COMPUTER ASSISTED INSTRUCTION & THE ACTFL PROFICIENCY GUIDELINES

Christina Huhn

Indiana University of Pennsylvania

Abstract

The current study presents the results of one mixed-method analysis of the implementation of communicative online activities into beginning Spanish classes, their effects on class performance, and evaluation of the student-generated work from the online module. Results show that although statistically there may be some doubt as to the effectiveness of an addition of a WebCT/Blackboard ® component into a beginning Spanish program, the environment did provide opportunities for students to demonstrate language proficiency. It is the analysis of student writing samples gathered from the hybrid course, using a rubric based on the ACTFL Writing Proficiency Guidelines, that adds depth and breadth to the study. The article also addresses common issues that arise in traditional analyses of technology implementations, and suggests ways that the ACTFL Writing Proficiency Guidelines can be useful in the evaluation of student work in qualitative research.

INTRODUCTION

Research in foreign language education and technology has long sought new and increasingly effective methods of assessing student performance. This article presents the results of one mixed-method study that suggests that traditional comparison based analyses alone may be more limiting than qualitative, proficiency-based evaluations. The study addresses common issues that arise in traditional analyses of technology implementations, and suggests ways that the ACTFL *Writing*

Proficiency Guidelines can be useful in the evaluation of student work in qualitative research.

The project that served as the basis for this article involved the implementation of communicative activities in a WebCT/Blackboard based module into beginning Spanish classes, its effect on the class performance, and the results of student-generated work from the online module.

LITERATURE REVIEW

CAI and CALL programs have also traditionally evaluated performance in a prescriptive fashion, by comparing programs "with" and "without" technology, or comparing students to other students , rather than comparing performance to pre-established benchmarks. Research in foreign language education and technology has included numerous studies evaluating the effectiveness of technology-based programs, however, in recent years, researchers have called for expanded evaluation of these studies. Traditionally, research in technological programs such as these has included three main areas of focus: student outcomes (grades and test scores), student attitudes about learning through distance education, and overall student satisfaction toward distance learning. (Pederson, 1987). Studies evaluating student performance have been less frequent. Sanders (2005) conducted an analysis of a re-designed, hybrid Spanish course (reduced face-to-face contact, increased online work), He measured student performance using the *Brigham Young University Web-based Computerized Adaptive Placement Exam (BYU WebCAPE),* the ACTFL *OPI (Oral Proficiency Interview)* and the *WPT (Writing Proficiency Test)*. He found that the course redesign was most successful in increasing enrollment opportunities and cost reduction – problems faced by that particular university. He also found that proficiency and student outcomes worsened over time, but he also states that students were still able to reach an "Intermediate" level of proficiency – a finding that should not be quickly discarded.

Chapelle (1997) and Burston (2003) both note that investigators are still trying to determine the optimal approach to analyzing research in CALL. Burston (2003) comments that while CALL and IT (Instructional Technology) continue to be called to appropriately demonstrate their effectiveness, research also needs to show what has been accomplished.

In the same vein, this expansion helps surmount many of the challenges inherent to traditional analyses. Burston (2003) also examines the problem of "no significant difference". In his article, he discusses the misinterpretation of "no significant difference" as "no difference exists" rather than "the significant difference cannot be readily identified". In quantitative evaluations of CAI and CALL programs, the validity of such studies has at times remained questionable, due to the external and uncontrolled variables abundant in many studies. Jamieson (1988), Chapelle (1997)

and Burston (2003), concur, and explain that many studies are still looking at cause and effect. Failure to design studies that offer a wider view of student learning could thus slow our efforts to measure the effects of improved teaching methods. Burston (2003) advocates that we need to broaden our views of our research beyond the "computer versus traditional" methodology. One way to accomplish all this is by enriching our assessment measures to include both quantitative and proficiency-based analyses.

General Trends in L2 Assessment

Several authors have advocated that we re-evaluate our views of assessment, including analyses of technological implementations. Shrum and Gilson (2005), and Phillips (2006) both examine this re-direction. Phillips (2006) presents the following trends in assessment practice:

- Moving away from analyzing only specific knowledge and isolated skills toward assessing full knowledge
- Spending less effort comparing students to other students and more effort comparing performance to pre-established benchmarks.
- Treating assessment not as <u>independent</u> of curriculum and instruction, but aligning assessment <u>with</u> curriculum and instruction.
- Basing inferences not on restricted or single sources of information but rather on multiple sources of evidence
- Shifting from assessment as conclusive to regarding assessment as continual.

With *broader assessment,* the responses are contextualized and individualized. Foreign languages often have more than one possible way to express an idea. The focus shifts to what the student *can* do, and where further learning is necessary. The addition of qualitative, descriptive results allows comparison of achievement to expectations. These assessments can occur at any point during learning – and their value as a feedback tool may even outweigh their value as a final measure of learning. (Phillips, 2006)

Traditional Analyses vs Broader Assessment

Written or computerized exams and traditional analyses have become a customary method of assessing student knowledge, and research shows a variety of investigations of their effectiveness (Salaberry, 1999; Burston, 2003; Chapelle, 2004 and Phillips, 2006, among others). While appropriately used traditional analyses still hold a place in assessing language learners, they provide only one dimension of our students' knowledge. Delett, et al (2001) indicates, "The recent emphasis in foreign

language education on student performance has resulted in a reevaluation of instruction and assessment approaches. In the area of assessment, teachers have turned to techniques that underscore student participation and progress [sic]" (p. 559). Active assessment measures such as portfolios, written lab reports, and communicative online activities are of equal or greater value than the traditional paper and pencil or computerized testing methods. By extension, the same holds true for our research agenda: we need to incorporate methods beyond quantitative measures.

