

Preliminary Evaluation of the Impact of a Writing Assessment System on Teaching and Learning*

Evaluación preliminar del impacto de un sistema de evaluación en la enseñanza y el aprendizaje

*Ana Patricia Muñoz***

*Martha E. Álvarez****

EAFIT University, Colombia

Motivated by the conviction that valid and reliable assessments can foster positive changes in instruction, we implemented a writing assessment system in the adult English program at the Language Center, EAFIT University, Medellín. To evaluate the impact of the system, we analyzed the improvement of 27 students' writings over time. We also examined the quality of their teachers' (n=35) teaching and assessment routines. Student progress was measured by looking at the syntactic complexity of the texts using an average number of words per T-unit. Teachers' instructional practices were examined through portfolio analyses. Results show some improvement in students' writings and that the teachers need to provide a better response to students by appropriately using the required assessment tools.

Key words: Writing assessment, prompts, scoring rubrics, washback

Con base en la creencia de que las evaluaciones válidas y confiables pueden motivar cambios positivos en la enseñanza y el aprendizaje, implementamos un sistema de evaluación para la escritura en el programa de inglés de un centro de idiomas en Colombia. Para evaluar el impacto del sistema, analizamos las mejoras de 27 escritos a través del tiempo. Igualmente se analizó la calidad de las prácticas de enseñanza y evaluación de 35 profesores. Las mejoras en los escritos se determinaron a través de la complejidad sintáctica usando como medida el promedio de número de palabras por texto entre unidades T. Las prácticas de evaluación y enseñanza se evaluaron mediante el análisis de portafolios de escritura. Los resultados indican ciertas mejoras en los escritos y la necesidad de proporcionar mayor entrenamiento y orientación para que los profesores utilicen el sistema de evaluación de manera más apropiada.

Palabras clave: Evaluación de la escritura, instrucciones, rúbricas para calificar, efecto de arrastre

* This paper reports the results of the first phase of the project, "Evaluating the impact of a writing assessment system on teaching and learning," developed by the Research Group of the Language Center at EAFIT University, Medellín. Code number: 818022.

** E-mail: apmunoz@eafit.edu.co

*** E-mail: ealvarez@eafit.edu.co

Address: Carrera 49 No. 7 Sur-50 Medellín, Colombia.

This article was received on October 24th, 2007 and accepted on December 13th, 2007.

Introduction

This article draws together results of the first phase of a two year project, “Evaluating the impact of a writing assessment system on teaching and learning,” which is being undertaken by the Research Group of the Language Center at EAFIT University, Medellin. The aim of the project is to evaluate the effect of a carefully designed set of writing assessment practices on the teaching and learning of writing in the adult English program. The English program has a total of 15 courses that can be completed in different schedules. Students can take intensive, semi-intensive or regular courses, depending on their needs. The intensive courses can be completed in a four-week period, working two hours a day, Monday through Friday. The semi-intensive courses are completed in eight weeks with the students attending three times a week. Regular courses are completed in a 10- week period, meeting twice a week. The current study is being conducted with students in the intensive schedule of the adult program.

In 2005, the research group designed and validated a writing assessment system (WAS) with the intention of improving teaching and learning writing practices. Intentional actions towards positive washback¹ require, as some ELT professionals have recommended, congruity between assessment and curriculum related objectives, authenticity of tasks, detailed score reporting, teachers’ understanding of the assessment criteria, and learner self-assessment (Hughes, 2003; Messick,

1996; Bailey, 1996; Shohamy, 1996). The design of the WAS closely followed these recommendations. First, each component of the WAS –writing standards per course, rubrics, conventions, and writing tasks– was explicitly connected; second, the writing tasks were designed by considering authenticity requirements (parallel with real life situations, consistency with classroom and curriculum related objectives, and interaction between tasks and students’ background (Bachman & Palmer, 1996; O’Malley & Valdez, 1996; Douglas, 2000; Widdowson, 1979); and third, the rubrics were designed to render consistent application, $r > 0.7$ (Muñoz et al., 2006).

The WAS consists of a set of writing rubrics² aligned with writing standards for each course, writing conventions to check grammar, vocabulary, punctuation, and spelling problems, and tasks for the writing section of the mid-term and final tests. The system was implemented during the first academic quarter of 2006 after teachers had received training to familiarize them with its appropriate use. A three-module course dealing with theory and practice was offered, including: 1) definition of writing ability, 2) planning and design of writing tasks, plus 3) consistent use of the rubrics and conventions. Moreover, a training course was held to guide teachers on how to teach writing and how to keep writing portfolios.

The teaching of writing at the institution focuses on three basic components: 1) the process students go through when writing

¹ Washback refers to the influence of assessment on teaching and learning (Hughes, 2003; Wall & Alderson, 1993).

² Scoring scales for different levels of proficiency. They are used to measure different aspects of writing ability: coherence and cohesion, grammar, vocabulary, spelling, and task completion.

(prewriting, drafting, revising, and editing); 2) the accuracy, content, and organization of the writing; and 3) the particular genre the students are producing (letters, essays, biographies, reports, etc.). We believe that an emphasis on the process, the product, and the genre can help students greatly improve their writing skills by considering the personal process, the accuracy of the language used, and the purpose of the piece of writing (Harwood, 2005, Badger & White, 2000).

The Language Center (LC) does not have courses exclusively designed for the teaching and learning of writing. This skill is part of the regular language program. Teachers are required to organize their classes using the writing standards established for each course. This requirement obviously leads to presenting the class with tasks that follow the writing process established by the LC: pre-writing, drafting, revising, editing, and providing detailed scoring and feedback. It is expected that at least one of the writing tasks will constitute a formal writing evaluation for the course.

