Palimpsesto, 2004.
Collage sobre lienzo, 200 x 100 cm

Developing a Coherent System for the Assessment of Writing Abilities: Tasks and Tools*[1]

Mg. Ana Muñoz**
Dr. Jonathan Mueller***
Esp. Martha Álvarez****
Mg. Sandra Gaviria*****

This article discusses the design and validation of a writing assessment system. Two main processes are addressed: validation of the construct and the writing test prompts, and design and validation of scoring instruments. Construct validation included defining and aligning the construct, the standards, and the nature of the assessment tasks. The validation of tasks was determined by analyzing task content against a set of criteria to design appropriate writing tasks and by surveying students' understanding of the prompts. Finally, scoring rubrics were designed based on the standards and the construct. To validate the rubrics, their alignment with the construct and the standards was established, as well as the estimation of their inter-rater reliability ($r = > 0,7$).

Key words: writing standards, construct validity, writing prompts, inter-rater reliability, rubrics

Este artículo presenta el proceso de diseño y validación de un sistema para la evaluación de la escritura. Se discuten dos procesos centrales: validación del constructo y de las instrucciones de escritura en las pruebas evaluativas, y diseño y validación de las herramientas de evaluación. La validación del constructo requirió su definición y alineación con los estándares de escritura y con las tareas evaluativas. Para validar las tareas se analizó su contenido, respondiendo a un conjunto de criterios para el diseño apropiado de actividades de escritura y encuestando a los estudiantes participantes sobre su comprensión de las instrucciones. Por último, se diseñaron rúbricas como herramientas de medición, y su validez se determinó mediante la alineación con los estándares y el constructo, y la estimación de la confiabilidad interna entre evaluadores ($r > 0,7$).

Palabras clave: estándares de escritura, validez del constructo, instrucciones para la escritura, confiabilidad interna, rúbricas.

Cet article présente le développement et la validation d'un système d'évaluation de l'écrit. L'on présente deux processus essentiels qui sont la validation du construct linguistique

Ana Muñoz *et al*

et des tâches évaluatives puis l'élaboration et la validation des instruments évaluatifs. La validation du construct a été déterminée à travers sa définition et son alignement avec les standards d'écriture par niveau de suffisance, et avec les tâches évaluatives. Afin de valider ces dernières, nous avons analysé leur contenu en répondant à des critères pour le développement des activités évaluatives appropriées et en demandant aux étudiants leur compréhension des instructions écrites. Nous avons élaboré des grilles d'évaluation et leur validité a été définie selon leur alignement avec les standards, le construct et l'estimation de fiabilité interne chez les évaluateurs (r > = 0.7).

Mots clés: standards d'écriture, validité du construct, tâches évaluatives de l'écrit, fiabilité interne, rubriques

INTRODUCTION

The ability to write in English is an important skill both for educational and professional purposes. In an educational setting, writing can be used as a tool to monitor students' progress in a subject matter by having them reflect, analyze, and synthesize knowledge. In this context, writing can be used not only as a means for communication, but also as a support skill because it allows teachers to examine if students have the cognitive skills to be successful in an academic field (Bereiter and Scardamalia, 1987; Purves et al., 1984). Moreover, writing effectively in English is highly valued in the academic world because of the possibilities it offers teachers, students, and researchers to extend their intellectual production to international communities. Professionally, the need to write in English has become essential in today's global community because it allows citizens from different cultures to communicate through letters, e-mails, business reports, web pages, etc. (Weigle, 2002).

Whenever we talk about the need to develop a certain ability, teaching and assessment become imperative. Writing well is not a naturally acquired skill; it is a process that needs to be taught, practiced, and assessed. It is therefore of utmost importance that educational institutions emphasize the teaching and assessment of writing in their language programs.

The current study was undertaken at the Language Center, Eafit University, Medellín, Colombia. The teaching of writing within this context is carried out through classroom and at-home practice, and its assessment through essays, mid-term, and final examinations. However, there is not a consensus among teachers on how to teach and assess this skill. One way of bringing agreement to teaching and assessment practices, and fostering positive instructional practices is by designing assessment tasks and scoring instruments that are valid and reliable. It is widely recognized that well designed assessments in which there is authenticity of tasks, congruence between assessment and educational goals, detailed score reporting, teachers and students' understanding of the assessment criteria, among others, are beneficial to learning and teaching (Bailey, 1996; Hughes, 1989; Messick, 1996).

Although we are aware of the existence of commercially available writing assessment instruments, we set out to develop a classroom assessment system that would reflect our teaching and assessment beliefs and practices, that is, a system explicitly connected to the curriculum, serving both formative and summative purposes. In this sense, the measurement of student achievement becomes integral to learning, rather than imposed by some external assessment. According to Troman (1989), classroom-based assessment tends to be more democratic, diagnostic, professional-led, and with more focus on the process as compared to internationally standardized assessments.

In this article, we describe the process of designing and validating a writing assessment system. The article is divided in two main sections: 1) validation of the construct and the writing test prompts, and 2) design and validation of scoring instruments. In section one the following procedures will be explained:

Definition of the writing purposes
Definition of the writing construct
Specification of writing standards for the different levels of proficiency
Revision of existent writing test tasks (prompts)
Estimation of the validity of writing test tasks

Section two will describe the procedures to:
Design scoring instruments (rubrics)
Estimate the validity and inter-rater reliability of the scoring instruments

THEORETICAL FRAMEWORK AND RELATED STUDIES IN THE FIELD

Just as good writing requires a clear purpose, meaningful assessment begins with identifying the purpose or purposes of the assessment (Mueller, 2004; Wiggins, 1998). The purposes of an assessment should directly prescribe the standards, which are more specific statements of what students should know and be able to do in a certain domain such as writing. Research has found that standard-based assessment effectively facilitates teaching and learning (Guskey, 2001). It is critical that the standards align with the purposes of the assessment and that the assessment itself directly measures what is captured in the standards (Carr & Harris, 2001; Marzano, Pickering, & Pollock, 2001).

Meaningful standards which ask students to apply knowledge and skills are typically best addressed through authentic tasks that capture real world performance (Wiggins, 1998). For example, writing prompts –questions or statements students will address in their writing– that require students to engage in real world communication, more directly reflect standards of effective writing. Thus, to meaningfully assess whether students are capable of communicating in written form, first they must be given ample opportunities to practice such communication, and be authentically assessed through prompts and tasks that reflect the writing skills that are valued in educational and professional settings. Therefore, task 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: 60-61).

Writing assessment tasks may vary in the amount of specification or wording of the prompt. For instance, the prompt may include the **discourse mode** or purpose (Weigle, 2002) of the writing. Traditionally, four discourse modes have been identified: narration, description, exposition, and persuasion. 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 include the **stimulus material** or source material such as a short reading, graph, or drawing which provide the content for students to write about (Weigle, 2002). Further, writing prompts may refer to **the pattern of exposition** (Hale et al, 1996), or the specific instructions to the students as, for example, to make comparisons, draw conclusions, contrast, etc. And finally, the prompt may state 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. It is important to note that construct definitions vary according to the goals and needs of an institution. For instance, if a business is interested in making inferences about employees' ability to communicate effectively in writing with customers, the

definition of the construct may include: correct topical knowledge for dealing with trade, knowledge of specialized vocabulary, appropriateness of register, among other abilities. Based on the literature presented above and on the definition of its writing construct, the Language Center considers that prompts at the institution should:

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.

Finally, the judgment of student work is inevitably a subjective one on the teacher's part. To reduce teacher bias and increase the value of assessment, a clear set of criteria must be identified and then applied consistently to each student's samples of writing. Instructors have found that a well-designed rubric (or scoring scale) can provide such a tool in promoting accurate, reliable writing assessment (Stansfield & Ross, 1988; Weigle, 1994). Additionally, teachers need to be trained to consistently apply the rubric. It is important that teachers and students recognize that the writing prompts and tasks assigned directly assess meaningful standards of authentic writing relevant to their future goals, and that student writing is assessed along clearly articulated levels of performance for criteria that are aligned with those goals. Then teachers will be more motivated to change instructional practices to both teach, and have students practice around these authentic assessments, and students will be more likely to accept the value of such work (Natriello & Dornbusch, 1984). Furthermore, teachers and students will recognize the benefits of using the detailed feedback afforded by the rubric to develop and refine future instruction and learning.

