An Empirical Investigation of the Effects of Blended Learning on Student Outcomes in a Redesigned Intensive Spanish Course

DOLLY JESUSITA YOUNG
University of Tennessee

ABSTRACT
Replacing face-to-face teaching with online instruction in higher education will become an increasingly attractive strategy for reducing instructional costs. Research on the effectiveness of online instruction, however, is sparse. The purpose of this paper is to present findings of a 2-year study that examined learner outcomes within the context of foreign language learning when contact time is replaced by online instruction.

KEYWORDS
Technology, Hybrid/Blended Learning, Technology-enhanced Language Learning (TELL), Computer-assisted Language Learning (CALL)

INTRODUCTION
Interest in the potential of computer technology as a medium for language learning has grown exponentially since the last quarter of the 20th Century. Now close to a decade into the 21st Century, advances in technology, alongside a better understanding of the learning process, has allowed for the growth of interactive uses of the computer and more sophisticated and integrated media and technology applications. Second generation web applications (e.g., blogs and wikis, and audio/video podcasts) and eLearning instructional applications (e.g., Wimba and Elluminate) are promising tools for foreign language learning. We are in the midst of a real transformation in education, something that seems to be happening overnight.

The research on the effectiveness of these advances in technology, however, has been unable to keep pace with these innovations. Only a decade ago, Conrad (1996) reviewed several major foreign-language learning journals for articles associated with computer-assisted language learning (CALL) published between 1992-1995. He found only a small percent of articles (1.4%) were empirically-based CALL studies. The 19 articles he found that were empirically based represented “almost as many different specific areas of CALL applications” (p. 160) thereby making generalizations difficult.

Liu, Moore, Graham, and Lee (2003) analyzed the ways in which computers were used to advance second language learning from 1990-2000. Their database included 246 articles from 21 refereed journals related to language learning. They found that CALL publications fell into two broad categories: research based and nonresearch based. Out of the 246 articles reviewed, only 70 were research based.

In spite of the insufficient number of empirically based studies to support the effectiveness of the use or integration of technology in learning, there is increasing pressure by
government-funded agencies (Bober, 2001), academic administrators, and the technology industry (e.g., Blackboard Inc.) to integrate technology into higher education for the purpose of reducing instructional costs (see Twigg, 1998) or to compete with the potential of a profit-seeking private sector that believes higher education needs a makeover (see Serbin Pittinsky, 2003). According to Twigg (2001, p. 23), "The highest cost component of instruction is faculty personnel." In short, in an effort to reduce costs of instruction, higher education is turning to computer-based technology.

To address the paucity of empirically based research, to design research where in-class pedagogy and use of technology are fully integrated for a common pedagogical purpose, to account for a diversity of variables that can affect learning, and to build a more robust and sound research base that validates (or not) the blended-learning format, the study reported here investigated the effectiveness of technology within the foreign language-learning context. Specifically, the study examined learner performance within a redesigned intensive Spanish transitional course that met 2 days per week, emphasized interactive, collaborative, and content-based tasks, and used an online workbook as in the same course taught under the conventional paradigm\textsuperscript{1} that met 3 days per week and used a hard copy workbook.

REVIEW OF THE LITERATURE

Research on computer technology associated with second language learning, such as computer-assisted language learning (CALL), technology-enhanced language learning (TELL),\textsuperscript{2} or computer-mediated communication (CMC), has served a broad spectrum of purposes. Conrad (1996), Liu et al. (2003), and Salaberry (2001) provide overviews of computer technology research in second-language (SL) or foreign-language (FL) learning in the latter part of the 20\textsuperscript{th} century.

Conrad (1996) found that between 1992-1995, most of the empirical research related to computer technology and language learning focused on software control, feedback, pronunciation, and reading. Later CALL research investigated computer-facilitated talk and interaction (CMC), cooperative learning, process writing, and online reference tools.