The writing component aims at developing different skills from elementary to upper intermediate levels of proficiency. For instance, students at the elementary level are expected to be able to fill in simple forms where personal information is required, write short, simple postcards, describe people, places, jobs or study experiences, write short, simple imaginary biographies, write simple personal letters, and narrate stories. At intermediate levels students are required to write short, simple essays on topics of interest, summarise, report and give opinions, write brief reports, write personal letters and notes

asking for or conveying simple information of immediate relevance. Finally, at more advanced levels, students are expected to write clear, detailed descriptions of real or imaginary events and experiences, write a review of a film, book or play, write an essay or report which develops an argument, and present an argument for or against a given topic (tasks adapted from the Council of Europe, 2001).

Through the accomplishment of these tasks, the LC seeks to prepare students for future academic or professional demands. Admission to academic programs, placement into different levels of a language program, exemption from certain course work or selection for a particular job will largely depend on how appropriately students master this mode of communication while involving different socio-cultural norms and cognitive processes.

In the current article we will first review some of the literature in the area of writing assessment, contending that meaningful assessment can motivate positive changes in the instruction and learning of writing. We will then describe the method and procedures involved in the realization of this study and present the findings and discussion for this stage of the project which includes preliminary results for two of the three hypotheses researched. In the final section, we will offer some conclusions and implications for the classroom.

Review of the Literature

The primary purpose of assessment is to make interpretations and decisions about students' language ability. In view of this, it

is essential, for a specific assessment system, to define the ability or construct to be measured. Construct definition is the most important consideration when assessing because it determines what aspects of the ability are to be measured and how they

are going to be measured. The definition of the construct for the LC includes the specification of writing standards for each course, the definition of the teaching approach to writing and the aspects of language knowledge and ability (see Table 1).

Aspects	Definition
Grammar and vocabulary (Linguistic competence)	The grammar and vocabulary domain addresses the control of grammar, vocabulary and sentence structure. It examines the appropriate use of language structures, effectiveness and range of lexical choice, and the appropriateness to context and to the demands of the task. It also covers the control of spelling and punctuation.
Coherence and cohesion (Discourse competence)	The coherence and cohesion domain addresses the logical development (organization) of the text that enables the reader to follow a thread of thought through the development of the written piece. Cohesion measures the students' ability to link ideas by using cohesive devices such as transitions and connectors. Such devices permit logical sequencing, establishment of time frames for actions and events, and the creation of structure of meaning by establishing main and supporting language units.
Task completion (Sociolinguistic competence)	The task completion domain addresses the students' ability to thoroughly complete the given task. It examines students' ability to achieve the specified writing performance indicators through their knowledge of functional uses of the language and the appropriate register. It also examines the extent to which the students are able to elaborate and provide sufficient details to illustrate ideas and go beyond the given task, avoiding digressions and irrelevancies.

Table 1. Definition of language ability (Based on ACTFL proficiency guidelines and IELTS writing descriptors).

After defining the construct, it is necessary to plan carefully how to measure it. This involves the design of the assessment tasks and the scoring methods. Therefore, the design of tasks calls for a specification of the prompt, which defines the task for student writing assignments. It refers particularly to the written instructions to the student. The prompt consists of the question or statement students will

address in their writing and the conditions under which they will write (O'Malley & Valdez, 1996). According to Hyland (2003), a prompt can include both contextual and input data. Contextual data relates to information about "setting, participants, purpose, and other features of the situation" (Douglas, 2000:55 as cited in Hyland, 2003). This type of information should be clearly and briefly stated in the prompt

and should be appropriate to the students' level of proficiency and background experience. Input data, on the other hand, refers to the "visual and/or aural material to be processed in a communicative task" (Douglas, 2000:57 as cited in Hyland, 2003). Different types of input data may be responding to a short reading text, analysing a table or chart, or describing a picture.

The wording of the prompt may include the purpose (or 'discourse mode') of the writing. It may also specify the genre, which refers to the expected form and communicative function of the written product, such as a letter, an essay, a report, etc. (Weigle, 2002). The prompt may also make reference to the pattern of exposition (Hale *et al*, 1996), which refers to the specific instructions to the students; for example, making comparisons, drawing conclusions, contrasting, etc. And finally, the prompt can mention the audience (the teacher, the classmates, general public), the tone (formal/informal), the length (100 words, one page, etc.), and time allotment (30 minutes, one hour). Weigle (2002) considers that a prompt should, at least, include the audience, the purpose and some indication of the length, but that the ultimate choice of specification depends on the definition of the construct.

Based on the literature presented above and on the definition of its writing construct, the LC considers that prompts at the institution should be as follows:

1. Be connected to the writing standards for any specific course.
2. Include the genre or the purpose of the writing.

3. Include the audience, either implicitly or explicitly.
4. Include the organizational plan or form of presentation which specifies how students are to develop the writing. It refers to the process or the steps students have to follow when developing a writing piece. It may include the number of words, time allotment, sequence, number of paragraphs, etc.

In addition to task design, an essential component of evaluation is determining the scoring methods. Since the judgment of student work is inevitably a subjective one on the teacher's part, a clear set of criteria must be identified and then applied consistently to each student's samples of writing in order to reduce teacher bias and increase the value of assessment. Teachers have found that a well-designed rubric can provide such a tool in promoting accurate, reliable writing assessment (Weigle, 1994; Stansfield & Ross, 1988). Additionally, teachers need to be trained to apply the rubric consistently. One source of unreliability in writing assessment is due to inconsistencies among different evaluators in scoring. Sufficiently high consistency in scoring can be obtained by means of proper training of the evaluators. Prior to proceeding to the scoring stage, examiners should understand the principles behind the particular rating scales they must work with, and be able to interpret their descriptors consistently (Alderson & Wall, 2001). This may be achieved by conducting meetings where a group of examiners get together, at the same time and place, to score samples and reach consensus. During the meetings,

raters compare their scorings and discuss any differences of opinion they might have.