It is therefore, of paramount importance that assessment be systematically developed according to specific guiding principles and explicit criteria to ensure that the construct being measured is as valid as the instruments designed to assess it. Poorly constructed assessment tools are likely to elicit unreliable and

unfair inferences, and, thus, be inadequate for decision-making. Consequently, the design and development of a writing assessment must be guided by a model that ensures its adequacy and utility. In other words, it is necessary to establish the reliability and the construct validity of such an instrument. Both test qualities will be briefly described in the remainder of this section.

Reliability is defined as consistency of measurement. It has been described as "the degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable and repeatable for an individual test taker" (Berkowitz, et al. 2000). The reliability of communicative language tests may be compromised given the qualitative rather than quantitative nature of communicative language assessment and the involvement of subjective judgments (Weir, 1990). Raters must agree on the marks they award and use the marking scheme in the way it was designed to be used. Sufficiently high rater reliability can only be obtained by means of proper training of the raters, the use of a functional rating scheme, and tasks that lend themselves to promoting agreement among raters. Prior to proceeding to the marking 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). As Green (2002) states, "in order to reconcile the problems of 'unreliable' judgments and the need for rigorous assessment it is necessary to develop a shared understanding of descriptions of performance." Therefore, the training and standardization of examiners to the procedures and scales employed require an accurate and unambiguous description of a set of criteria to assess the students' performance.

In addition to reliability, an assessment instrument must also have construct validity, which examines the degree to which the assessment instrument measures the language ability that it is supposed to measure. According to Messick (1996), the construct validation process includes the definition of the construct to be assessed and the nature of the assessment tasks. Bachman and Palmer (1996) point out that if the definition of the construct is not a complex one, high levels of reliability and construct validity can be expected. If the characteristics of the tasks are "relatively uniform," high levels of reliability can also be expected. To determine construct validity, according to Bachman and

Palmer, it is necessary to consider the extent to which the test tasks correspond to course content and instructional activities.

Messick (1989) describes two possible sources of invalidity: construct under-representation, and construct-irrelevant variance. The former indicates that the assessment tasks overlook important dimensions of the construct. The latter indicates that the assessment tasks contain too many variables, many of which are irrelevant (either too easy or too difficult) to the interpreted construct. Task design is therefore of utmost importance, since it may affect the interpretation of a test score.

A critical quality of assessment tasks is authenticity, defined as the degree of correspondence between the assessment tasks and the set of tasks a student performs in a non-test (instructional) situation (Bachman & Palmer, 1996), and also the interaction of the students' background language knowledge with the test task (Douglas, 2000; Bachman and Palmer, 1996; Widdowson, 1979). Therefore, assessment tasks must be authentic so that they (1) include all the important aspects of the theoretical construct(s) and (2) promote a positive, affective and cognitive response from the test taker. Language learners are more motivated when they are presented with situations faced in the real world and have to construct their own responses.

If assessment has high degrees of validity and reliability, then it will have enlightening effects on language curricula: the information obtained from assessment can become an integral part of instruction because it is possible to use the results for improvement and repair of instruction (Rea-Dickins & Germaine, 1992). Thus assessments are not viewed as a summative but rather as a formative part of instructional processes.

VALIDATION OF THE CONSTRUCT AND THE WRITING TEST PROMPTS

Participants and setting

The Language Center (LC) offers courses in English to children, adolescents, and adults. The current study focused on the adult and adolescent programs.

The children program was not considered because children are in the school period in which they are just developing reading and writing skills, therefore, requiring a different definition of the construct.

The adult and the adolescent English programs are comprised of 14 and 10 courses respectively, plus different advanced courses that focus on one of the four language abilities. These courses were carefully scaled according to the proficiency levels of the Common European Framework (Council of Europe, 2001). The Common European Framework (CEF) provides a common basis for the explicit description of standards, content, and assessment. The framework also defines levels of proficiency which allow learners' progress to be measured at each stage of learning. The table below shows the alignment of the CEF levels of language proficiency with the LC proficiency levels and courses.

Table 1. Alignment of CEF Proficiency levels with LC levels				
CEF Proficiency levels		Language Center proficiency levels	Language Center Adult Courses	Language Center Adolescent Courses
Proficient User	C2			
	C1	low advanced	5 advanced	
Independent User	B2	high intermediate	11, 12, 13, (14)*	5 topic-based courses
	B1	low intermediate	(7)*, 8, 9, 10	7, 8, 9, 10
Basic User	A2	high beginner	2, 3, 4, 5, 6	3, 4, 5, 6,
	A1	low beginner	N, 1	1, 2

  * (speaking and listening courses)

Three groups of students participated in the current study: one group from the adult program courses N-6 (low and high beginners), another group from the adult program courses 8-13 (low and high intermediate), and a third group from the adolescent program courses 1-10 (low and high beginners and low

intermediate). Courses 7 and 14 of the adult program rely on speaking and listening exclusively, therefore they were not included in the study.

Procedures

Definition of writing purposes

The study began by specifying the purposes of the LC writing tests. Two types of purposes were identified: 1. achievement: identifying the degree to which students have met specific instructional goals for decisions on grading, promotion and modification of instruction and curriculum at the classroom level; and 2. diagnosis: identifying students' strengths and weaknesses to tailor teacher's instruction to meet students' needs, and to assist and encourage student self-assessment.

Definition of the construct

Once the purposes were identified, the researchers defined the construct writing language ability or communicative language competence. The approach adopted by the LC to define writing ability derives from the work of Hymes (1972), Canale and Swain (1980), and Bachman (1990), who divide language knowledge into three types: linguistic knowledge, discourse knowledge, and sociolinguistic knowledge. Each competence was subsequently separated into different aspects as can be seen in Table 2 below.

Table 2. Taxonomy of Language Knowledge (Adapted from Grabe and Kaplan, 1996: 220)

I.   Linguistic competence
    A.   Knowledge of syntactic/structural patterns
    B.   Knowledge of vocabulary
    C.   Knowledge of the written code
        1.   Spelling
        2.   Punctuation
II. Discourse competence
    A.   Knowledge of cohesive devices
    B.   Knowledge of organizational structures
III. Sociolinguistic competence
    A.   Knowledge of functional uses of written language
    B.   Knowledge of register and situational parameters

The definition of the construct for the adult and adolescent programs includes the following components of language knowledge:

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 through the answer. Cohesion measures the students' ability to link ideas by using cohesive devices such as transitions and connectors. Such devices permit logical sequencing; they establish time frames for actions and events; they create 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 standards, through their knowledge of functional uses of the language and 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.

Definition of writing standards

The definition of the construct also involved the specification of writing standards for the English programs and the alignment of these standards with the writing standards proposed by the CEF. To construct the standards, we first considered the different domains -personal, public, occupational, and educational- described in the CEF. The domains contextualize the teaching

and assessment of languages within particular situations, in which the learners will need to perform. Based on the domains, we identified the discourse modes and genres that were most appropriate for the LC student population. Both the discourse mode and genres increase in difficulty as the proficiency level increases. We also reviewed the four types of written communication identified in the CEF - creative writing, reports/essays, correspondence, and notes/messages/forms - and, from each, we selected the standards that were most suitable for our context.

Revision of existent writing test tasks (prompts)

The estimation of task validity involved the collection of qualitative information to analyze the existing mid-term and final test tasks. Two academic coordinators analyzed all the prompts from these tests, 90 in the adult program (48 mid-term and 42 final exams) and 28 (14 mid-term and 14 final exams) in the adolescent program. The number of prompts in the adult program corresponds to existent tests forms A and B for each course in the program; in the adolescent program to form A tests. The analysis involved answering a set of 'Yes/No' questions (See Table 3) that incorporate the criteria to which the LC adheres in the design of appropriate writing prompts -questions 1 to 4- plus questions 5 to 8 (adapted from Reid and Kroll, 1995) that helped to complement the analysis.

Table 3. Question Guides to Analyze the Appropriateness of Prompts

Does the prompt:
1. Match the writing standards of the course?
2. Include the genre (biography, letter, etc.) or the purpose of the writing (narration, persuasion, etc)?
3. Include the audience, either implicitly or explicitly?
4. Include the organizational plan or form of presentation (steps, number of words, sequence, number of paragraphs, etc.)?

Is the content of the prompt:
5. Relevant to the students' lives (authentic)?