In their literature review, Liu et al. (2003) report that the majority of publications between 1990-2000 on computer technology and SL learning were nonresearch based and tended to showcase the potential uses of CALL for developing SL/FL skills and cultural knowledge such as the following:

1. specific software programs to build specific language skills;
2. audio-video access to enhance reading and listening comprehension;
3. UseNet and email to access authentic cultural contexts;
4. electronic dictionaries for vocabulary acquisition;
5. voice interactive software for developing speaking skills;
6. interactive, real-world capabilities that more closely authenticate language uses;
7. multimedia benefits in distant education;
8. increasing interactive ability of technologies to simulate real-world tasks; and
9. easy access to authentic texts for cultural content.

In addition, much of their research describes uses and effects of existing software tools that enhance the SL-learning process, such as multimedia-authoring software, word-
processing software, the internet, and speech recognition software. Moreover, much of their research serves to discuss the advantages and disadvantages of using CALL as a tool for SL learning and testing.

Liu et al. (2003) reveal that of the 70 research-based papers on computer technology and SL learning, 65 tended to be quantitative and five were qualitative. The majority of publications examined topics such as text glossing, culture learning, learning strategies and styles, modes of interaction, or the ways in which computer technology promotes speaking, listening, reading, grammar, writing, or a combination of these skills. In general, the literature advocates the application of sound pedagogical and design principles, such as interactionist learning theory and learner-centered design guidelines and models, but most of this research does not permit generalization because of its broad range of research purposes. Chapelle’s (1997) review of CALL literature related to English as a second language (ESL) resonates with other researchers in that the ESL CALL research has also developed in diverse directions. Moreover, many studies on FL/SL learning measure only short-term outcomes. Lastly, much of the research treats CALL as a supplement to an existing curriculum, a tool that can enhance development of a specific skill, as opposed to a fully incorporated component of the FL program.

Salaberry (2001) examines articles published in the Modern Language Journal since 1916 and offers a succinct critique of CALL research. In Salaberry’s literature review, like those of Conrad (1996) and Liu et al. (2003), computer-mediated communication is spotlighted for its pedagogical, psychological and linguistic benefits, particularly the studies related to the Interchange component of the Daedalus Integrated Writing Environment (Beauvois 1995, 1998a, 1998b; Chun, 1994; Kern, 1995). Salaberry, however, adopts a more analytical perspective in his review as he underscores the complexities involved in research on CALL because of the disparity between technological sophistication and pedagogical effectiveness and the difficulty in separating variables in empirically-based research designs. He suggests using “carefully designed qualitative studies for the analysis of the multitude of factors that should be taken into account” and getting away from “technology driven pedagogy … to principle-oriented uses” (p. 51).

Thus, we are far from providing a body of evidence illustrating learning consequences of CALL. Moreover, we are hard pressed to find published articles that evaluate entire programs “intended to access language skill development and the integration of technology into the curriculum” (Adair-Hauck, Willingham-McLain, & Youngs, 2000, p. 270).

Research on Blended Learning

Recently the term “blended learning” has been used to describe the type of program that provides synchronous classroom contact with asynchronous online learning for the purpose of replacing some of the synchronous classroom seat time (Albrecht, 2006; Goertler & Winke, 2008). Blended learning can take a variety of forms due to (a) the array of functions it can serve, (b) the fluidity of technologies that exist, (c) the numerous ways to apply technology, and (d) the diversity of disciplines and ways courses are organized (Voos, 2003).

Research that replaces face-to-face contact time by online time was practically non-existent until 1997 when Adair-Hauck et al. (2000) set out to evaluate the effectiveness of integrating technology into a second-semester college-level French course. Two classes of second-semester French served as a control group (n = 16) and a treatment group (n = 17). The treatment group met 3 days per week and completed multimedia assignments, including TELL activities, outside of class. The control group met 4 days per week, one of those days
consisted of completing the same multimedia assignments in class. Both groups were taught by the same instructor and used the same textbook and supplementary materials. Students in both groups took five achievement tests measuring listening, reading, and writing skills and a quiz on knowledge of Francophone culture. The findings of this research indicated that “students in the treatment group performed equally well as the control group in listening and speaking and better on reading and writing achievement measures” (p. 269).

