursos de inglés. La encuesta tiene una duración promedio de 10 a 20 minutos; por favor tómese su tiempo y complete el siguiente cuestionario con **total sinceridad**, garantizamos el manejo confidencial de sus respuestas.

1. Como resultado de su trabajo en clase, ¿Qué ganancias ha tenido en la comprensión de los siguientes aspectos?

1.1 Los principales conceptos y estructuras explorados en clase.	Ninguna ganancia	Alguna ganancia	Ganancias significativas
1.2 Las correlaciones que hay entre las estructuras y elementos del idioma.	Ninguna ganancia	Alguna ganancia	Ganancias significativas
1.3 La forma como las ideas y temáticas de las clases de inglés en EAFIT se relacionan con otras áreas del conocimiento.	Ninguna ganancia	Alguna ganancia	Ganancias significativas
1.4 La forma como las temáticas de las clases de inglés en EAFIT se relacionan con situaciones de su vida cotidiana.	Ninguna ganancia	Alguna ganancia	Ganancias significativas
1.5 La forma como el aprendizaje del inglés le permitirá afrontar situaciones de la vida real.	Ninguna ganancia	Alguna ganancia	Ganancias significativas
1.6 Explique cómo ha cambiado su comprensión del idioma a partir de las clases de inglés en EAFIT.			

2. ¿Qué ganancias ha tenido en las siguientes competencias y habilidades, como resultado de su trabajo en clase?

2.1 Identificar patrones y estructuras gramaticales en textos orales o escritos.	Ninguna ganancia	Alguna ganancia	Ganancias significativas
2.2 Reconocer las relaciones entre los diferentes conceptos y estructuras.	Ninguna ganancia	Alguna ganancia	Ganancias significativas
2.3 Expresar ideas propias de manera oral.	Ninguna ganancia	Alguna ganancia	Ganancias significativas
2.4 Expresar ideas propias de manera escrita.	Ninguna ganancia	Alguna ganancia	Ganancias significativas
2.5 Escribir diferentes tipos de textos (oraciones, párrafos) con estilo y vocabulario adecuados al contexto.	Ninguna ganancia	Alguna ganancia	Ganancias significativas
2.6 Interactuar efectivamente con otras personas en inglés (entendiendo y dando a entender sus ideas).	Ninguna ganancia	Alguna ganancia	Ganancias significativas
2.7 Preparar y realizar presentaciones orales.	Ninguna ganancia	Alguna ganancia	Ganancias significativas
2.8 ¿Qué habilidades ha obtenido como resultado de su trabajo en clase?	Puedo...		

3. Como resultado de su trabajo durante las clases, ¿qué ganancias ha tenido en los siguientes aspectos?

3.1 Su entusiasmo por el idioma.	Ninguna ganancia	Alguna ganancia	Ganancias significativas
3.2 Su interés por discutir o practicar el idioma con amigos y familia.	Ninguna ganancia	Alguna ganancia	Ganancias significativas
3.3 Su confianza en que comprende el idioma.	Ninguna ganancia	Alguna ganancia	Ganancias significativas
3.4 Su confianza en que puede aprender el idioma.	Ninguna ganancia	Alguna ganancia	Ganancias significativas
3.5 Su nivel de comodidad expresándose en inglés.	Ninguna ganancia	Alguna ganancia	Ganancias significativas
3.6 Su disposición a buscar la ayuda de otros (docentes, compañeros) para resolver sus dudas.	Ninguna ganancia	Alguna ganancia	Ganancias significativas
3.7 ¿De qué manera ha contribuido esta clase a modificar su actitud hacia el idioma inglés?			

4. ¿Qué ganancias ha tenido al integrar los siguientes elementos, como resultado de su trabajo en la clase?

4.1 Conectar ideas y estructuras claves de las clases con otros conocimientos de su área de estudio o trabajo.	Ninguna ganancia	Alguna ganancia	Ganancias significativas
4.2 Aplicar las estructuras y lo aprendido en clase, a otros contextos y situaciones.	Ninguna ganancia	Alguna ganancia	Ganancias significativas
4.3 ¿De lo aprendido en el curso, qué podrá aplicar en otras asignaturas o en contextos de su cotidianidad?			

5. ¿Qué tanto han contribuido los siguientes aspectos de la clase a su aprendizaje?

5.1 Explicaciones por parte del profesor.	Ninguna ayuda	Alguna ayuda	Ayuda significativa
5.2 Participar o escuchar debates o disertaciones durante la clase.	Ninguna ayuda	Alguna ayuda	Ayuda significativa
5.3 Participar en trabajos en grupo durante la clase.	Ninguna ayuda	Alguna ayuda	Ayuda significativa
5.4 Hacer actividades prácticas durante la clase.	Ninguna ayuda	Alguna ayuda	Ayuda significativa
5.5 ¿De qué manera han contribuido las actividades de clase a su aprendizaje?			

6. ¿Cuánto han aportado los siguientes aspectos del proceso evaluativo a su aprendizaje?

6.1 Actividades evaluativas en general durante la clase.	Ninguna ayuda	Alguna ayuda	Ayuda significativa
6.2 Proyectos grupales evaluables.	Ninguna ayuda	Alguna ayuda	Ayuda significativa
6.3 Oportunidades de repaso en clase (dados por el profesor).	Ninguna ayuda	Alguna ayuda	Ayuda significativa
6.4 La cantidad de exámenes y tiempo entre éstos.	Ninguna ayuda	Alguna ayuda	Ayuda significativa
6.5 La relación entre los contenidos de la clase y de los exámenes.	Ninguna ayuda	Alguna ayuda	Ayuda significativa
6.6 La forma como el sistema de evaluación (las rúbricas, las convenciones de corrección de escritura, la claridad en cuanto a lo que se evalúa y cómo) le ayudó a comprender en qué aspectos debe mejorar.	Ninguna ayuda	Alguna ayuda	Ayuda significativa
6.7 La retroalimentación que recibió después de las evaluaciones o actividades.	Ninguna ayuda	Alguna ayuda	Ayuda significativa
6.8 ¿De qué manera han contribuido las actividades evaluativas a su proceso de aprendizaje?			

7. ¿Qué tanto le han ayudado los siguientes recursos en su proceso de aprendizaje?

7.1 El texto guía.	Ninguna ayuda	Alguna ayuda	Ayuda significativa
7.2 La plataforma My English Lab.	Ninguna ayuda	Alguna ayuda	Ayuda significativa
7.3 Enlaces, presentaciones e información enviada, recomendada o utilizada en clase por el profesor.	Ninguna ayuda	Alguna ayuda	Ayuda significativa
7.4 Recursos audiovisuales utilizados en la clase.	Ninguna ayuda	Alguna ayuda	Ayuda significativa
7.5 ¿Cómo han contribuido los diferentes recursos a su proceso de aprendizaje?			

Ha finalizado la encuesta, agradecemos mucho su participación.