Is the topic of the prompt:
6. Easy to be accomplished within the assigned time? (approx. 20 min.)
7. Within the expertise, experience of the students?

Is the language of the instructions:
8. Clear?

Coordinators' opinion was analyzed using crossed-frequency tabulation as can be seen in Figures 1 and 2 below. These figures show the appropriateness of mid-term and final test prompts for the following categories: Relationship (prompt and writing standards), genre, audience, organizational plan, relevance, difficulty of the topic, expertise, and clarity of the instructions.

Figure 1. Coordinators' Opinion on the Appropriateness of Prompts - Adult Program

Figure 1 shows that most of the prompts from the adult program tests complied with the criteria for designing adequate prompts, with more than 80% of the prompts rated as appropriate by the coordinators for all but two categories: the relationship between the prompts and the standards, and the organizational plan. Specifically, 46.6% of the prompts (42 out of 90) were identified as not adequately related to the standards. Likewise, 23.3% of the prompts (21 out of 90) indicated poor organizational plan, that is, they lacked specificity in either the number of words, paragraphs, steps, or sequence to develop the writing. Consequently, the prompts were revised and adjustments made in these two areas to help students focus more. Here is an example of an inappropriate prompt and how it was modified:

Writing standard (from a low beginner course): "Describe a place to a friend or pen pal"

Inappropriate prompt: *Write about your dream house. What does it look like?*

Appropriate prompt: *Read this note from your friend, Elizabeth.*

Have a nice holiday!
Please send me a holiday postcard.
Tell me where you are, describe the place, the things you are doing, and what the weather is like.

Elizabeth

Write Elizabeth a postcard. Answer her questions about your holiday. Write 20 – 25 words.

The revised prompt is very specific about what the writer has to do and how he/she has to do it, whereas the inappropriate prompt does not provide the student with the necessary scaffolding to complete the task. To help students successfully fulfill a writing task, it is crucial to state the conditions or requirements of the task (audience, genre, steps, etc.). Specifying the conditions reduces the task complexity by structuring it into manageable chunks and thus increasing successful task completion.
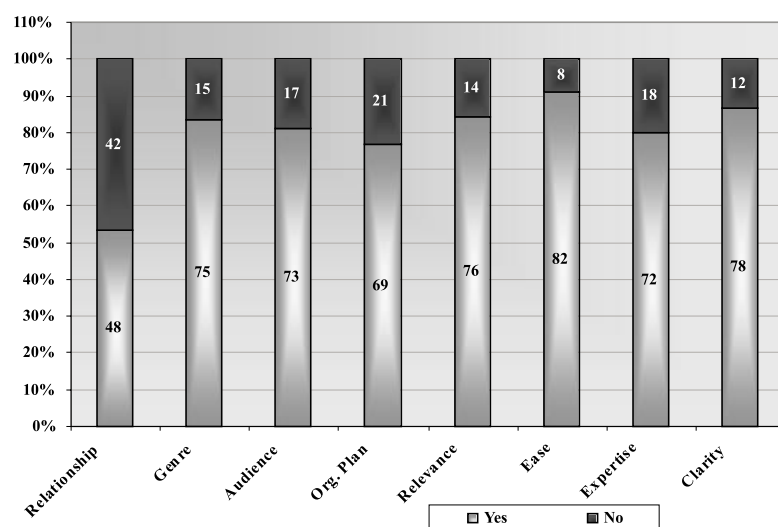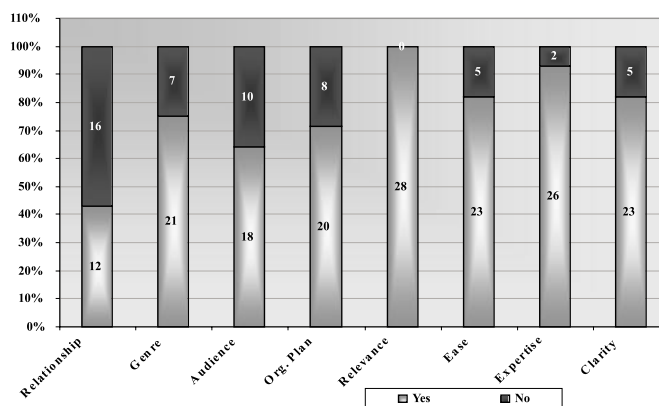

Figure 2. Coordinators' Opinion on the Appropriateness of Prompts - Adolescent Program

Figure 2 illustrates that the prompts from the adolescent program tests did not comply with the following criteria: Relationship, genre, audience, and organizational plan, with percentages lower than 80%. The relationship of the prompts with the standards was especially poor; more than half of the prompts (16 out of 28) did not match the standards. Lack of specificity of the audience or possible readers (35.7% of the prompts) may also affect the development of the writing. When writers have a specific audience in mind, they have to make ideas understandable to the reader. They also need to decide on the tone (formal/informal) to address the audience. By knowing the audience, the purpose (as stated by the standard), and the organizational plan, the writers can make their writing more focused and effective. Based on these results, adjustments were made in the aforementioned categories.

Data analysis for test prompts validation

Once the changes were implemented, the prompts were given to students for further analysis. One hundred and thirty seven students from courses N-13 of the adult program, and 126 from courses 1-10 of the adolescent program were given a survey in Spanish (Appendix 1) to examine their understanding of test prompts, their opinion on the difficulty of the topics given, and their opinion on the congruence between the writing tasks and the writing standards. Surveys were conducted after students had taken the writing section of the mid-term or final tests. Crossed-frequency tabulation was carried out to analyze the data.

Results and discussion

Students' survey - adult program

Students reported their opinion on a Likert scale regarding: 1. their understanding of prompts; 2. the degree of difficulty of the given topic; and 3. the relationship between the writing test tasks and writing standards.

Considering all the courses, results revealed that 22 out of the 137 students surveyed did not understand the prompts, either partially or completely (See Table 4). Nineteen students reported having problems understanding the prompts because of language interference (LI) and three students because of

their emotional state (Mood). Language interference covered aspects such as the wording of the prompt and unknown words. Additionally, 28 students reported that it was not easy to write on the topic assigned because the topic was unknown (UT), not interesting (NI), or because they lacked vocabulary to express ideas (LV).

Since one of the objectives of this study was to revise the writing prompts, we conducted a course-by-course analysis to determine which prompts needed to be redesigned. Results are presented only for the courses where three or more students reported some problems with the prompts.

Table 4. Understanding of Instructions – Adult program

Course	Opinion	LI	Mood	Frequency	Students per course
6	Partially	5	1	6	12
	Not at all	3	0	3	
12	Partially	3	1	4	10
	Not at all	0	0	0	
13	Partially	4	0	4	12
	Not at all	4	1	5	
Total Frequency		19	3	22	

Table 5. Difficulty of the topic – Adult program

Course	Opinion	UT	LV	NI	Frequency	Students per course
5	Partially	0	2	1	3	5
	Not at all	0	0	1	1	
6	Partially	1	5	0	6	12
	Not at all	1	2	1	4	
11	Partially	0	0	6	6	16
	Not at all	0	0	0	0	
12	Partially	2	0	1	3	10
	Not at all	1	0	0	1	
13	Partially	1	2	0	3	12
	Not at all	0	1	0	1	
Total Frequency		6	12	10	28	

As can be seen in Tables 4 and 5, both instructions and topic for courses 6, 12, and 13, prompts need to be improved. On a general basis, a representative number of students in courses 6, 12, and 13 reported partial or no understanding of the prompts due to language interference (LI). Some students in these courses also claimed having difficulties writing on the assigned topic because they did not have enough vocabulary to express their ideas (LV) and considered that the topic was either unknown (UT) or not interesting (NI). Additionally, prompts for courses 5 and 11 may also need refinements since some students reported difficulty in writing due to LV and lack of interest in the topic.

In general, most of the difficulties, either understanding the instructions or developing the writing tasks, were related to lack of language knowledge. Students considered that they lacked vocabulary to express their ideas or that the wording of the prompts contained unknown vocabulary. Therefore, it will be necessary to revise the prompts and adjust the language to the students' level of proficiency. Regarding aspects such as mood or interest in the topic, it was not within the scope of the study to investigate what specific emotional variables may have interfered with the development of the writing.