Although different studies have been conducted on the accuracy and validity of large scale writing assessments (Novak, et al., 1996; Walberg & Ethington, 1991), little has been investigated concerning the impact of writing assessment on teaching and learning. For instance, Stecher et al. (2004) studied the effects of a test – the Washington assessment of student learning (WASL) – and a standards-based system on writing instruction in Washington schools. As a result of analyzing statewide surveys of both principals and teachers, the researchers found that although the approach to writing, a process approach, changed little before and after the test was instituted, curriculum (writing conventions, emphasis on audience, purpose, styles and formats) and instructional methods (greatest emphasis on WASL rubrics for student feedback) did change. The study concluded that the WASL influenced instruction. In another study, Lumley & Wenfan (2001) examined the impact of the Pennsylvania assessment policy on writing instruction and teaching methodology. The findings indicate that even though teachers agreed with the type of scoring and characteristics of effective writing proposed by the Pennsylvania Holistic Scoring Guide, they were reluctant to use the state rubric, descriptors, and writing samples. The authors concluded that there may be some deficiencies in the support material; that teachers may be using their own evaluation tools, or that they are not adopting the suggested writing approach.

As can be seen from the above studies, a host of elements beyond the assessment

itself needs to be considered. According to Wall (1996), different factors might prevent positive washback effects: teachers' lack of understanding of the exam, resistance to change, and exam content. She also refers to other factors such as the gap between test designers and teachers, and lack of well trained teachers.

The aim of the research described in the current article is to evaluate the impact of writing assessment practices on the teaching and learning of writing in English as a foreign language. More specifically, following the implementation of a writing assessment system, it is hypothesized that:

1. Student writing will significantly improve from pre- to post test;
2. Teacher writing instruction will significantly improve;
3. Student and teacher perceptions of the WAS will be positive.

Method

Participants

Twenty seven university students aged 17 to 20 participated in the study. Most of them enroll in the LC because they need to comply with a bilingualism policy established by the university. Others take classes because of academic or professional requirements. The adult English program has a total of 69 teachers; from these, 35 (20 females and 15 males) participated voluntarily in the study. Eighteen of the participant teachers had undergraduate degrees in language education or in translation from local universities. The others had degrees in other areas such

as administration or engineering. Most of them teach an average of 28 hours a week. The teachers had little experience in the teaching of writing. Therefore, they received training on how to teach and assess this skill. Additionally, they were instructed on how to keep writing portfolios for their students.

Data Collection

Before describing the data collection procedures, it is essential to visualize the current composition of courses for the adult English program, the distribution of courses by proficiency level, and the type of writing that students are expected to produce (see Table 2).

Proficiency levels	Courses for adults	Type of writing expected
Advanced	5 topic-based courses	Informative, descriptive, narrative persuasive (Upper intermediate to advanced)
Upper intermediate	10, 11, 12, 13, 14*	
Intermediate	6, 7*, 8, 9	Informative, descriptive, narrative (Pre-intermediate to intermediate)
Pre-intermediate	2, 3, 4, 5	
Elementary	N, 1	Informative (Elementary)

* Speaking and listening courses

Table 2. Distribution of courses by level of proficiency for expected writing abilities.

To test hypothesis 1, a longitudinal study involving a pre and post-test design for the pre-intermediate, intermediate and upper intermediate proficiency levels of the intensive schedule is being conducted. We want to observe students' writing improvement from course 2 to course 5, from course 6 to course 9; and from course 10 to 13. Measuring progress within levels of proficiency was imperative due to, on the one hand, the significant fluctuation of the student population. A great number of students interrupt their English classes at different periods during the year, mainly because of mid-term/final exams or summer/Christmas breaks at the university. Therefore, we cannot expect that the same number of students goes from course 1 to 13 without interruption. On the

other hand, existing research suggests that improvements can take place in a period as short as eight weeks (Arthur, 1980).

Part of the rationale for conducting a longitudinal study is that effective training of teachers does not happen overnight. Similarly, student progress needs to be examined over time. In the second stage of the project, we can compare improvement of student writing and teacher instruction in terms of where it was at the end of the first phase. Hopefully, teacher practice will improve over that time and, as a result, student writing should also improve.

For the pre and post-tests, writing tasks were designed according to levels of proficiency. A narrative type of task was designed for pre-intermediate, a narrative-descriptive task for intermediate, and a

persuasive task for the upper intermediate students (See Appendix 1).

The pretest writing tasks were applied initially to 126 students in courses 2 (n=57), 6 (n=47), and 10 (n=22) in May, 2006. Students were given 30 minutes to complete the task and no dictionaries were allowed. In August, the same tasks with the same instructions were given as the post-test to students who reached courses 5, 9, and 13. The process was repeated in July with students in courses 2, 6, and 10. The same tasks were given as the post-test to those who reached courses 5, 9, and 13 in October, when the collection of data for the year 2006 ended. By this time, the final sample population was 27 students who had taken the pre- and post tests. These students received writing instruction from the 35 teachers involved in the study. All the information gathered from the 35 teachers was taken into consideration because, at some point, they taught at least one of the 27 students.

Teachers' quality of writing instruction (hypothesis 2) was examined by analyzing 35 writing portfolios gathered from the teachers in the intensive schedule from May to September, 2006. Quality was defined based on the aspects that are deemed important in the LC approach to writing instruction and assessment. These are constituted by 1) congruence between task and writing standards for the course; 2) appropriateness of the prompt; 3) explicitness and elaboration of idea generation technique; 4) understanding of writing conventions; and 5) detailed scoring and feedback. As stated in the introduction to this report, teachers were instructed on how to keep the portfolios. The portfolios were distributed at the

beginning of each course. Inside each folder, steps were specified to guide teachers in the filing process (See Appendix 2). Teachers were expected to submit the folders at the end of each course including students' first drafts and final texts. Although teachers were to file students' writings, the purpose of the portfolio was to evaluate teachers' understanding of the writing process and scoring procedures as reflected in the writings. In other words, the interest was placed on teacher instruction and not on student performance per se.

Measures

- a. Pre- and post-test writing tasks to estimate student progress over time.
- b. Portfolios to assess teachers' quality of writing instruction.