Additionally, The American Council of Teachers of Foreign Languages (*ACTFL*) has developed Proficiency Guidelines that identify specific levels of language proficiency, and we are required to show that students have reached those levels– not that group A was better than group B, or method C was better than method D. For example, reviews of teacher education programs for recognition by NCATE (National Council for Accreditation of Teacher Education) no longer rely on numerical data such as exam scores or grades, but rather on evidence that students (our future teachers) can demonstrate proficiency. Given these trends, traditional analyses will provide only part of the picture –more detail and strength of assessment is required to determine what our students know, what they need to learn, and where they are in the process.

ACTFL Writing Proficiency Guidelines

The ACTFL Writing Proficiency Guidelines (hereafter Writing Guidelines) were intended as "global characterizations of integrated performance" (Breiner-Sanders, et al, 2002, p. 9). In other words, these guidelines are intended to be used to identify what they *can* do, not what they *cannot* do. Despite the efforts put into these guidelines, few studies have utilized these guidelines in empirical research. In the preface to the *Writing Guidelines:*

> *The committee invites the profession to use these guidelines to assess writing proficiency and to consider the implications of these revisions on instruction and curricular design. The committee also invites the profession to continue to study, discuss, and carry out research on these writing guidelines so that they can be further refined to more precisely describe writing performance.* (ACTFL, 2001)

The ACTFL *Oral Proficiency guidelines* and the *Oral Proficiency Interview* have benefitted from criticisms and empirical studies on discourse analysis, and discussion of reliability and validity issues, among others. This kind of attention in our research agenda would be expected, yet the *Writing Guidelines* have not experienced the same kind of consideration. Thompson (1996) attempted to find correlations between proficiency and experience levels, and found that while writing

improved significantly, there is much overlap and variation between language students. Hayden-Roy (2004) used the proficiency guidelines as a foundation for discussing text choices and learning goals. Henry (1996) used the *Writing Guidelines* to analyze student writing, but states, "Although many have accepted the *Writing Guidelines* fairly uncritically, their validity and reliability remain untested" (p.321). The most compelling literature on the *Writing Guidelines* was completed by Valdes,et al. (1992). Valdes gave detailed attention to the history and elements of the Writing Guidelines, and illuminates many assumptions made in their development. She then details a study conducted using Spanish student writing samples in Spanish, using the *Writing Guidelines* in an attempt to further the development of a theory of L2 writing, and raises important issues related to these guidelines that must be further addressed in the research. She concludes her article by saying:

> *In spite of whatever limitations they may have, the Guidelines have caused us to examine progressions and sequences of development that had not seemed relevant before… it is our hope that FL professionals will join their ESL colleagues in carrying out the kind of research that can inform both the teaching and the assessment of writing in languages other than the first.* (Valdes, p. 348).

The current study utilizes the *Writing Guidelines* as part of a mixed method study, in order to demonstrate one way these guidelines may help us better evaluate language proficiency in a Computer Assisted Instruction module.

DESCRIPTION OF THE STUDY

This study was designed as an assessment of the Beginning Spanish program at a large, land grant Midwestern university. Prior to the study, the Spanish 102 course was a traditional classroom-based course. In the fall semester, a new technology component was added, using WebCT/Blackboard for delivery.

In addition to online lab work, students completed independent tasks focusing on culture and writing. These tasks coincided with the currently accepted practice in foreign language education of Communicative Language Teaching (CLT), which emphasizes real-life situations, active learning, communication, functional language, and interaction with the language, rather than memorization, repetition and grammar-driven teaching methods. CLT attempts to replicate the way we learn our first language by exposing us to written and spoken language as early as possible in the learning process. The intention of the additional component was to allow students an opportunity to demonstrate active language skills, with the hopes that this additional experience would result in stronger knowledge that could be seen throughout the course. A sample writing activity is included in Appendix B.

The study was conducted over the course of two semesters of language instruction in Spanish 102: one spring semester, and the following fall semester. The population included all students enrolled in Spanish 102 during those semesters. (Appendix A contains full demographic information). The population was chosen in part because the researcher had taught Spanish 101 during the pilot testing of a similar technology component, affording a strong comprehension of the technology component, and a unique viewpoint into the population and environment of the study.

Data collection and analysis included grade books (which contained lab scores, final exam and final course grades) from all sections for both semesters, mid-term and final exam written paragraphs for both semesters, and all student generated WebCT/Blackboard responses to the writing activities for the fall semester. All data was gathered during the natural course of instruction, and maintained by the teaching assistants who taught each of the 30 sections. The mixed-method study began with a traditional statistical analysis, followed by a qualitative analysis. The study was designed to look at the Beginning Spanish program as a whole, but this article will look specifically at the following research questions:

1. What effect did the WebCT/Blackboard supported classroom have on overall student performance?
2. How did it affect student individual performance on Midterm and Final exams?
3. Was there a significant difference in the scores on the lab work?
4. Was there a significant difference in the scores on the midterm and final exams?
5. Was there a significant difference in the final course grades after adding the WebCT/Blackboard component?
6. Were students able to demonstrate language proficiency through their work in the WebCT/Blackboard environment?

METHODOLOGY

As is often seen in empirical research in CAI and CALL, the study was conducted to compare a traditional course with instruction supported by technology, in order to determine if the active and independent learning activities in the online environment helped students develop language skills. The quantitative data included scores from midterm (covering the first half of the course, and the first four chapters of the course) and final exams (covering the entire course, but not comprehensively), the Lab Component scores and overall course grades.