Table 6. Relationship Between Test Tasks and Writing Standards– Adult program

Course	Partially	Not at all	Students per course
6	3	5	12
11	4	6	16
13	4	2	12
Total	11	13	

Table 6 presents students' opinion on the relationship between the test task and the writing standards of the course. Eight students out of twelve, in course six considered that the writing task was partially or not related to the standards. In course 11, ten students, out of sixteen, agreed that the task needed to be connected to the standards. Half of the students in course 13 also found the task to be unconnected to the standards. It is possible that students who reported these discrepancies could also have had difficulties when writing due to lack of vocabulary or lack of interest in the assigned topic, as described in Table 5.

Students' survey - adolescent program

The same survey given to adult students was used for the adolescent program. Results revealed that 31 out of 126 students surveyed did not understand the prompts either partially or completely. Twenty-eight students reported lack of understanding due to language interference and three due to emotional state. Additionally, 42 students reported that it was not easy to write on the assigned topic either because it was unknown (UT) or not interesting (NI) or because they lacked vocabulary (LV) to express their ideas (Tables 7 and 8). It might be possible that students' difficulties are related to age and background knowledge factors. Nonetheless, it is necessary to revise the meaningfulness of prompts to students' lives.

Table 7. Understanding of Instructions

Course	Opinion	LI	Mood	Frequency	Students per course
1	Partially	2	1	3	7
1	Not at all	0	0	0	7
2	Partially	6	1	7	29
2	Not at all	4	0	4	29
3	Partially	2	0	2	11
3	Not at all	1	0	1	11
4	Partially	5	0	5	26
4	Not at all	2	0	2	26
5	Partially	3	0	3	17
5	Not at all	1	0	1	17
7	Partially	2	1	3	13
7	Not at all	0	0	0	13
Total Frequency		28	3	31	

Table 8. Difficulty of the topic – Adolescent Program

Course	Opinion	UT	LV	NI	Frequency	Students per course
1	Partially	0	3	0	3	7
	Not at all	0	0	0	0	
2	Partially	0	8	0	8	29
	Not at all	0	3	1	4	
3	Partially	0	3	0	3	11
	Not at all	0	2	0	2	
4	Partially	1	5	0	6	26
	Not at all	1	1	1	3	
5	Partially	0	1	0	1	17
	Not at all	1	0	1	2	
7	Partially	0	2	0	2	13
	Not at all	1	2	1	4	
10	Partially	2	0	2	4	12
	Not at all	0	0	0	0	
Total Frequency		6	30	6	42	

Tables 7 and 8 illustrate that most of the prompts of the adolescent program test tasks were not understood because of language knowledge. Students considered that the vocabulary used in the prompts was unknown, especially in courses 2, 4 and 5. Language knowledge also interfered with students' ability to write on the given topic. Thirty students reported that they did not have enough vocabulary to express their ideas, especially in courses 2, 3, 4, and 7.

Table 9. Relationship Between Test Tasks and Writing Standards – Adolescent Program

Course	Partially	Not at all	Students per course
4	3	2	26
5	2	1	17
6	3	1	7
7	1	2	13
Total	11	6	

Ana Muñoz *et al*

Table 9 presents students' opinion on the relationship between the test tasks and the writing standards of the course. Eleven students reported partial connection between the test tasks and the standards and six indicated that they were not connected at all. Consequently, the writing tasks for these courses need to be revised.

DESIGN AND VALIDATION OF SCORING INSTRUMENTS

Based on the writing standards, two assessment rubrics were designed for the adult English program: one rubric for courses N-6, corresponding to levels A1 and A2 of the CEF (Basic Users), and another rubric for courses 8-13 (except for course 7 and 14 whose focus is on speaking and listening), corresponding to levels B1 and B2 of the CEF (Independent Users), For teachers' reference, on the back of each rubric assessment, specifications per course are provided: standards, grammar and vocabulary, genre, discourse mode, and expected written production (See Appendices 2 & 3). The adult program rubric for Basic Users was slightly modified for courses 1-10 of the adolescent program. This decision was based on the idea that adolescents at beginner and intermediate levels do not have the same academic and personal background that adults in beginning levels may have. Likewise, their strategic and pragmatic competences are different from adults at the same level. Adolescents are unprepared to write the way adults do, not because they are slow, indifferent to or incapable of academic excellence, but because they are novice learners. Therefore a Basic User rubric may be a suitable instrument for the type of knowledge and abilities students have at this level.

The validity of the scoring instrument and the construct was estimated by establishing: 1. the alignment of the rubrics with those components of writing that were included in the definition of the construct, 2. the connection of the rubrics to the CEF, 3. the alignment of the rubrics with the standards and tasks, and 4. the estimation of inter-rater reliability.

Alignment of the rubric with the construct

By defining the construct, we selected the aspects of writing that were included in the rubrics: Coherence and cohesion, grammar and vocabulary, and task

completion. The same aspects are measured along the different levels of proficiency. However students' output requirements change as they move from one level to the other as can be seen in the descriptions below:

A1 students –low beginners at the Language Center– should be able to write one paragraph using short simple sentences linked with connectors such as 'and,' 'or,' 'but.' They should also be able to control a few simple grammatical structures and sentence patterns and have a basic repertoire of words and simple phrases related to personal details and particular concrete situations. Students at this level can produce different text types or genres: Personal letters, personal E-mails, forms, and postcards.

A2 students –high beginners at the Language Center– should be able to write short, simple pieces (from one to two paragraphs) using connectors to link sentences such as: 'because,' 'also,' 'first,' 'then,' 'later,' 'finally,' 'for example,' 'in conclusion,' 'for that reason,' 'consequently,' 'compared with,' 'similarly,' 'in contrast,' 'on the contrary.' They should be able to combine simple structures correctly and groups of words and formulae in simple everyday situations. Students at this level should be able to write simple notes, instructions, biographies, questionnaires, simple reports, and simple survey reports.

B1 students –low intermediate at the Language Center– should be able to write continuous, intelligible text –from two to three paragraphs– in which elements are connected using cohesive devices like: 'to summarize,' 'therefore,' 'as a result,' 'since,' 'on the other hand,' 'besides,' 'in addition,' 'however,' 'such as,' 'meanwhile,' 'in other words,' 'with regard to.' Students should also be able to use a reasonably accurate repertoire of grammatical structures and sufficient vocabulary to express ideas on familiar topics. Students should be able to write in genre such as letters, personal E-mails, reports, simple essays, short stories.

B2 students –high intermediate at the Language Center– should be able to write from three to four paragraphs on different topics and purposes using a variety of linking words to mark clearly the relationships between ideas, using: 'For,' 'so that,' 'although,' 'in order to,' 'in fact,' 'moreover,' 'nevertheless,' 'in

reality,' 'otherwise,' 'to sum up,' 'curiously,' 'coincidentally,' (un)fortunately,' 'ironically,' 'in theory,' 'obviously,' 'presumably.' Students should also demonstrate a good grammatical control without having to restrict what they want to write, adapting a level of formality appropriate to the circumstances and use circumlocution and paraphrasing to cover gaps in vocabulary and structure. Students should be able to write reports, notes, notices, diary entries, guides, essays, and guided articles.

Alignment of the rubrics with the CEF

The communicative language competence model proposed by the CEF (2001:13) is based on three components: linguistic, sociolinguistic and pragmatic (also known as discourse competence). The language aspects contained in the LC rubrics derive from these competences as explained in the Design and Results section above.

Alignment of the rubrics with the standards and tasks

We compared the CEF writing scales with existing LC writing standards. First of all, we aligned our standards with the CEF proficiency levels which gave us an idea of what writing proficiency students will have once they finish the English programs. We then distributed the LC standards into the different adult and adolescent programs.

As we have already explained, the aspects contained in the rubrics give an account of the language competences considered in our teaching approach. As a result, we made sure that the standards relate to those competences and, therefore, to the aspects of the rubrics. Additionally, the suggested writing tasks were designed based on the standards so that students' written performance can accurately be measured.

Estimation of inter-rater reliability

Inter-rater reliability refers to the predisposition of different raters to give the same scores to the same samples of writing. Therefore, we invited three different groups of teachers –three teachers in each – to score the same pieces of writing.

Prior to the study, the teachers participated in two calibration sessions where the rubric descriptors were reviewed and writing samples scored. Results were then compared and discussions carried out in order to reach consensus on the assessment criteria.