Data Analysis

Impact on Student Writing

Measurement of student progress was done by examining the syntactic complexity of the pre- and post test writing tasks. Complexity is mainly judged by the higher frequency of complex sentences in a text. Wolfe-Quintero et al. (1998 in Polio, 2001, p. 96) highlight the idea that syntactic complexity means "that a wide variety of both basic and sophisticated structures are available... whereas a lack of complexity means that only a narrow range of basic structures are available..." According to Hunt (1970), the ability to combine more and more sentences is a sign of syntactic maturity. Moreover, syntactic complexity is the most

common feature for determining the effects of a program or intervention (Polio, 2001).

Different techniques have been used in writing research to measure text complexity, ranging from counting words, clauses, sentences or T-units in a text to averages of the number of words, clauses, or sentences per T-unit. A T-unit stands for 'minimal terminable unit' and is defined as a main clause plus all subordinate clauses and non-clausal structures attached to or embedded in it (Hunt, 1970). In simplified terms, a T-unit is the shortest unit which can stand alone as a sentence. For example: 'He stopped and he sat down on the soft grass.' has two T-units because there are two complete sentences which can 'stand alone'. There is a great deal of evidence that the T-unit has some special status as a meaningful unit of discourse which serves as a measure of syntactic complexity and cognitive maturity in a writer³. For instance, Hunt (1970) examined how the correlation of sentence length and academic maturity worked. He looked at the writing of fourth, eighth and twelfth graders and educated adults and found that 4th graders averaged 8.60 words per T-unit, 8th graders averaged 11.50, 12th graders averaged 14.40, and educated adults averaged 20.20 words per T-unit.

Using an average of number of words per T-unit, we analyzed the syntactic

complexity of the writings of the 27 students who completed both pre and post-tests (a total of 54 writings: 27 pre-tests and 27 post-tests). All the writings were coded to protect students' identity and typed in order to avoid problems due to illegible handwriting. The complexity ratio was obtained by counting the number of T-units per text and dividing it by the total number of words in the text. The number of words in the pre-test and post-test writing texts was balanced so that comparisons could be done between texts of similar length. The calculation of average words per T-units was done by two raters using Polio's 'Guidelines for T-units, clauses, word counts and errors' (1997). The count was done, first, individually. Then, the two raters compared results by naming in unison the final word of each T-unit. For the 54 writings, the average T-unit count for the two raters was 16.5 and 15.7 (P-value = 0.57) showing no significant differences between the evaluators at a 10% level of significance and a correlation of 0.97. The data were further analyzed using a signed rank test at a level of significance of 10% which shows the difference between complexity ratios. The level of significance was set at 10% due to the small size of the sample.

Impact on Writing Instruction

The analysis of portfolios was conducted using a rubric designed and validated⁴ for this purpose (See Appendix 3). The

3 The following example (taken from Hoelker & Hashi (2005), while correct, would be syntactically underdeveloped.

Fatima walked to the store (one T-unit).

Fatima walked slowly (one T-unit).

Fatima bought some bread (one T-unit).

Fatima returned home (one T-unit).

When combining sentences, transformations are performed on the sentences. A syntactically mature sentence, containing one T-unit reads: "Fatima walked slowly to the store to buy some bread and returned home."

4 To determine validity, the aspects measured by the rubric were aligned to the writing construct as defined for the Language Center (Muñoz, et al. 2006). Further, the descriptors for each aspect in the rubric were progressively adjusted by evaluating different portfolios used for piloting purposes.

rubric measures the aspects that are central to the LC approach to teaching writing and assessment: 1) congruence between task and writing standards for the course; 2) appropriateness of the prompt; 3) explicitness and elaboration of idea generation technique; 4) understanding of writing conventions; and 5) detailed scoring and feedback. The overall presentation of the portfolio (drafts and final version properly dated and organized) was also analyzed. Each aspect was evaluated on a 1-3 scale, where 3 = excellent, 2 = satisfactory and 1 = unsatisfactory. Two evaluators conducted the analysis of the 35 portfolios, first, individually, and then together to compare ratings. The table

below shows the percentage of agreement between the raters for each of the evaluated aspects. For instance, in relation to prompt design, the evaluators gave the same ratings to 29 portfolios (82.9%).

Based on these percentages, discrepancies were discussed and consensus reached for the final ratings.

Results and Discussion

Impact on Student Writing

The tables below present the results of the syntactic complexity analysis of the pre- and post test writing tasks for the 27 students.

Prompt	Congruence	Idea generation	Conventions	Scoring	Overall
82.9% (29)	97.1% (34)	77.1% (27)	94.3% (33)	91.4% (32)	94.3% (33)

Table 3. Percentage and frequency of agreement between raters.

Course 2				Course 5			
Code	Text words	T-units	Complexity ratio	Text words	T-units	Complexity ratio	Gains
029	127	15	8.4	125	9	13.8	5.4
002	110	15	7.3	109	10	10.9	3.6
038	80	10	8.0	78	8	9.75	1.8
007	158	13	12.1	160	13	12.3	0.2
035	52	8	6.5	52	8	6.5	0.0
036	112	9	12.4	111	9	12.3	-0.1
010	112	11	9.3	106	10	8.1	-1.2
041	124	6	20.7	127	8	15.9	-4.8
	109.4*			108.5*			

*Average of balanced number of words between texts

Table 4. Average number of words per T-unit (Courses 2 – 5).

The signed rank test for data in Table 4 showed that the number of students whose writings increased in level of complexity, students 029, 002, 038, and 007, is not significantly different from the number of students whose writings decreased in complexity, students 036, 010 and 041 (P-value = 0.62). Table 5 below illustrates that there are more students whose writings increased in syntactic complexity –Students 067, 044, 284, 045, 186, 182, 192, and 175– than writings where syntactic complexity decreased –046, 048, 179, 196–. However, the difference is not statistically significant (P-value = 0.12).

The data in Table 6 show that students' writings 099, 095, 202 and 204 increased in syntactic complexity whereas the rest of the students produced less mature writings in the post test.