In the spring course, Spanish 102 did not include a web-based enhancement. Instruction was primarily classroom-based, using the communicative language

teaching method. Students were scheduled for one class day every two weeks in a computer laboratory, and the Laboratory component was completed individually, without textbook, dictionary, collaboration, or any other assistance. In fall course, the same types of data were gathered for Spanish 102, but the lab component was now assigned on-line (as LabWeb), and students now made the choice of how to complete LabWeb. The content of the lab component was equivalent to that of the spring semester. The exams were also equivalent – as both courses were using the same edition of the same textbook (Dímelo Tú, Harcourt, Fourth Edition, 2001), and the exams were being constructed by the same, experienced coordinator, and were only minimally modified between semesters. As a result, the exams were relatively consistent between the two semesters.

Population Demographics

The population for this study is composed of all students who enrolled in Spanish 102, whether as a requirement for their degree or for some other reason. The samples are a convenience sample composed of all enrolled students in the spring and the following fall semesters. These samples are designated as S and F, respectively.

In summary, the demographics of the two populations are relatively equal in composition. See Appendix A for full population details.

Data Collection Procedures

The data for this project were collected in a natural environment, during the normal course of instruction. There was no intervention on the part of the researcher. The data were collected in electronic format, and provided to the researcher on CD-Rom and in paper format. For the main quantitative analysis, the entire population was observed. For the final exam analysis, a random sample was taken due simply to the volume of responses.

Data Preparation

The Spanish 102 course sections were coordinated, meaning that all Teaching Assistants and instructors use the same course content, syllabus, and grading scale. An electronic grade book was created for use in tracking student grades. This grade book for both spring and fall semesters was the source of the data for this study.

Once these data were prepared, they were entered into an Excel spreadsheet, and the data were analyzed using SPSS software.

Data Description

The initial component of the WebCT/Blackboard exercises is the LAB. During the spring semester, these exercises were administered in a computer lab, with no books or notes allowed. In the fall semester, LAB became LABWeb and was assigned as independent work on the WebCT/Blackboard site (the content remained unchanged). The mean score for the spring semester was 84.84%, while for fall it dropped to 71.86%. This would provide the impression that the students did not perform as well on the Lab components when they were placed in an online, independent work environment.

Table 1: Initial Descriptive Analysis Results

	N	Mean	SD
LAB Spring Semester (Lab)	421	84.8409	19.78871
LAB Fall Semester (Web)	669	71.8685	21.33586
Midterm Exams Spring	421	76.7886	12.58572
Midterm Exams Fall	669	73.2123	13.45907
Final Exams Spring	421	74.0048	17.60262
Final Exams Fall	669	74.0105	20.74247
Final Course Grade Spring	421	80.3646	12.25555
Final Course Grade Fall	669	78.9575	14.93823

The next element of interest is the midterm exams. This exam is given at approximately the midpoint of the semester. The mean for the midterm exams for the Spring Semester was 76.79%, whereas Fall Semester the Mean was 73.21%. Although the mean is lower for the fall semester, the difference is smaller than that for the LABWeb exercise.

The third element is the final exam scores for both semesters. For this semester, the means are approximately equal: 74.00% (S) and 74.01% (F). This could be

interpreted as a non-negative change – that is, the addition of the web-enhanced course components did not have a notable effect on the students' final exam grades.

The final and perhaps most important element (at least in the eyes of our students) is the final course grade. The mean was 80.36% for the spring semester and 78.96% for the fall. This shows a slight, but still notable decrease in the overall course grades with the additional of the web-enhanced course.

However, the results above are simply descriptive in nature, and before we can accept them as statistically significant, some further analyses are needed.

Data Analysis Procedures

The first step in analyzing the quantitative data (lab, midterm exam, final exam and overall course grades) was to perform Levene's test for equality of variances. The means were then compared using independent sample t-tests.

Additionally, qualitative data (final exam writing selections) were also analyzed. For the qualitative statistical analyses, sample sizes were too small for Chi-Square Tests of homogeneity; therefore, Fisher's Exact Test had to be used.

Levene's Test of Variance

We can further compare these variances by using Levene's Test of Variance. Levene's test was developed in 1960, by H. Levene and expanded by Brown and Forsyth (1974). This test compares the variability of the populations, and gives the probability that the two populations were primarily homogeneous in terms of their variances. In order to conclude that these variances are equal, Levene's test should have a P-value greater than a significance level of .05 (p>.05). These hypotheses can be expressed as follows:

Table 2: Hypotheses for Levene's Test

Ho: $\sigma(Spring) = \sigma(Fall)$	The variances of the populations are equal
Ha: $\sigma(Spring) \neq \sigma(Fall)$	The variances of the populations are not equal

In looking at the data for this project, the following results were observed:

Table 3: Results of Levene's Test

Levene's Test for Equality of Variances			
	F	P-value	
LAB Lab Component	56.962	.000	p < .05
Midterm Exam Grade	.000	.994	p > .05
Final Exam Grade	4.386	.036	p < .05
Final Course Grade	5.883	.015	p < .05

*Note: When Lab components were broken down into those prior to the midterm and those after the midterm, the P-value remained the same.

In Levene's test, if the P-value is less than .05, the two variances are significantly different. If it is greater than .05, the two variances are not significantly different. In terms of the LAB Lab Component, the final exam Grade, and the final course grade, the p-value was less than .05 (p < .05) and therefore shows that the two variances are significantly different. In these components, we must therefore reject the null hypothesis that these populations do have equal variances. We therefore accept the alternate hypothesis that these variances are unequal. We will therefore treat this data as "Equal Variances Not assumed" in calculating our t-tests. The P-value of .994 is very large. From this, we can conclude that we should NOT reject the null hypothesis and the variances may indeed be equal. This is a bit of a caveat, and hints at the fact that there might be a type II error. Further analysis using T-tests will help to clarify this.