To collect the data on inter-rater reliability, the first group scored writing samples from courses N - 6 of the adult program, the second group scored samples from courses 8 – 13 of the adult program and the third group scored samples from courses 1-10 of the adolescent program. The writing samples were taken from students' actual performance either in mid-term or final tests. The writings were transcribed and given to teachers to score on the aspects of the rubrics (See one sample in Appendix 4). Each of the three teachers in the first group marked 24 writing samples using the 1 - 6 rubric of the adult English program, for a total of 72 scoring samples. In the second group, each teacher scored 23 samples using the course 8 - 13 rubric of the adult English program, for a total of 69 samples. Using the rubric for courses 1-10 of the adolescent program, the third group of teachers marked 20 writing samples each, for a total of 60 samples.

Data analysis for rubric validation

The scored writing samples were analyzed using ANOVA (analysis of variance) and Pearson correlation. ANOVA can be used to compare the scores given by a set of raters and to determine whether there is any statistical difference between the mean scores; that is, "if some raters tend to give higher or lower scores than other raters" (Weigle, 2002: 135). Correlation can be used to indicate the strength of the relationship between sets of scores.

Results and discussion

As can be seen in Tables 10, 11, and 12 there were no significant differences among the mean scores given by each evaluator to the different language aspects in the two programs. For instance, the mean scores for coherence and cohesion in courses N – 6 (Table 10) were:

Evaluator 1 = 3.22
Evaluator 2 = 3.27
Evaluator 3 = 3.18

There are no statistical differences between these mean scores (P-value = 0.9084). The other aspects also have P-values higher than 0.10 meaning that teachers assigned similar scores, and even, in some cases, the same scores. The correlations were also moderately high – almost all above 0.7 – indicating that there is consistency in the assignation of scores. It is worth noting that when this study was conducted the rubric for courses 8 – 13 did not include spelling as a separate aspect but as part of grammar and vocabulary, consequently, this aspect was not contemplated in Table 11.

Table 10. Inter rater Reliability for Rubric Courses N – 6 Adult Program

Aspect	Means and 95% Confidence Intervals	Pearson Correlation
| Coherence and Cohesion | ```
Evaluator Mean Stnd. error Lower Upper
 score limit limit
_____
Ev1 3.22 0.1378 2.944 3.514
Ev2 3.27 0.1311 2.999 3.542
Ev3 3.18 0.1339 2.910 3.464

_____
P-value = 0.9084
``` | ```
 Ev1 Ev2
----------------------
Ev2 0.8279
 0.0000
Ev3 0.9044 0.7510
 0.0000 0.0000
----------------------
Correlation
P-Value
``` |
| Grammar and Vocabulary | ```
Evaluator Mean Stand. error Lower Upper
 score limit limit
_____
Ev1 2.97 0.1606 2.646 3.311
Ev2 3.10 0.1378 2.819 3.389
Ev3 2.97 0.1063 2.759 3.199

_____
P-value = 0.7578
``` | ```
 Ev1 Ev2
----------------------
Ev2 0.8839
 0.0000
Ev3 0.9007 0.8408
 0.0000 0.0000
----------------------
Correlation
P-Value
``` |
| Spelling | ```
Evaluator Mean Stnd. error Lower Upper
 Score limit limit
_____
Ev1 2.85 0.1104 2.625 3.082
Ev2 3.06 0.1177 2.818 3.061
Ev3 3.04 0.0947 2.845 3.237

_____
P-value = 0.3301
``` | ```
 Ev1 Ev2
----------------------
Ev2 0.7613
 0.0000
Ev3 0.8037 0.8014
 0.0000 0.0000
----------------------
Correlation
P-Value
``` |

<table>
<tr><td rowspan="2">Task Completion</td><td>

```
Evaluator Mean Stnd. error Lower Upper
 score limit limit
_____
Ev1 3.12 0.1320 2.851 3.398
Ev2 3.41 0.0935 3.223 3.610
Ev3 3.25 0.1206 3.000 3.499

_____

P-value = 0.2136
```

</td><td>

```
 Ev1 Ev2
---------------------
Ev2 0.8064
 0.0000
Ev3 0.8832 0.6835
 0.0000 0.0000
---------------------
Correlation
P-Value
```

</td></tr>
</table>

Table 11. Inter rater Reliability for Rubric Courses 8 – 13 Adult Program

Aspect	Means and 95% Confidence Intervals	Pearson Correlation
Coherence and Cohesion	```Evaluator Mean Stnd. error Lower Upper score limit limit _____ Ev4 3.06 0.1188 2.818 3.311 Ev5 2.97 0.1735 2.618 3.338 Ev6 2.97 0.1706 2.623 3.332 _____ P-value = 0.9022```	``` Ev4 Ev5 -------------------- Ev5 0.6158 0.0018 Ev6 0.5511 0.8667 0.0064 0.0000 -------------------- Correlation P-Value```
Grammar and Vocabulary	```Evaluator Mean Stnd. error Lower Upper score limit limit _____ Ev4 2.91 0.1118 2.681 3.144 Ev5 2.67 0.1281 2.408 2.939 Ev6 2.73 0.1535 2.420 3.057 _____ P-value = 0.4224```	``` Ev4 Ev5 -------------------- Ev5 0.8474 0.0001 Ev6 0.7669 0.8714 0.0003 0.0000 -------------------- Correlation P-Value```
Task Completion	```Evaluator Mean Stnd. error Lower Upper score limit limit _____ Ev4 3.36 0.1052 3.151 3.587 Ev5 3.19 0.1286 2.928 3.462 Ev6 3.13 0.1226 2.876 3.384 _____ P-value = 0.3474```	``` Ev4 Ev5 -------------------- Ev5 0.8519 0.0000 Ev6 0.8253 0.7096 0.0000 0.0001 -------------------- Correlation P-Value```

Table 12. Inter rater Reliability for Rubric Courses 1 - 10 Adolescent Program

Aspect	Means and 95% Confidence Intervals	Pearson Correlation
| Coherence and Cohesion | ```
Evaluator Mean Stnd. error Lower Upper
 score limit limit
_____
Ev1 3.10 0.2071 2.666 3.533
Ev2 3.12 0.2402 2.622 3.627
Ev3 2.85 0.2149 2.400 3.299

_____

P-value = 0.6258
``` | ```
 Ev1 Ev2
------------------
Ev2 0.9387
 0.0000
Ev3 0.9194 0.8345
 0.0000 0.0001
------------------
Correlation
P-Value
``` |
| Grammar and Vocabulary | ```
Evaluator Mean Stnd. error Lower Upper
 score limit limit
_____
Ev1 3.05 0.2309 2.566 3.533
Ev2 3.02 0.1685 2.672 3.377
Ev3 3.30 0.2020 2.877 3.722

_____

P-value = 0.5710
``` | ```
 Ev1 Ev2
------------------
Ev2 0.9492
 0.0000
Ev3 0.8884 0.7845
 0.0000 0.0000
------------------
Correlation
P-Value
``` |
| Spelling | ```
Evaluator Mean Stnd. error Lower Upper
 score limit limit

_____
Ev1 3.32 0.1592 2.991 3.658
Ev2 3.30 0.1597 2.965 3.634
Ev3 3.22 0.1830 2.841 3.608

_____

P-value = 0.9084
``` | ```
 Ev1 Ev2
------------------
Ev2 0.8584
 0.0000
Ev3 0.9511 0.7557
 0.0000 0.0001
------------------
Correlation
P-Value
``` |
| Task Completion | ```
Evaluator Mean Stnd. error Lower Upper
 score limit limit

_____
Ev1 3.67 0.1787 3.300 4.0490
Ev2 3.77 0.1641 3.431 4.1185
Ev3 3.67 0.1823 3.293 4.0566

_____

P-value = 0.8973
``` | ```
 Ev1 Ev2
------------------
Ev2 0.8780
 0.0000
Ev3 0.9195 0.7945
 0.0000 0.0000
------------------
Correlation
P-Value
``` |

Conclusions and implications

This article has discussed the process of designing and validating a writing assessment system. Two main topics were addressed: validation of the construct and the writing test prompts, and design and validation of scoring instruments.