In general, considering all the proficiency levels, 62% percent of the students showed an increase of syntactic complexity. Although for the majority the gains were not considerable, some students' syntactic complexity increased significantly: students 029, 002, and 038 from course 2 to course 5; students 284 and 182 from course 6 to 9; and 099 and 204 from course 10 to 13. Below, four excerpts taken from the writings are presented to exemplify the increase in syntactic complexity of students 002 and 284.

This is my friend Jose, # he is eighteen year old, he live in Sabaneta with his family, # he live in one big house out the city, # his family is very nice,# the house is beautiful,# there is bathroom, living room, kitchen, garage, bedroom,# in the bedroom there is one big bed and closet.#
Code 002 - Course 2

There are 7 T-units in this excerpt.
As can be seen the writer used very short

Course 6				Course 9			
Code	Text words	T-units	Complexity ratio	Text words	T-units	Complexity ratio	Gains
067	271	38	7.1	276	34	8.1	1.0
044	183	25	7.3	183	21	8.3	1.0
284	114	12	9.5	118	10	11.8	2.3
045	251	34	7.4	251	29	8.7	1.3
046	177	19	9.3	178	21	8.5	-0.8
048	150	22	6.8	148	24	6.2	-0.6
186	232	23	10.0	235	22	10.7	0.7
182	157	14	11.2	167	12	13.9	2.7
179	106	9	11.7	95	12	7.9	-3.8
196	174	18	9.7	171	18	9.5	-0.2
192	176	21	8.3	175	19	10.1	1.8
175	243	30	8.1	248	27	9.1	1.0
	186.2			187.1			

Table 5. Average number of words per T-unit (Courses 6 – 9).

Course 10				Course 13			
Code	Text words	T-units	Complexity ratio	Text words	T-units	Complexity ratio	Gains
099	252	18	14.0	257	17	17.0	3.0
206	157	9	17.4	162	10	16.2	-1.2
095	151	13	11.6	160	13	12.3	0.7
205	228	12	19.0	228	22	10.4	-8.6
202	153	9	17.0	148	8	18.5	1.5
211	307	24	12.8	296	25	11.8	-1.0
204	265	19	14.0	263	16	16.4	2.4
	216.1			216.3			

Table 6. Average number of words per T-unit (Courses 10 – 13).

sentences. In the post test, sentences were joined using coordination with subject deletion or subordination, producing a more complex text.

Many years ago, was a children what lived in a world where not exist material things.# They only could play with friend in the street, because not had television and computers.# They went at school # after they arrived at house to make the homework. # After they can go out to play soccer, and other plays.#

Code 002 - Course 5

The number of T-units here is 5. The student writes more complex sentences by using relativization and subordination. Despite the grammatical errors, the first T-unit shows the use of a relative pronoun 'what*' (instead of 'who') and 'where' to make this sentence longer [Many years ago, was a children what lived in a world where not exist material things]. The second T-unit uses a subordinating conjunction to make a more mature sentence [They only could play with friend in the street, because not had television and computers].

Here is another sample from a student pre-tested in course six and post-tested in course nine.

It happened the last Friday on night. # I was in my house with my cousins and my aunt. # I was seeing a movie with my aunt # and my cousins were in the kitchen. # We were in the first floor # suddenly we heard a sound in the second floor. #

Code 284 - Course 6

There are 6 T-units here. Most sentences are short and there is one attempt to combine through coordination without subject deletion [I was seeing a movie with my aunt # and my cousins were in the kitchen].

Many days ago I was walking to my house in the night and knew a girl that was walking in the same way. # She told me that her house was near of my house and asked me if I could walk whit her because she doesn't want to be alone. #

Code 284 - Course 9

This passage contains 2 T-units revealing more mature writing. In the first T-unit the subject of the second verb has been deleted

to avoid redundancy since it is the same as the subject of the first sentence [I]. In the first T-unit, there is also a relative pronoun [that] which makes the sentence longer and more sophisticated. The second T-unit uses again subject deletion and subordination yielding a more mature sentence pattern.

If we consider the average of all the complexity ratios in the last courses of the proficiency levels as shown in Table 7, it is possible to observe that the averages increased between course 5 and 13 (11.2 to 14.7) but dropped considerably in course 9 (9.4). It is likely that the writing assignments between courses 5 and 9 lacked emphasis on the subordination and relativization elements necessary to produce more complex syntactic patterns. It is also possible that consolidation of more complex structures, which are taught at this level, is taking place and therefore were difficult to produce.

It is also important to note that more complex sentences may be indicative of maturity, but not necessarily of quality. Too many complex sentences may be a problem because of an uncontrolled use of subordination which may reduce communicative effectiveness and the grammaticality of sentences.

Impact on Writing Instruction

Table 8 below presents the results of portfolio analysis. The portfolios were evaluated using the rubric designed for this purpose (see *Data analysis*, section *b*).

As indicated in the table, 17.5% of the teachers designed excellent prompts. This means that the prompts clearly followed the requirements for prompt design at the LC; in other words, specification of genre or discourse mode, audience, and organizational plan. 42.5% of the teachers

Course	Count	Average	SD	Min.	Max.
5	8	11.2	3.0	6.5	15.8
9	12	9.4	2.0	6.2	13.9
13	7	14.7	3.1	10.4	18.5

Table 7. Average of complexity ratios for courses 5, 9, and 13.

Aspect	Excellent	Satisfactory	Unsatisfactory
Prompt	17.5	42.5	40.0
Congruence	57.5		42.5
Idea generation	15.0	47.5	37.5
Convention	22.5	47.5	30.0
Scoring	2.5	40.0	57.5
Overall presentation	15.0	72.5	12.5

Table 8. Percentages of adequacy of writing assessment and instruction.

designed satisfactory prompts because they omitted one of the requirements or worded the prompt somewhat awkwardly. 40% did not include any of the specifications. Providing students with well-designed prompts is obviously an important aspect of assessment because students' successful performance greatly depends on how well teachers and test developers design the tasks. Therefore, prompt design becomes crucial to "allow all candidates to perform to the best of their abilities and to eliminate variations in scoring that can be attributed to the task rather than the candidates' abilities" (Weigle, 2002, p. 60-61).