Our final condition, that the samples are independent of each other can be demonstrated by simply looking at our populations. While it is hoped that the dependent variable (fall Semester – when the "treatment" of including the web-enhanced course components was added) will show a change, the two samples were not matched, as they are two separate groups of students. This leads us to use Independent Sample T-tests.

Independent Sample T-Tests

Now that the conditions for using Independent Sample t-tests have been met, we need to state our hypotheses. These can be expressed in terms of a null and alternate hypothesis (S symbolizes the spring semester, F symbolizes the fall semester):

Table 4: Hypotheses for Statistical Analyses

Ho: Mean(S) = Mean (F)	The means for the two semesters are equal, and the treatment (the addition of the web-enhanced component) appears have had no significant effect
Ha: Mean (S) ≠ Mean (F)	The means for the two semesters are not equal, and the treatment (the addition of the web-enhanced component) has had some effect.

QUANTITATIVE RESULTS

A number of important quantitative results were collected through the course of the study. The table below shows the results of the Independent Sample t-tests.

Table 5: Overall results of Independent sample t-tests

	t	P-Value
LAB Lab Component	9.051	.000
Midterm Exams	4.446	.000
Final Exams	-.005	.996
Final Course Grade	1.694	.091

LAB Component

As the above table shows, the P-values for the LAB component are very small (close to 0), which provides us with very strong evidence that we should reject the null hypothesis. In other words, the means for the two semesters are not equal, and the treatment (the addition of the web-enhanced component) appears to have had *some effect* on the performance of our students.

A look at the means of the LAB Exercises in Table 1 for the two semesters shows that the mean dropped sharply in the fall semester – from 84.8409% to 71.8685%. This tells us that the effect on the LAB Web Exercises appears to be a negative one.

Midterm Exam

The midterm Exam grades show a p-value of .000 (p <.05) which provides us with very strong evidence that we should reject the null hypothesis. In other words, the means for the two semesters are not equal, and the treatment (the addition of the web-enhanced component) appears to have had *some effect* on the performance of our students. A comparison of the means for the midterm exams shows that the means for the Spring Semester was 76.79%, whereas Fall Semester the Mean was 73.21%. This shows that there was a negative effect on the midterm grades in the fall semester.

Final Exam Scores

The final exam, however, shows something quite different. The p value for these scores is quite large (p-value = .996 – notably greater than .05).This provides strong evidence that we should indeed not reject the null hypothesis, and conclude that the addition of an online course management tool had no demonstrable effect on the final exam scores.

Final Course Grade

The final course grade shows a p value of .091. This indicates that there is not strong enough evidence to reject the null hypothesis and we therefore conclude that the addition of the online course management tool has had no significant effect on the final course grade.

Quantitative Analysis

The table below shows a summary of the results of the quantitative analysis

Table 6: Summary of Results

	Effect
LAB Component Overall.	Significant effect – negative
Midterm Exams	Significant effect – negative
Final Exams	No significant effect
Final Course Grade	No significant effect

The results of the statistical analyses were either negative or "no significant effect". This may lead us to conclude that the changes made to the program were ineffective – specifically that the addition of the WebCT/Blackboard component did not increase student language knowledge. However, there is some question as to the validity of these results. As Burston (2003) cautions, results of this nature are often misinterpreted, causing researchers to feel that the program has not been successful in achieving the desired goals, when the more accurate explanation may have been that the this method of investigation simply did not show the change.

These statistical analyses would seem to indicate that the addition of the new component to a beginning Spanish course is not notably effective, and this apparently negative result did leave those involved in this project pondering the reasons behind that result. As previously mentioned, the researcher in this case had a unique perspective and insight, having taught a previous level course using a similar WebCT/Blackboard component.

The researchers went back to the data, only to discover, as Burton (2003) suggests, that the validity of the original study had been jeopardized by a number of factors unknown to and uncontrollable by the researcher. A detailed discussion of the problems is beyond the scope of this article, but included issues of student non-compliance, record keeping, TA training, and supervision and support for large language programs. These concerns were disheartening at best – destructive to the project at worst.

These results also seemed inconsistent with typical performance for beginning language learners. In teaching the earlier course, a pattern of responses appeared that made these analyses even more puzzling. This pattern – in which students demonstrated significant language abilities when completing interactive tasks in the new module, drove the researcher to take a deeper look at the remaining part of the project.

QUALITATIVE RESULTS

The disappointing initial results underscore the previously noted importance of not limiting ourselves only to one type of analysis, but rather look at broader measures of student language proficiency. In the fall semester, in addition to moving the lab component to an online environment, additional activities were incorporated. One assignment was a communicative writing task, which provided students an opportunity to express themselves. Student responses consisted of two parts: a list of short answer responses, and then a paragraph composed using those answers. The assignments followed the sequence of course content, and increased in difficulty as the semester progressed. A sample exercise can be found in Appendix B.

It is important to note here that although students were divided into groups, some students chose to work individually, and some chose to collaborate. The writing activities were independent tasks, completed outside of the language classroom. Students therefore chose the method of completion. WebCT/Blackboard recorded all data generated by the students.

Data Analysis of Writing Exercises

The writing exercises provided an opportunity for students to use their language skills to express themselves. The prompts were based on the textbook section entitled ¿Te Gusta Escribir? ("*Do you like to write*?"), and were appropriate novice/intermediate level writing tasks. These assignments focused on planning, organizing, and writing. There were 1203 responses received from 19 instructors, and all identifying data were removed to prevent identification of students and instructors.

Rather than analyzing this student-generated work statistically, using scores or grades, this work can be more appropriately analyzed by selecting a *typical case sample* (Patton, 1990). A typical case sample is a qualitative sample based on one or more "typical" cases. The researcher brought a unique perspective to this project as a teacher of beginning language students. This allowed selection of a typical case sample that demonstrates student proficiency. The purpose of such a sample is intended to be illustrative – in this case, it is intended to exemplify the language capabilities of the subjects of the study, rather than "prove" proficiency.