The definition of the construct required, first, the specification of the LC writing purposes, second, the specification of language components to be included in the construct, and third, the specification of the writing standards for each level of proficiency. The validity of the construct was demonstrated by aligning the LC standards with the CEF writing standards, language domains, and types of written communication, without disregarding the local needs and context in order to guarantee that the tasks are representative of the type of writing that students will need in real-life situations.

The revision and validation of tasks were assessed by analyzing task content against a set of criteria to design appropriate writing assessment tasks, and by establishing whether students understood the prompts. Results of the test tasks analysis by coordinators revealed that some tasks in the adult and adolescent programs needed to be redesigned. The areas requiring improvement were the relationship between tasks and course standards, the organizational plan or form of presentation, the genre, and the audience. Lack of specification of these areas may cause students' misunderstanding of the prompts: "Students may be misled by vague instructions, a directionless topic or the fact that the task cannot be completed in the time allotted" (Hyland, 2003:224). Therefore, the redesign of the prompts needs to be guided by a set of criteria that limit the possibilities of misunderstandings. Students will then have a clearer direction and more rhetorical structure when writing. It is crucial that teachers explicitly know what they want from the assessment because instructions can be more clearly communicated and task performance can be more easily assessed.

The revised tasks were given to students for further analysis. Results showed that the majority of the students understood the prompts and found it easy to write on the assigned topics. However, some students in both programs

reported difficulties regarding the interference of vocabulary in understanding the instructions and developing the tasks. It is possible that the vocabulary used in the prompts was too complex for the students' level of proficiency. Consequently, we need to examine whether the vocabulary used in the prompts matches the required lexica for each of the courses and standards and make changes accordingly. Additionally, adolescent students reported lack of interest in the topic, perhaps because they perceived it is irrelevant to their lives. Consequently, topics that are more meaningful to students' lives need to be considered when designing prompts for the adolescent program.

Similarly, some students related the misunderstanding of the prompts and the difficulty to write on the given topics to affective factors such as mood and lack of interest. Undoubtedly, students' performance may be influenced by their physical or emotional state. However, the findings of this research did not provide information on the specific factors that may affect writing: students' own competences, individual characteristics, or the conditions under which the task is carried out. Further research is therefore needed into the affective domains that influence writing.

Finally, the design of the scoring rubrics was based on the writing standards and the construct. To validate the rubrics, their alignment with the construct, the CEF, and the LC writing standards was established. Validation of the rubrics also required the estimation of inter-rater reliability. Results in this area indicated that raters' scores were in agreement ($r > 0.7$) when applying the relevant rubric, meaning that the rubrics allow for consistent marking.

In light of the findings further development and investigation of the programs will include:

Revision and design of writing tasks based on the criteria for appropriate writing task development.

Training of teachers on teaching and assessing writing. An in-service training program will be offered to instruct teachers how to design writing tasks, and how to reliably score students' writing. In addition, teachers also need to be trained on how to provide enough practice based on a process-oriented approach to writing.

Ensuring that vocabulary corresponds to the required lexica for each of the courses and standards, and that the language used in prompts does not exceed students' knowledge of language.

Examination of the effect of writing assessment on the teaching and learning process. Once we implement the writing assessment system we will start investigating its impact on instructional practices.

In sum, the process of construct, task, and rubric validation allowed us to design a coherent system for the assessment of writing. We hope that the procedures undertaken in this study, may serve as a model for the design of valid assessments at different language teaching institutions in Colombia. In our country the area of evaluation is rapidly moving from traditional to alternative modes of assessment that are geared more toward thinking and away from rote memory. Therefore, there is a need for assessments that adequately represent the curriculum, the constructs, and the teaching and learning philosophy of the institutions, and that promote accurate and reliable inferences about students' performance.

REFERENCES

Alderson, J. C. & D. Wall, 2001, *Language test construction and evaluation,* Cambridge, Cambridge University Press.

Bachman, L. & A. Palmer, 1996, *Language testing in practice: Designing and developing useful tests*, Oxford, Oxford University Press.

Bachman, L., 1990, *Fundamental considerations in language testing*, Oxford, Oxford University Press.

Bailey, K. 1996, "Working for washback: a review of the washback concept in language testing", *Language Testing* 13, 257–79.

Bereiter, C. and M. Scardamalia, 1987, *The psychology of written composition*, Hillsdale, NJ, Lawrence Erlbaum Associates.

Berkowitz, D. et al., 2000, *The use of texts as part of High-Stakes decision-making for students: A Resource Guide for educators and policy-makers,* Washington, DC, U.S. Department of Education.

Canale, M. and M. Swain, 1980, "Theoretical bases of communicative approaches to second language teaching and testing", *Applied Linguistics* 1*,* (1), 1- 47.

Carr, J. F., & D. E. Harris, 2001, *Succeeding with standards: Linking curriculum, assessment, and action planning*, Alexandria, VA, ASCD.

Council of Europe, 2001, *Common European framework of reference for languages: Learning, teaching, assessment*, Cambridge, Cambridge University Press.

Douglas, D., 2000, *Assessing language for specific purposes*, Cambridge, Cambridge University Press.

Grabe, W. and Kaplan, R. B., 1989, "Writing in a second language: Contrastive rhetoric". In D. M. Johnson and D. H. Roen (eds.), *Richness in writing*. New York and London: Longman.

Green, S., 2002, Criterion referenced assessment as a guide to learning – The importance of progression and reliability. *A paper presented at the association for the study of evaluation in education*. Southern Africa International Conference.

Guskey, T. R., 2001, "Helping standards make the grade", Educational Leadership, 59, 20-27.

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.

Hughes, A. 1989, *Testing for language teachers,* New York: Cambridge University Press.

Hyland, K., 2003, *Second language writing,* Cambridge: Cambridge University Press.

Hymes, D. 1972, "On communicative competence", in J. Pride and H. Holmes (eds.), *Sociolinguistics*, NY, Penguin, 269 – 93.

Marzano, R. J., et al., 2001, *Classroom instruction that works: Research-based strategies for increasing student achievement*, Alexandria, VA, ASCD.

Messick, S. 1996, "Validity and washback in language testing", *Language Testing 13(3),* 241-256.

Messick, S.**,** 1989, "Meaning and values in test validation: the science and ethics of assessment", *Educational Researcher* 18 (2), 5-11.

Mueller, J., 2004, *Authentic Assessment Toolbox*, <http://jonathan.mueller.faculty.noctrl.edu/toolbox> [Retrieved: March, 2005]

Natriello, G., & S. M. Dornbusch, 1984, *Teacher evaluative standards and student effort*, New York, Longman.

Purves *et al.,* 1984, "Towards a domain-referenced system for classifying assignments", *Research in the Teaching of English*. 18 (4), 385 – 416.

Rea-Dickins, P. & K. Germaine, 1992, *Evaluation*, Oxford: Oxford University Press.

Reid, J. and B. Kroll, 1995, "Designing and assessing effective classroom writing assignments for NES and ESL students", *Journal of second language writing* 4 (1), 17-41.

Stansfield, C., J. & Ross, 1988, "A long-term research agenda for the Test of Written English", *Language Testing,* 5, 160-186.

Troman, G. 1989, "Testing tension: The politics of educational assessment", *British Educational Research Journal*, 15, 279-295

Weigle, Sara, 2002, *Assessing Writing*, Cambridge, Cambridge University Press.

Weigle, S., 1994, "Effects of training on raters of ESL compositions", *Language Testing,* 11, 197-223.

Weir, C. J., 1990, *Communicative language testin*g, London, Prentice Hall.

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

Wiggins, G., 1998, *Educative assessment: Designing assessments to inform and improve student performance*, San Francisco, CA, Jossey-Bass Publishers.