With regard to congruence between prompts and writing standards, it was observed that while 57.7% of the teachers used writing tasks directly related to the writing standards, 42.5 % used activities that had little or no relation to the standard. Even though the standards are clearly defined for each course, teachers had difficulties making this connection. This might be due to the preference of some activities by the teacher or the students without regard to the objectives of the course. More awareness needs to be raised regarding the connection between these two aspects. When teachers and students recognize that the writing tasks directly assess the standards and that writing is assessed along clearly articulated levels of performance, teachers will be more motivated to change instructional practices both to teach and have students practice around these authentic assessments, and students will be more likely to buy into the value of such work (Natriello & Dornbusch, 1984).

The analysis shows that 15% of the teachers explicitly gave evidence of the technique used to generate ideas, such as brainstorming, listing, mind mapping, etc. The technique was clearly presented, elaborated, and reflected in students' writings and, therefore, evaluated as excellent. Although 47.5% of the teachers indicated the technique used, they did not fully elaborate on it or the technique was partly evidenced in the students' writing. Still, 37.5% of the teachers gave no evidence of the technique used. It is then necessary to encourage the use of pre-writing techniques. These are important because they spark general ideas on the topic in a draft form. Pre-writing helps students to focus on the topic by breaking it down into manageable sections and organizing ideas. In other words, it enables the writer to prioritise ideas.

Regarding the use of conventions, the data revealed that 22.5% of the teachers made an excellent use of the conventions, providing students with precise and appropriate feedback. These teachers seem to have a clear understanding of the symbols. Similarly, 47.5% of the teachers applied the conventions in an appropriate way, but still confused some of the symbols or used them inconsistently. 30% of the teachers appeared not to understand the symbols or not to use them at all. It is essential that all teachers use the conventions appropriately. Suitable use of these symbols may affect students' writing in a positive way because, when editing their writings, students need to exercise higher-order thinking skills such as analysis, synthesis, and evaluation in

order to improve their writings (Bereiter & Scardamalia, 1987; Elbow, 1983).

When scoring writings, 2.5% of the teachers were very specific providing scores for each aspect –coherence & cohesion, grammar & vocabulary, and task completion– and descriptor (each aspect is specified by descriptors a. b. and c.) of the rubric and offering qualitative comments to help students understand the score. 40% provided satisfactory scorings, meaning that they assigned scores for each aspect but did not give scores for each descriptor. Likewise, they included some useful comments for students. A great number of teachers, 57.5%, assigned global grades and did not comment on the students' writings.

Score reporting may be an influential factor in performance. Several studies confirm that global skills assessments seem to be less reliable than skill specific or behaviour specific descriptors (Chapelle & Brindley, 2002; Strong-Krause, 2000). Furthermore, it is crucial that teachers do not simply respond to grammar or content by means of scores but by making more personalized comments so as to maintain a meaningful dialogue with the student. Likewise, comments need to be related to the text itself rather than to general rules (Bates et al., 1993 in Hyland, 2003).

Finally, the analysis of the overall presentation of the portfolios shows that 15% of the teachers included all the required portfolio elements, mainly, specification of the prompt, presentation of the generation of ideas technique, inclusion of first draft with conventions, dated and commented, inclusion of second draft properly scored, dated and organized. Seventy two per cent of the teachers

presented satisfactory portfolios, meaning that some of the required elements were missing and dates were also sometimes omitted. The rest of the teachers (12.5%) omitted numerous pieces of writing or presented the material in a disorganized manner making it hard to analyze the data.

Conclusions and Implications

The results of the current study indicate some improvements in the syntactic complexity of students' writings. Sixty two percent of the students produced more mature texts by combining sentences using relative pronouns, subordinate conjunctions, and subject deletion. Factors such as lack of specific teaching and assessment of any of these elements may have affected the production of a higher number of syntactically mature writings. Since the pre and post-test writing tasks were designed according to proficiency levels, it is not possible to consider this as a factor that could have influenced syntactic maturity.

Based on the fact that teaching and assessment at the LC are more focused on the process than the product of writing, it became evident that syntactic complexity cannot fully account for student progress in writing. Being able to produce more complex sentences does not necessarily mean that there are less grammatical errors or that ideas are more coherently connected. Consequently, improvement cannot only be measured quantitatively. Other factors need to be considered for a more comprehensive view of language improvement. This is even more relevant if we consider that finding an objective measure of student progress

is difficult because a precise definition of improvement becomes impossible. According to Casanave (1994), examining quantitative changes in student writing overtime is, in one sense, a failure because a picture of student writing will necessarily be very incomplete. Therefore, it is important to examine aspects such as coherence, cohesion, revision, and task completion, which might provide better information about the process rather than the product of writing. These aspects may be difficult to evaluate in terms of face validity for those who want to see objective measures. However, coherence and discourse features are concerned with the quality of a text's organization, and therefore with how clear students are able to communicate an idea. The relationship between teacher feedback and student revision is also important to examine in order to determine the degree to which students address the teacher feedback and the degree to which revisions are related to teacher comments. By analysing task completion, we can observe how completely students develop the prompts and accomplish the writing standards.

Additionally, a measure of language improvement needs to match our assessment system as used in the context of our own classes. In the current study, teachers emphasized the teaching of grammar and, accordingly, students were able to produce more complex sentences. However, in order to have more clarity of the effects of the WAS on learning, we need to consider all the aspects involved in this system, going beyond all the implicit complications involved in measuring qualitative variables.

Regarding the teaching and assessment of writing as evidenced by the portfolios, it is clear that the teachers have not been able to implement the WAS as established by the LC. Different factors may account for this: when innovative assessments are proposed it may take teachers some time to adjust to changes. It is possible that some aspects of the assessment system are not yet internalized or clear to teachers. Therefore, we need to offer more training opportunities and stimulus that will motivate teachers to participate in the assessment process in a more committed manner.