To obtain the sample, a typical set of responses from one instructor's artifacts was chosen, based on the researcher's unique perspective in this program. The *Writing Guidelines* were used as the foundation for analysis, based on the patterns observed in the previous semester. The *Writing Guidelines* do not equate to a particular level for a Spanish 102 course. However, a closer look, in light of the limited studies discussed above, and the first year course objectives outlined by this

specific university (See Appendix C) would lead us to expect students at this level to be at progressing from the Novice to the Intermediate level of language proficiency.

The writing prompts used in the activities were novice to intermediate level communicative tasks, but related to the individual chapter topics. The evaluation rubric (Table 7) utilized in this study was developed based on the Novice and Intermediate proficiency descriptors in the *Writing Guidelines*. Student generated responses were then compared to the rubric for evaluation (See Table 7).

In comparing the student responses (Table 8), to the proficiency based rubric (Table 7), it can be seen that students are approaching the proficiency level that might be expected of a student in a Spanish 102 at this university.

Table 7: Rubric based on selected criteria from the ACTFL Writing Guidelines

	Approaching	Novice/Intermediate	Exceeds
Topics	Can write lists, short messages, within original context only	Can create statements and formulate questions based on familiar material	Writes short, simple communications, compositions, descriptions, and requests for information based on familiar vocabulary
Grammatical Structure	Recombine learned vocabulary and structures into simple sentences. May see verbatim copying from original materials.	• Conversational style, basic word order; • Mostly in present time. Many sentences are basic subject-verb-object word order. • May be basic errors in grammar, word choice, punctuation, and spelling.	• Mostly present tense – other times are inconsistent. • Begin to demonstrate control of syntax in basic sentences, and basic verb control • Hints at stylistic elements
Vocabulary	Common, discrete, familiar vocabulary only.	Limited to highly predictable content and personal information tied to limited language experience.	Personal preferences, daily routines, common events, and other topics related to personal experiences and immediate surroundings.
Structure and Coherence	• Unable to sustain writing.-errors are prevalent. • Generally comprehensible to natives used to the writing of non-natives, but gaps in comprehension may occur.	• When attempting to write at a higher level, quality deteriorates. • Writing is understood by natives used to the writing of non-natives, although additional effort may be required.	• When attempting to write at a higher level, quality deteriorates. • Can be easily understood by natives used to the writing of non-natives.

Table 8: Selected student writing responses

Sample #1	*Approaching*
Grade: 7	Comments: a. …un cantante. b. Nació… c. … y cantante. I want more information.
Answer: a. Ricky Martin es el cantor. b. Nacié en Puerto Rico. c. Ricky Martin se hizo actor y cantor.	
Sample #2	*Approaching*
Grade: 7	Comments: You needed to write a parragraph, not just 3 very short answers.
Answer: a. Violeta Barrios de Chamorro b. 1930 c. la silla presidencial	
Sample #3	*Novice/Intermediate*
Grade: 10	Comments: b. ¿Qué servían …? c. ¿A qué hora era….? La fiesta era en …. Sirvieron Macs grandes y malteada de chocolate en la fiesta. La fiesta fue el …
Answer: a. ¿Donde estaba la fiesta por la tarde? b. ¿Que sirviá a la fiesta? c. ¿ Que tiempo era la fiesta? La fiest estaba en McDonalds. Habia Grande Macs y el chocolate malt sirvió en la fiesta. La fiesta era sábado en el mediodía	
Sample #4	*Approaching*
Grade: 8	Comments: Bien!
Answer: El español era probablemente orgulloso. También, feliz de tener la tierra nueva. El local son probablemente todavía triste. Su faught de la gente duramente. El emperador estaba triste perder. Era su tierra. Muchos muerto.	
Sample #5	*Exceeds*
Grade: 10	Comments: Muy bien!!!
Answer: Fumar a. El aspecto positivo: 1. Bueno para la imagen: Es algo más adulto. Muestra independencia. 2. Comformidad con el grupo: Todos los amigos fuman. b. El aspecto negativo: 1. Malo para la salud: Causa cánc de pulmones. Puedes morir. Aumenta el riesgo de tener un ataque de corazón 2. Molesta a muchas personas: Nc se permite en muchos lugares. Afecta dónde puedes sentarte. c. Obviamente, el aspecto negativo del fumar más daña que el aspecto positivo El fumar mata los centenares de miles de gente cada año. Todos los dias, miles de jóvenes empiezan a fumar porque ellos piensan es bueno para su imagen, los ayuda a probar que ellos son un adulto. A veces, los jóvenes empiezan a fumar porque todos sus amigos fuman, asi que ellos deben también. Antes usted lo empieza a fumar debe saber que ese fumar puede causar el cancer de pulmón, aumenta su oportunidad de tener un ataque de corazón. También marca muchas gente molesta cuando usted fuma alrededor ellos. Actualmente, muchos restaurantes, edificios federales han prohibido el fumar.	
Sample #6	*Exceeds*
Grade: 10	Comments: Muy bien!!
Answer: La historia de Cuba De 1511-1898: la Colonia de España En 1898 : Fue cedidos a EE.UU. como resultado de la duerra de 1898. La primera mitad del siglo XX: un período de mucha inestabilidad polítaca y socia para Cuba. Durante la segunda mitad del siglo XX : bajo el poder del dictator military Fulgencio Batista. 31 de diciembre de 1958: Joven abogado Fidel Castro tomó control del gobierno. En 1960: Fidel Castro proclamó a Cuba como una república socialista. Desde que 1960, 10 por ciento de la población han dejado Cub para estados unidos En el tarde diecinueve siglo, Cuba era uno de las últimas colonias del imperio magnífico de España. En 1898, Cuba y el Puerto Ricos fueron rendidos a estados unidos como el resultado de la guerra de 1898. Durante el primer tiempo del vigésimo siglo, Cuba luchó diplomáticamente y socialmente bajo el poder d dictador militar Fulgencio Batista, que favoreció los intereses extranjeros, especialmente los estados unidos. El 31 de diciembre de 1958, Fidel Castrol tomaron el control del gobierno. Dos años posteriores, él declaró Cuba como una República de socialista. Como resultado, los estados unidos han impuesto una prohibición en Cuba. También, casi 10 por ciento de la población han dejado Cuba para los estados unidos.	