THE AUTHORS

**   Ana Muñoz is the Research and Teacher Education Coordinator at the Language Center, EAFIT University. She holds an M.A. in TESOL

from Eastern Michigan University, USA and has a B.A. in Languages – English and French from Universidad Pontificia Bolivariana, Medellín, Colombia. E-mail: apmunoz@eafit.edu.co

*** Jonathan Mueller is a professor of psychology at North Central College, Illinois. He holds a Ph.D in Personality and Social Psychology from Northern Illinois University. For more than ten years he has consulted with teachers, schools and districts, at the K-12 and post-secondary levels, on the development, evaluation and revision of assessments, goals and standards. E-mail: jfmueller@noctrl.edu

**** Martha E. Alvarez is an Industrial Engineer from Universidad Nacional, Medellín. She holds a postgraduate degree in Information Systems from Universidad EAFIT and has been a teacher at the School of Engineering in this institution since 1992. She has also taught and coordinated the Statistics I and II programs. E-mail: ealvarez@eafit.edu.co

***** Sandra Gaviria is the Coordinator of the Adult English Program and researcher at the Language Center, EAFIT University. She holds an M.A. in TESOL from West Virginia University, USA and has a B.A. in Modern Languages from Universidad Pontificia Bolivariana, Medellín, Colombia. E-mail: sgaviria@eafit.edu.co

APPENDIX 1

Curso_____   Examen_____ Parcial_____ Final_____ Formato____

Queremos saber su opinión acerca de la actividad en la sección de ***WRITING*** que acaba de realizar en el examen. Por favor, para cada pregunta, elija la opción que mas se adecua a su experiencia. Califique de 1 a 5, siendo:

5 completamente de acuerdo y **1 completamente en desacuerdo.**

	1	2	3	4	5
1. Entendí claramente las instrucciones para realizar el escrito					
2. Me pareció fácil escribir sobre el tema propuesto					
3. La actividad de escritura propuesta se relaciona con los temas vistos en clase					

Si respondió 1, 2 ó 3 en la pregunta 1, elija la razón para no entender las instrucciones:

[ ]Contenía palabras desconocidas
[ ]La redacción no era entendible
[ ]El lenguaje utilizado en las instrucciones era muy avanzado
[ ]Mi estado de ánimo me impidió entender bien las instrucciones
[ ] Otra. ¿Cuál?_____

Si respondió 1, 2 ó 3 en la pregunta 2, elija la razón por la que le pareció difícil escribir sobre el tema:

[ ] El tema me era desconocido o demasiado especializado
[ ] El tema no me pareció interesante
[ ] No encontré cómo expresar lo que quería decir
[ ] Otra. ¿Cuál?_____

GRACIAS POR SU VALIOSA COLABORACIÓN

APPENDIX 2

Rubrica para la evaluación de la escritura - Cursos N, 1, 2, 3, 4, 5, 6 (ADULTOS)

Aspectos / nota	5	4	3	2	1
Coherencia y Cohesión	a. El texto es desarrollado, **en su totalidad,** de una manera lógica y secuencial. b. Utiliza conectores y transiciones básicas **de manera excepcional.**	a. **La mayor parte** del texto es desarrollado de una manera lógica y secuencial. b. La **mayor parte** del texto es conectado mediante el uso variado y apropiado de conectores y transiciones básicas	a. Tiene **dificultades** para desarrollar el texto de una manera lógica y secuencial b. En **algunas partes** del texto hay un uso variado y adecuado de conectores y transiciones básicas.	a. El desarrollo del texto es en su mayor parte **ilógico** y falto de secuencia b. Uso **mínimo** de conectores y transiciones básicas.	a. El texto **no** tiene un desarrollo lógico ni secuencial. b. **No hay uso de conectores ni transiciones.**
Gramática y Vocabulario	a. El texto presenta un **control total** de estructuras requeridas para el curso. b. Utiliza **gran** variedad de vocabulario básico en contextos familiares	a. **La mayoría** del texto presenta un uso adecuado de las estructuras requeridas para el curso. b. La **mayor parte** del vocabulario básico es variado y **apropiado** al contexto	a. El texto presenta un **control parcial** de estructuras requeridas para el curso b. Utiliza un rango **limitado** de vocabulario y **algunas veces inapropiado** para el contexto.	a. **Control mínimo** de las estructuras requeridas para el curso. b. Es evidente la **falta** de vocabulario y su **frecuente** uso inapropiado.	a. **No** utiliza las estructuras del curso. b. El vocabulario se limita a unas cuantas palabras usadas **inapropiadamente.**
Ortografía	El texto es fácil de leer gracias a: a. **Excelente** dominio de la puntuación b. **Excelente** dominio de la ortografía.	a. **Buen** dominio de la puntuación b. Comete **pocos** errores de ortografía que no causan interferencia con la comprensión del texto.	a. **Inconsistente** uso de la puntuación b. Algunas ideas del texto son **un poco confusas** debido a inconsistencias ortográficas.	a. Comete errores **muy frecuentes** en la puntuación b. El texto **es en su mayor parte** incomprensible debido a la incorrecta ortografía.	a. Errores **severos** y **persistentes** en la puntuación. b. Hay interferencia con la claridad y comprensión del texto debido a la incorrecta ortografía.

Cumplimiento de la Tarea	a. **Todos** los elementos requeridos por el prompt están presentes b. El escrito **no** presenta irrelevancias o desviaciones del tema c. **Excepcional** elaboración y detalles para ilustrar las ideas.	a. Desarrolla unas partes del prompt mas que otras b. El escrito presenta mínimas irrelevancias o desviaciones del tema. c. **Buena** elaboración y detalles para ilustrar las ideas.	a. **Omite algunos** de los requerimientos del prompt b. El escrito presenta **algunas** irrelevancias o desviaciones del tema. c. El desarrollo y la elaboración de las ideas es **superficial.**	a. **Pocos** de los requerimientos del prompt están presentes b. El escrito presenta **muchas** irrelevancias o desviaciones del tema. c. **No** desarrolla o elabora las ideas.	a. La realización del escrito **no** tiene relación con los requerimientos del prompt.

*Ver requisitos de evaluación en el reverso

For the Eafit Language Center Use ONLY - 2006

ASSESSMENT SPECIFICATIONS – Adult Program

	WRITING STANDARDS	Grammar & Vocabulary	Genre	Discourse Mode	Output Requirement
N	• Provide personal information (e.g. name, telephone number, address, and occupation)	Verb "Be" Body parts, simple illnesses, parts of the house, the colors, classroom objects, places in the city, occupations, days of the week, familily members	Personal letter Personal E-mail Form Postcard	• Informative • Descriptive	■ One paragraph Short simple sentences linked with connectors such as 'and,' 'or,' 'but.' Control of a few simple grammatical structures and sentence patterns Basic repertoire of words and simple phrases related to personal details and particular concrete situations
1	Provide self, family members or friends' personal information	*Verb be, Subject pronouns, Possessive adjectives, Questions using verb be, Indefinite articles, Verb be + age, Present simple, Yes/no/ Wh questions, Subject + be* for description. Numbers 1–20, Alphabet, Nationalities, Months, Adjectives			

	Functions	Grammar and vocabulary	Text types	Text modes	Writing
2	• Describe a familiar place • Describe physical appearance	There is/are, Some and any, Adverbs of frequency, Can and can't, Present progressive, Prepositions of place, Present progressive as Future. Room objects, Colors, Adjectives, Days of the week, Time words, Adverbs of frequency, Clothes, Weather	Personal letters Personal E-mails Postcards Simple notes Instructions Biographies Questionnaires Simple reports Survey reports	• Informative • Descriptive • Narrative	• **One to two paragraphs** Short, simple pieces of writing using connectors to link sentences such as: 'because,' 'also,' 'first,' 'then,' 'later,' 'finally,' 'for example,' 'in conclusion,' 'for that reason,' 'consequently,' 'compared with,' 'similarly,' 'in contrast,' 'on the contrary.' Use of simple structures correctly Use of groups of words and formulae in simple everyday situations
3	• Describe one's leisure and everyday activities • Express thanks and apology	Count/non-count nouns, Imperatives, Would like, Past simple regular and irregular verbs, Going to, Should (not). Foods, Recipes and cooking, Historical Journeys, Sequence markers, Vacation activites, Important life events			
4	• Describe habits and routines • Describe future plans	Question formation, Present simple, Much, many, a lot of, Review of possessive Adjectives, Gerunds, Present progressive, Subject + *be* / *have*+ description, Object pronouns, Present simple and frequency adverbs. Countries, Occupations, Travel, Nationalities, Describing places, Physical /Personality description, Sport and leisure			

5	• Describe job or study experiences • Compare people, objects, and places	Prepositions, Gerunds and infinitives, Past progressive, Going to / will /won't, Review of past simple, Comparatives, Superlatives, Can/could in requests, Too/either. Directions, Sports and leisure, Parties, Business, Life histories, Crime history
6	• Narrate a past personal experience • Report on familiar topics that have been read about or watched	Possessive pronouns, Should/need to/have to, Past tense, Present perfect with *Ever/ never*, Verb forms, Do/make, Present perfect with *For/since* and *how long?*, *Tell/ask/ want*, Review of tenses, Present progressive, going to/might, Will/won't. Health, Life events, Fashion, Weekend breaks and plans