It is also possible that teachers are somehow reluctant to implement the WAS. It is likely that the suggested assessment system makes new demands on the teachers' competencies and beliefs. It is also possible that the design of the WAS presumes that teachers have certain beliefs about the nature and goals of evaluation. This would obviously lead us to the field of teacher cognition, defined by Borg (2003, p. 81) as the "unobservable cognitive dimension of teaching –what teachers know, believe, and think." In fact, different research studies indicate that teachers have complex beliefs about pedagogical matters which, according to Borg (2003), create a structured set of principles. These principles, he points out, are derived from teachers' prior experiences, school practices, and individual personalities. In the field of English language teaching beliefs have been studied to see how they have informed the instructional practices and decisions of teachers (Borg, 2003; Burns, 1992). Furthermore, other research literature suggests that beliefs and practice are related,

and that teachers may hold beliefs that are not compatible with the practices called for in institutional plans (Bliem & Davinroy, 1997; Borko et al., 1997). It then follows that meaningful change in assessment practices may require change in teachers' beliefs about such practices. This would be an important topic for future research: how to introduce change in educational settings and how to involve teachers in it.

In addition to the abovementioned observations, it is also imperative to consider that the effects of teacher training programs take place over time. In the second phase of the project, after more training is provided and time has elapsed, it will be possible to verify this connection.

Overall, the results of this study provide a clear knowledge of the areas that need further improvement; this, in turn, will lead to the implementation of corrective measures in these areas as well as constitute a comparison point for the data gathered in the second phase of the study. It will then be possible to compare teacher instruction and student progress in writing. It is expected that writing instruction will improve and, as a result, student writing will also improve.

References

- ACTFL proficiency guidelines writing (2001). Retrieved July, 2005 from <http://www.actfl.org/files/public/writingguidelines.pdf>
- Alderson, J. C., & Wall, D. (2001). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Arthur, B. (1980). Short-term changes in EFL composition skills. In C.A. Yorio, & J. Schachter (Eds.). *On TESOL '79: Focus on the learner* (pp. 330-342). Washington, DC: TESOL.
- Bachman, L., & A. Palmer. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Badger, R., & White, G. (2000). A process genre approach to teaching writing. *ELT Journal*, 54(2), 153-60.
- Bailey, K. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13, 257-279.
- Bates, L., Lane, J., & Lange, E. (1993). *Writing clearly: Responding to ESL composition*. Boston: Heinle & Heinle.
- Bereiter, C., & Scardamalia, M. (1987). An attainable version of high literacy: Approaches to teaching high-order skills in reading and writing. *Curriculum Inquiry*, 17(1), 9-30.
- Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, 36(2), 81-109.
- Borko, H., Mayfield, V., Marion, S. F., Flexer, R. J., & Cumbo, K. (1997). *Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development* (Report 523). Los Angeles: Center for the Study of Evaluation.
- Bliem, C. L., & Davinroy, K. H. (1997). *Teachers' beliefs about assessment and instruction in literacy* (Report 421). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Burns, A. (1992). Teacher beliefs and their influence on classroom practice. *Prospect*, 7(3), 56-66.
- Casanave, C. P. (1994). Language development in students' journals. *Journal of Second Language Writing*, 3(3) 179-201.
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Chapelle, C.A., & Brindley, G. (2002). Assessment. In Schmitt, N. (Ed.), *An introduction to applied linguistics*, (pp.267-88). London: Arnold.
- Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.

- Elbow, P. (1983). Teaching thinking by teaching the writing process. Stony Brook, NY: State University of New York.
- Hale, G., et al. (1996). A study of writing tasks assigned in academic degree programs, (TOEFL Research report No. 54). Princeton, NJ, Educational Testing Service.
- Harwood, N. (2005). The sample approach: Teaching writing with Cambridge Examination classes (online). Retrieved December 7, 2007, from <http://privatewww.essex.ac.uk/~nharwood/sampapproach.htm>
- Hoelker, J., & Hashi, Z. (2005). Successful EFL Writers from the Gulf. Retrieved November, 2006 from http://www.englishaustralia.com.au/index.cgi?E=hcfunc&PT=sl&X=getdoc&Lev1=pub_c06_07&Lev2=c05_hoelke
- Hughes, A. (2003). Testing for language teachers. New York: Cambridge University Press.
- Hunt, K. W. (1970). Syntactic maturity in schoolchildren and adults. *Monographs of the Society of Research in Child Development*, 35 (1, Serial No. 134).
- Hyland, K. (2003). Second language writing. Cambridge: Cambridge University Press.
- IELTS writing descriptors (undated). Retrieved July, 2005 from http://www.ielts.org/_lib/pdf/UOBDS_WritingT1.pdf
- Lumley, D.R. & Wenfan, Y. (2001, April 10). *The impact of state mandated, large-scale writing assessment policies in Pennsylvania*. Paper presented at the annual meeting of the American Educational Research Association. Seattle: WA.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256.
- Muñoz, A, Mueller, J., Alvarez, M., & Gaviria, S. (2006). Developing a coherent system for the assessment of writing abilities: Tasks and tools. *Íkala, revista de lenguaje y cultura*, 11(17), 265-307.
- Natriello, G., & Dornbusch, S. M (1984). Teacher evaluative standards and student effort. New York, Longman.
- Novak, J.R., Herman, J.L., & Gearhart, M. (1996). Establishing validity for performance based assessments: An illustration for collections of student writing. *The Journal of Educational Research*, 89(4), 220-233.
- O'Malley, M., & L. Valdez. (1996). Authentic assessment for English language learners. Addison-Wesley Publishing Company.
- Polio, C. (2001). Research methodology in L2 writing research. In Silva, T. & Matsuda, P.K. (eds.), *On second language writing* (pp. 91-115). New Jersey: Lawrence Erlbaum Associates.
- Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language learning*, 47(1), 101-143.
- Stansfield, C. W., & Ross, J. (1988). A long-term research agenda for the Test of Written English. *Language Testing*, 5, 160-186.
- Shohamy, E., Donitsa-Schmidt S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13, 298-317.
- Stecher, B., Chun, T., & Barron, S. (2004). The effects of assessment-driven reform on the teaching of writing in Washington state. In Chen L. & Watanabe, Y. (Eds.), *Washback in language testing* (pp.147-170). New Jersey: Lawrence Erlbaum Associates.
- Strong-Krause, D. (2000). Exploring the effectiveness of self-assessment strategies in ESL placement. In Ekbatani, G. & Pierson, H. (Eds.), *Learner-directed assessment in ESL* (pp. 255-78). Mahwah, NJ: Lawrence Erlbaum.
- Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10, 41-69.
- Wall, D. (1996). Introducing new tests into traditional systems: Insights from general education and from innovation theory. *Language Testing*, 13, 231-354.
- Walberg, H.J., & Ethington, C.A. (1991). Correlates of writing performance and interest: A U.S. national assessment study. *Journal of Educational Research*, 84(4), 198-203.
- Weigle, S. (1994). Effects of training on raters of ESL compositions, *Language Testing*, 11, 197-223.
- Weigle, S. C. (2002). Assessing writing. Cambridge: Cambridge University Press.