Qualitative Analysis

In comparing these writing samples using the proficiency-based rubric students are indeed able to demonstrate a novice/intermediate level of language writing proficiency in these activities. One paragraph is a novice/intermediate level of writing, three other responses are approaching that level, and two have already exceeded it.

The paragraphs reveal some inconsistencies, and in some cases, characteristics of Novice learners are still clearly visible, such as verbatim copying from the source materials and errors. Students at this level were able to complete the tasks with some language comprehension, and able to demonstrate an appropriate level of language proficiency. Errors were prevalent, but that is an expectation of beginning language learners.

DISCUSSION OF FINDINGS

In evaluating the findings of this study, it is beneficial to revisit our original research questions:

1. What effect did the WebCT/Blackboard supported classroom have on overall student performance?
2. How did it affect student individual performance on Midterm and Final exams?
3. Was there a significant difference in the scores on the lab work?
4. Was there a significant difference in the scores on the midterm and final exams?
5. Was there a significant difference in the final course grades after adding the WebCT/Blackboard component?

The response to the first group of questions was discussed in detail, and the results can be seen in Table 6. The initial quantitative analysis showed that the technology-enhanced program did not reveal any notable effects on the lab work, midterm and final exam scores, or final course grades, giving the impression that the addition of the new online module was ineffective. However, the final question holds more promise:

6. Were students able to demonstrate language proficiency through their work in the WebCT/Blackboard environment?

By using proficiency-based rubrics to analyze a small subset of the qualitative data, student knowledge is more visible. Students were able to demonstrate language proficiency through the writing activities. The results of the study – and the benefits of these multiple analyses can be seen more clearly when comparing the traditional analyses of the implementation and proficiency based assessments (Table 9).

Table 9: Comparison of Traditional versus Expanded Assessment

	Traditional, comparison based assessment	Proficiency-based evaluation
Details	Negative effect or no significant effect on student performance	Students were able to complete the assignments, comprehend and respond to materials in the target language and demonstrate language proficiency.
Knowledge of student proficiency	Unable to identify specific student knowledge.	Student knowledge is visible and identifiable in the results obtained.
Conclusion	The WebCT/Blackboard ® component did not enhance the Spanish 102 courses, or the data and detail gathered is insufficient.	By providing students with an opportunity to express themselves in an open-ended, technology-supported environment, students are able to demonstrate language proficiency at an appropriate level.

In reviewing these two sets of results side by side, we can see the strength and depth gained by the proficiency based assessment. Contrary to the traditional analyses, the broader language proficiency analysis shows that students were able to demonstrate an appropriate level of language proficiency. This comparison tells us that while we are unable to identify whether or not the added technology component increased student knowledge as a whole, we can conclude that the students enrolled in the WebCT/Blackboard enhanced course do demonstrate some level of proficiency – that of novice/intermediate level language skills – as expected of a student of this level. From this we can also conclude that the online module did not hinder students language proficiency development. Further investigation may reveal exactly how it provided them the means to demonstrate that proficiency.

In the research conducted for this article, the traditional, quantitative data analysis left many questions unanswered. A deeper, qualitative analysis of a smaller portion of the population provided depth and strength of assessment that we as educators are being called to provide.

Phillips (2006) suggestions go a long way in helping us develop much deeper and more meaningful assessment:

- The proficiency-based analyses used in this current study demonstrate an assessment of the *full knowledge* of our students, by looking more closely at the language they generated rather than just their scores and grades.
- By aligning our assessment measures with the established benchmarks for foreign language learners, *comparing our student-generated work to the Writing Guidelines rather than to other students*, we can gain a clearer picture of our students' abilities.
- Currently, Communicative Language Teaching is commonly used as a teaching method in foreign language education. This is an active teaching method, and assessment of student-generated work created

> from similar activities *aligns much better with this instructional method.*

- Expanded assessments also generate *multiple sources of evidence*. In this project, we now have statistical, qualitative, and proficiency-based data. Where the statistical data are unconvincing, the qualitative, proficiency-based data allows us to see that our students are actually performing much as we would expect for beginning language learners. It is important not to lose sight of the fact that we are still working with beginning learners here, and that language is an internal and varied process – statistical results may not demonstrate what our students can accomplish.

- Finally, assessment *needs to be continual*. This is ananlysis of an initial semester of a new technology implementation. A similar analysis could also be performed on future semesters to investigate these same elements over time.

LIMITATIONS

All research has limitations, and this study is no exception. In addition to the compromises in the quantitative data, there are several areas that may merit further investigation.

Generally the populations of spring and fall language classes, while demographically equal, may have inherent differences related to scheduling, repetition of courses, and other factors, and this could have an effect on the performance of that population. Further research is necessary to investigate this issue.

Additionally, as this work was done in an online environment, there is no way to know if students worked independently or collaboratively – this may also be a useful point for further research.

Finally, qualitative, proficiency-based analysis is more time and labor intensive, and as such, requires smaller sample sizes. This can make results difficult to generalize and replicate. The sample chosen for the qualitative measure in this study was selected using typical case sampling (Patton, 1990) that illustrates our need to not limit ourselves to statistical analysis. However, samples of this type may be more difficult to replicate or apply more broadly to other context. It also validates the need for further research and investigations of this nature.