APPENDIX 3

Writing Assessment Rubric – Courses 8, 9, 10, 11, 12, 13 (ADULT)

Aspects/Score	5	4	3	2	1
Coherence & Cohesion	**a.** Develops a unified and coherent text that relates **all** ideas together **b.** Uses **meaningful** transitions and connectors **c. Progression** is smooth and controlled	**a.** Develops a unified and coherent text that relates **most** ideas together **b.** For the **most part** transitions are meaningful and links between ideas are logical **c.** The progression of ideas is **rarely** hindered by factors such as wordiness, repetition, fragment, run-on sentences, or unfocused ideas	**a. Some** overall organization but **some** ideas are **illogical** or **unrelated.** **b.** The misuse, overuse, or omission of cohesive devices sometimes **affect** the logical development of the writing. **c.** The progression of ideas is **sometimes** hindered by factors such as wordiness, repetition , fragment, run-on sentences, or unfocused ideas	**a. Most** of the text is illogical and lacks organization **b.** The text lacks meaningful transitions to establish a **clear link** between ideas **c.** The progression of ideas is **hindered** by factors such as wordiness, repetition , fragment, run-on sentences, or unfocused ideas	**a.** The writing is **not** coherently arranged. The reader can not follow the thread of ideas **b. Inadequate** or **no** use of cohesive devices. **c.** Lacks progression

Grammar & Vocabulary d. only in courses 11- 13	a. Exact and outstanding control of language and sentence variety (full range). b. Uses a **rich** variety of vocabulary and some idiomatic expressions. Consistently **correct** and **appropriate** use of vocabulary d. Paraphrases or uses circumlocution successfully	a. Sentence structures are **generally** accurate and varied b. Vocabulary is **mostly** varied and used accurately and appropriately d. Paraphrasing or circumlocution is **not always** successful	a. **Inconsistent** control of language and sentence variety b. **Good** control of vocabulary but **some** confusion and incorrect word choice does occur. The writer restricts the range of structures and vocabulary to control his/her production d. Paraphrases or uses circumlocution **with difficulty**	a. **Limited** control of sentence structure and variety. b. **Lacks** variety of vocabulary. May **frequently** use vocabulary inappropriately or inaccurately d. **Rarely** paraphrases or uses circumlocution	a. **Lack** of control of sentence structures b. **Inappropriate** choice of word. d. **Never** paraphrases or uses circumlocution
Spelling & Punctuation	a. **Spelling** and punctuation mistakes are **rare.** Consistent command of spelling, punctuation, capitalization,	a. **Few** errors of spelling and punctuation. Demonstrates a very good command of spelling, punctuation, capitalization,	a. **Some** errors of spelling and punctuation.	a. **Many** errors in spelling and punctuation cause the writing to be difficult to read.	a. **Severe** errors in spelling and punctuation affect understanding of the text
Task Completion	a. **All** the elements required by the prompt are present. b. There are **no** irrelevancies or digressions. c. **Exceptional** elaboration and details to illustrate ideas.	a. **Some** parts of the prompt are more developed than others. b. There are **minimal** irrelevancies or digressions. c. Good elaboration and details to illustrate ideas.	a. **Some** of the requirements specified in the prompt are **missing.** b. There are **some** digressions and irrelevancies. c. Development remains **superficial** and **inconsistent.**	a. **Few** of the requirements of the prompt are present. b. There are **frequent** digressions and irrelevancies. c. **Little** elaboration and details in the development of the task.	a. The writing is **not** related to the given task.

* See assessment requirements in the back

For the Eafit Language Center Use ONLY - 2006

ASSESSMENT SPECIFICATIONS – Adult Program

Course	WRITING STANDARDS	Grammar & Vocabulary	Genre	Discourse Mode	Output Requirement
8	• Give an opinion on any given topic • Describe an experience or event in some detail	Past progressive vs. past simple, Used to, Review of might/will/going to for plans & arrangements, Review of comparatives and superlatives, Will/won't and going to for predictions, The first conditional, Review of present perfect vs. past simple, Too and enough; too much/too many. Life events, Leisure activities, Cognates, The body and its functions, Loan words, Country and city life, Problems of urbanization	Letters Personal E-mails Reports Short stories Simple essays	▪ Informative ▪ Descriptive ▪ Narrative	▪ **Two to three paragraphs** Continuous, intelligible text in which elements are connected using cohesive devices like: 'to summarize' 'therefore,' 'as a result,' 'since,' 'on the other hand,' 'besides,' 'in addition,' 'however,' 'such as,' 'meanwhile,' 'in other words,' 'with regard to,' 'concerning…' Use of reasonably accurately a repertoire of grammatical structures Sufficient vocabulary to express ideas on familiar topics
9	• Summarize factual information on familiar topics • Explain reasons and purposes for actions	Verb complementation, Reasons and purposes: because/ so/ so that/ to, Indefinite pronouns, Relative clauses, Present passive, Expressions followed by verb + -ing or to + verb, Causatives: have/get something done, Must, may/might, can't. Collocations, Money, Illness, Shopping, Movie vocabulary, Television, Festivals, Foods & ingredients, Symptoms, Remedies			

	Descriptors	Language content	Text types	Assessment criteria
10	• Describe hopes, dreams, and ambitions • Describe feelings and reactions on familiar topics • Explain problems with reasonable precision	Wishes about the present, hopes for the future, The second conditional, Reporting statements and opinions, Requesting and stating information Past passive, Was/were going to, Present perfect vs. present perfectProgressive. Wish/hope/expect/wait, Virtual reality, Travel, Industry, Word formation, Technology, Synonyms, Advertising		
11	• Summarize the plot and sequence of events in a book, film, or play • State conditions and results of real and unreal situations	Review of past tenses and present perfect, Conditionals: zero, first and second, Review of passives, Review of relative clauses, Gerunds, Question tags. Personal information, Life events, Leisure activities, Cognates, The body and its functions, Loan words, Country and city life, Problems of urbanization	Reports Notes Notices Diary entries Guided articles Guided essays	■ Informative ■ Descriptive ■ Narrative ■ Argumentative ■ Three to four paragraphs ■ A variety of linking words to mark clearly the relationships between ideas, using: 'For,' 'so that,' 'although,' 'in order to,' 'in fact,' 'moreover,' 'nevertheless,' 'in reality,' 'otherwise,' 'to sum up,' 'curiously,' 'coincidentally,' '(un)fortunately,' 'ironically,' 'in theory,' 'obviously,' 'presumably.' Good grammatical control without much sign of having to restrict what he/she wants to write, adapting a level of formality appropriate to the circumstances. Use circumlocution and paraphrase to cover gaps in vocabulary and structure

Ana Muñoz *et al*

12	• Write a review of a film, book or play • Explain the advantages and disadvantages of different situation from one's point of view	Past perfect, Review of past tenses, Past reported speech: statements and questions, Adjective order, Quantifiers, Obligation/necessity, Prepositions of time. Collocations, Money, Shopping, Movie vocabulary, Television, Festivals, Foods & ingredients, Culture, Illness, Symptoms, Remedies
13	• Develop an argument in support of or against a particular point of view • Provide detailed solutions to a given problem	Third conditional, Wish, Past modals: would, should, could, Past modals for deduction: could, might, may, must have, Future progressive vs. future simple, Connectors, Future perfect, Review: future forms Wish/hope/expect/wait, Virtual reality, Travel, Industry, Word formation, Technology, Synonyms, Advertising

APPENDIX 4

Course: II ADULT

Evaluator's Name_____

Scores: C&C:_____, G&V:_____, OR:_____TC:_____, Total:_____

Prompt

You are on vacation. Write an E-mail to a friend describing the place you are visiting. Mention the weather, people, and tourist places. Also include where you are staying and what activities you are doing. Write a minimum of 50 words. (10 points)

Student answer

Hello, I am in orlando, it is very good, the beachs are very nice, the sea is blue and very big, the weather is very good, is hot, in the morning I go to the beach and in the nigth I go to the parties, I dance every nigth, the people is very nice. This is very good.

Bye.