Widdowson, H. (1979). *Explorations in applied linguistics*. Oxford: Oxford University Press.

Wolfe-Quintero, K., Inagaki, S., & Kim, H.Y. (1998). *Second language development in*

writing: Measures of fluency, accuracy, and complexity (Tech. Rep. No. 17). Honolulu, HI: National Foreign Language Resource Center.

About the Authors

Ana Muñoz is the Coordinator of Research and Teacher Education at the Language Center, EAFIT University, Medellin. Her current research interests cover assessment and evaluation. Her recent publications include *Students' objectivity and perception of self assessment in an EFL classroom* (2007) and *Developing a coherent system for the assessment of writing abilities: Tasks and tools* (2006).

Martha E. Alvarez is a professor in the School of Engineering, EAFIT University, Medellin. She holds a postgraduate degree in Information Systems from Universidad EAFIT. Her recent publications in EFL include *Students' objectivity and perception of self assessment in an EFL classroom* (2007) and *Developing a coherent system for the assessment of writing abilities: Tasks and tools* (2006).

Appendix 1: Writing Tasks for Pre- and Post Tests

Courses 2 - 5 Narrative:

Write a short story about children who live in a world where there is no such thing as television, computers, or electronic games. Consider aspects such as family life, school, and friends.

Courses 6 - 9 Narrative + description:

Write an account about a real or imaginary scary moment in your life. Your readers want to know what happened and where it happened. Be sure to include how the place looked, and how the people involved felt and reacted. Readers also want to know why this incident sticks in your memory.

Courses 10 -13 Persuasive:

Write an essay for the university's newspaper stating your opinion about whether or not homework should be required. Your goal is to convince your readers that your opinion is "right." Be sure to include the effect doing or not doing homework has on your learning, as well as other arguments to prove your opinion.

Appendix 2: Instructions for Writing Portfolios

Writing can be a powerful tool not only for communicating ideas but for improving language skills. Grammar, vocabulary, mechanics, coherence and cohesion are some of the aspects that students can develop through writing practice. One way of fostering appropriate teaching and assessment of writing is by keeping a portfolio of students' work.

What is a portfolio?

A portfolio is a purposeful collection of students' writing (or any other type of work). It is used to keep track of the process students go through when developing a writing piece. The portfolio contains drafts, revisions, and final written work.

Why keep a portfolio?

Through the portfolio we can do the following:

- Examine if students improve their writing over time
- Observe the writing process, not just the final product
- Promote students' self assessment
- Become aware of one's teaching writing practices
- Become aware of the connection between curriculum and instruction

How to keep the portfolio?

At the end of the course, you will hand in the portfolio to the Coordinator of Research. The writings need to evidence the steps described below:

1. **Selecting the topic:** File in the portfolio the prompt for the selected topic.
2. **Generating ideas:** On the first page of the portfolio you will find a sheet listing some techniques to get students started on the topic: brainstorming, listing, mind mapping, cubing, outlining, etc. Tick the one you used for the writing activity. Make sure the technique is reflected on the students' writings. Also, date the sheet.
3. **Drafting:** Have students write and hand in a draft. Make sure students write their names on the drafts.
4. **Revising:** Check the drafts using the writing conventions. Make a copy⁵ of the drafts, date them and file the ORIGINAL drafts in the portfolio.
5. **Editing:** Give students the checked copy of the drafts. Have them rewrite the draft based on your feedback.
6. **Scoring:** Collect edited versions and grade them using the appropriate rubric.
7. **Filing final product:** Make copies of the edited writings, date them, and file ORIGINAL versions in the portfolio. Give the copies back to students.

5 To make copies at no cost, please pick up an authorization slip at the Research office.

Appendix 3: Portfolio Rubric

Portfolio Rubric			
	Excellent	Satisfactory	Unsatisfactory
Prompt	All the elements required for a prompt ⁶ are clearly stated.	Most of the elements required for a prompt are present.	Few of the elements are present OR not clearly worded.
Congruence	The task directly relates to the writing standard(s) of the course.		Does not match standard or is not clear.
Idea generation	Explicit evidence of technique is used, clearly presented and elaborated. Technique sparks exceptional generation of ideas.	Technique clearly presented, but not fully elaborated. Technique elicits good generation of ideas.	Little evidence provided.
Use of conventions	Precise and appropriate. Shows clear understanding of the symbols.	Confuses some of the symbols OR inconsistent use of the symbols.	Inappropriate use of the symbols OR does not use them.
Scoring	Scores given for each aspect and descriptor of the rubric. Provides comments to help student understand the score.	Scores given only for each aspect of the rubric. Some useful comments are provided.	Scores given only for each aspect. No comments provided.
Overall presentation	All required portfolio elements are included. Properly dated and organized.	Some of the required elements are missing. Dates are sometimes omitted.	Numerous pieces omitted OR Material disorganized and hard to analyze.

⁶ Specification of genre or discourse mode, audience, and organizational plan (Muñoz et al., 2006).