CONCLUSION & IMPLICATIONS

The topics discussed in this article, and the components of this study support the need for expansion of our research methodologies in foreign language research, in particular in our research using Computer Assisted Instruction. Our research studies must not be restricted to traditional comparisons of groups of students such as students in one semester versus students in another, but must include broader, expanded assessment measures.

Similar to Sanders (2005), the students in this study were able to complete intermediate level language tasks. The assessment in this study provides us a depth and breadth of knowledge that the traditional analysis leaves unaddressed. Knowing that our students can comprehend and respond to authentic texts, and even achieve an appropriate level of proficiency gives us a significant picture of what our students can do, and where further learning is needed – as well as providing us as researchers with an understanding of where further investigation is imperative. This knowledge will prove far more valuable in furthering research in foreign language assessment than the more restrictive analyses. When designing research studies evaluating technology-based programs it is important to remain grounded in current practices and research in pedagogy, including knowledge of the proficiency guidelines, language standards, and similar topics.

Additionally, this study demonstrates how the *Writing Guidelines* can be useful in analyzing the work completed by post-secondary students. Studies using the ACTFL Proficiency Guidelines in post-secondary second language research remain limited.

Finally, this current study also identifies a number of new paths for continued research:

- Do writing selections from student compositions or the midterm and final exams also demonstrate this proficiency?
- How do students perform on listening and reading comprehension tasks when analyzed in a similar manner?
- Can the *Writing Guidelines* be further used in our research, in order to gain additional insight into our students' abilities?
- Can these concepts be expanded into other fields – in particular those who are held accountable to external governing bodies?

With increased accountability to outside organizations such as NCATE (National Council for Accreditation of Teacher Education) and others, deeper, stronger assessment of our students has become essential. We will be required to demonstrate not that our students can pass an exam, or show improvement in a course grade, but that they can demonstrate proficiency and can demonstrate

proficiency levels developed by the appropriate content specific organization (ACTFL in the current study). This project demonstrates the necessity and benefits of shifting our focus away from traditional comparison based assessment alone, which will likely fall short in demonstrating what our students (and our future teachers) are able to do.

REFERENCES

American Council on the Teaching of Foreign languages. (2001) ACTFL Proficiency Guidelines. Hastings-on-Hudson, NY: ACTFL

Breiner-Sanders, K., Swender, E., & Terry, R. (2002, January 1). Preliminary proficiency guidelines--writing revised 2001. *Foreign Language Annals*, *35*(1), 9-15.

Burston, J. (2003). Proving IT works. *CALICO Journal,* Vol 20, No. 2 p.219-226.

Chapelle, C. (1997). CALL in the year 2000: still in search of research paradigms?. *Language Learning & Technology*, Vol 1, No. 1, pp. 19-43

Chapelle, Carol A. (2004). Technology and second language learning: expanding methods and agendas. *System*. Vol 32, no. 4, pp. 593-601.

Delettt, Jennifer S.; Barnhardt, Sarah; Kevorkian, Jennifer A.; A framework for portfolio assessment in the foreign language classroom. *Foreign Language Annals*, v34 n6 p559-68 Nov-Dec 2001

Hayden-Roy, P. (2004, Spring). Well-Structured Texts Help Second-Year German Students Learn to Narrate. *Unterrichtspraxis/Teaching German*, *37*(1), 17-25

Henry, K. (1996, Autumn). Early L2 Writing Development: A Study of Autobiographical Essays by University-Level Students of Russian. *Modern Language Journal*, *80*(3), 309-326

Jamieson, Joan & Chapelle, Carol. (1988). Using call effectively: what do we need to know about students. *System.* Vol 16, no. 2, pp. 151-162.

Lafayette, Robert C., ed. (1996). *National standards: a catalyst for reform.* ACTFL Foreign Language Education Series. Lincolnwood, IL: National Textbook.

National Standards in Foreign Language Education Project (1999). *Standards for foreign language learning in the 21st century.* Yonkers, NY: Author.

Patton, M.Q. (1990). Qualitative evaluation and research methods (2nd ed.). Newbury Park, CA: Sage.

Pederson, Kathleen Marshall. Research on CALL. In *Modern Media in Foreign Language Education: Theory and Implementation*. ACTFL Foreign Language Education Series. (pp.99-131). Lincolnwood, IL: National Textbook Company.

Phillips, June K. (1998) Media for the Message: Technology's Role in the Standards. *Calico Journal* Vol 16 no.1 pp. 25-36.

Phillips, June K. (2006). Assessment now and into the future. In *ACTFL 2005-2015: Realizing Our Vision of Languages for All*. ACTFL Foreign Language Education Series. (pp 75-103). New Jersey: Pearson Prentice Hall.

Robinson-Stuart, Gail L. (1998) New directions in CALL: getting to the heart of it. Calico Journal Vol 16 no.1 pp. 11-23.

Salaberry, M. Rafael. (1996). A theoretical foundation for the development of pedagogical tasks in computer mediated communication. *Calico Journal.* v14 n1 p5-34 Fall 1996.

Salaberry, Rafael (1999). Call in the year 2000: Still developing the research agenda. *Language Learning & Technology*, Vol 3, No. 1, (pp. 104-107)

Salaberry, M. Rafael. (2000). Pedagogical design of computer mediated communication tasks: learning objectives and technological capabilities. *The Modern Language Journal. 84, i.*

Salaberry, M. Rafael. (2001). The use of technology for second language learning and teaching: a retrospective. *Modern Language Journal. Vol. 85, I.*

Sanders, R. (2005, December 1). Redesigning Introductory Spanish: Increased Enrollment, Online Management, Cost Reduction, and Effects on Student Learning. *Foreign Language Annals*, *38*(4), 523-532.

Samaniego, Fabián A., Blommers, Thomas J., Lagunas-Solar, Magaly, Ritzi-Marouf, Viviane, Rodríguez-Nogales, Francisco. (2001) *¡Dímelo tú! 4thd Edition.* Harcourt College Publishers: Fort Worth

Smith, Wm. Flint,, editor. (1987). *Modern Media in Foreign Language Education: Theory and Implementation*. ACTFL Foreign Language Education Series. Lincolnwood, IL: National Textbook Company. (p. 99-132)

Thompson, I. (1996, Spring). Assessing Foreign Language Skills: Data from Russian. *Modern Language Journal*, *80*(1), 47-65

Valdes, G., et al. (1992). The Development of Writing Abilities in a Foreign Language: Contributions toward a General Theory of L2 Writing. Technical Report No. 61.

APPENDIX A

Detailed Population Graphics

	Spring		Fall	
Female	383	53%	229	48%
Male	336	47%	248	52%
Total	719	100%	477	100%

Year of Birth	Spring		Fall	
26 or older	18	<1%	15	<1%
25	15	3%	8	1%
24	28	6%	19	3%
23	67	14%	56	8%
22	115	24%	91	13%
21	135	28%	111	15%
20	97	20%	220	31%
19	2	<1%	191	27%
Total:	477		719	

	Spring		Fall	
FRESHMAN	163	34%	370	51%
SOPHOMORE	157	33%	185	26%
JUNIOR	118	25%	104	14%
SENIOR	38	8%	57	8%
MASTER	0	0%	1	0%
PROFESSIONAL	1	0%	2	0%
Totals	477	0%	719	100%

	Spring		Fall	
AFRICAN AMERICAN,NONHISPANIC	21	4%	53	7%
ASIAN AMERICAN/PACIFIC ISLANDER	16	3%	16	2%
CAUCASIAN AMERICAN,NONHISPANIC	418	88%	620	86%
HISPANIC AMERICAN	11	2%	24	3%
INTERNATIONAL	7	1%	4	1%
NATIVE AMERICAN/ALASKAN NATIVE	4	1%	2	0%
	477	100%	719	100%

Note: These demographics were obtained from university statistics, and vary slightly from the actual number of participants, most likely due to failure to follow official course drop procedures and similar issues.

APPENDIX B

Writing Prompt Sample

You are logged in as xxxx Logout
View *Capítulo 13, Paso 2*
Spanish Menu Question List

Type	writing
	Paso 2 **Persuadir (pp. 470-471)** **Basándote en la definición del arte de persuasión (A. Persuadir), escribe una lista de la información que vas a necesitar para escribir un ensayo a favor o en contra de un tema particular. (Puede ser los efectos de fumar, de beber bebidas alcohólicas, etc...) Organiza tu lista alrededor de los tres temas siguientes.** *[Based on the definition of the art of persuasion (A. Persuading), write a list of the information that you will need to write an essay in favor of or against a particular topic. For example, effects of smoking, drinking alcoholic beverages, etc. Organize your list around the three following themes:]*
Question	**TEMAS** **a.El aspecto positivo** *[The positive aspect.]* **b.El aspecto negativo** *[The negative aspect.]* **c.¿Cuál tiene más valor que el otro?** *[Which has more value than the other?]* **Luego escribe, en un breve párrafo (usando tu lista de temas), cuál es tu opinión personal sobre el tema. Ten points total.** *[Then, write, in a brief paragraph (using your list of topics), what is your personal opinion on the topic.]*
Links	none

APPENDIX C

Objectives for First Year Spanish Courses (through Spanish 102)

CHAP.	OBJECTIVES	STRUCTURE
9	You will learn how to describe the weather, and how it affects you. You will also learn how to describe your daily routine as well as a typical weekend. Finally, you will learn how to give detailed directions.	Weather expressions, mucho versus poco, reflexive verbs, por versus para, and affirmative "tú" commands
10	You will learn how to ask for help in case of an emergency, and how to answer questions about a crime. You will also learn how to describe a car accident and a robbery.	Adverbs derived from adjectives, preterit of irregular verbs, negative and indefinite expressions, and hacer in expressions of time.
11	You will learn how to describe what you and others used to do in the past (childhood, past experiences, etc...) You will also learn how to talk about what you have or have not done. Finally, you will discuss Ecology.	Imperfect of regular verbs, imperfect of "ser, ir, and ver," preterit versus imperfect, and present perfect.
12	You will learn how to describe what you will do in the future (during your next vacation, f.i.) You will also learn how to talk about conditional situations (What if...?) Finally, you will learn how to give advice, instructions, and orders.	Future, conditional, and the negative "tú" commands
13	You will learn how to give advice (tell someone what to do or not to do.) You will also learn how to express emotions and doubts.	Subjunctive with "¡Ojalá!" impersonal expressions, expressions of persuasion & emotion, and "Usted" & "Ustedes" commands
14	You will learn how to express fears, hopes, and opinions. You will also learn how to report what others say	Subjunctive with expressions of doubt, denial, uncertainty, and with adjective and adverb clauses

Note: These objectives were developed by the course coordinator in conjunction with the textbook. The objectives were grammar abased, rather than proficiency based, and presume that all material in the previous course had been mastered. "Gustar" is a review concept from Spanish 101.

ABOUT THE AUTHOR

Christina Huhn is an Assistant Professor of Spanish. Her doctorate is in Foreign Language Education and Technology from Purdue University and her research interests include technology, beginning learners, second language writing, and program assessments for both academic and local school districts. She also serves as an NCATE Program Reviewer.