In his doctoral research, Kunz (1997) investigated replacing one hour of face-to-face contact time with a computerized tutorial; his research focused on the students’ study strategies, attitudes, achievement, and retention of German under several conditions. The subjects in his study were 100 students of German at the University of Texas enrolled in five beginning classes. He divided the students into three groups. The first group followed the conventional format of the course, which used a hard copy of the workbook; the second group used both the hard copy of the workbook and the computerized version of the workbook, referred to as Computerized Lernheft (CL), but spent less than 6 hours per chapter on CL. The third group also used the hard copy of the workbook but spent over 6 hours per chapter on CL. Kunz found that the students in the second and third groups had significantly higher gains in measures of German short-term achievement scores but that the difference was not significant for long-term achievement scores. The latter findings contribute to the “no significant difference” phenomenon that is also consistent in much of the CALL research.

Around the same time, Musumeci (1999) reported results of a grant-funded project designed to address expanding demand of Spanish courses at the University of Illinois. A 16% increase in the demand for Spanish courses over several years in conjunction with an impending campus-wide foreign language requirement far exceeded the university’s seating capacity and the department’s pool of graduate teaching assistants (GTAs). Musumeci proposed a reconfiguration of two lower level courses—Spanish 122, an intensive first-year Spanish course and Spanish 210, a Spanish grammar course—which traditionally met 4 days per week. She proposed the replacement of two face-to-face meeting times by online instruction. The grant allowed the university to convert the hard copy of the workbook used in the basic Spanish classes to an online version using Mallard developed at the University of Illinois (Twigg, 1999). In this way, in-class instruction could focus on advancing speaking skills and refining listening skills. Writing, using online journals via FirstClass (a commercial course-management system that includes a conferencing feature), reading, and additional listening activities were moved to the online component. The online component also provided grammatical explanations and practice with instant feedback to students. The GTAs continued to have 4 contact hours per week by teaching two classes per week instead of one, thereby meeting the needs of 40 students instead of 20.

To assess student performance in Spanish 122, researchers compared students’ scores in the blended-learning format and the conventional instructional format on the university’s Spanish placement exam and the departmental exams. Student gain scores on the Spanish placement exams “reveal that students in the technology-enhanced format of Spanish 122 show significantly larger gains on the placement exam than did students in the conventional format” (Musumeci 1999, pp. 5-6). No significant differences in students’ scores were found, however, between the two formats on the departmental listening and writing tests. No data were gathered on speaking skills. For Spanish 210, Musumeci (1999, p. 5) reported that the students in the blended-learning course performed significantly better than those in the conventional course in one semester but not in another semester.

By the time of the new millennium, an upsurge of blended-learning research surfaced supported by university grants or national foundations. In the spring of 2000, the Language Online project at Carnegie Mellon investigated the effectiveness of online learning. Recogniz-
ing the importance of individual variables (Dunkel, 1991), Chenoweth and Murday (2003) investigated students’ FL learning and their satisfaction with their learning in two instructional formats taught by two different instructors. A traditional elementary French class \( (n = 12) \) met for four 50-minute classes per week and used material created or photocopied by the instructor. An online elementary French class \( (n = 8) \) met for 50 minutes once a week and were required to participate in an hour long chat session in addition to corresponding with classmates via email and discussion board postings. Results of listening, reading, and speaking tests were not significantly different between the groups. However, the students in the online group obtained significantly higher median scores (34.5) on writing than those in the conventional group (31.5).

Chenoweth, Ushida, and Murday (2006) designed an extensive 2-year study on the effectiveness of learning foreign languages in a blended format. They examined learner performance outcomes of two elementary and intermediate French and Spanish courses in blended versus conventional formats, again as part of the Language Online Project at Carnegie Mellon. A total of 13 sections followed the blended format, and 21 sections continued with their conventional format. The elementary conventional courses met 4 days per week and the intermediate conventional courses met 3 days per week. In the blended format, learners met face to face for 1 hour a week and 20 minutes a week with a FL major or native speaker undergraduate. In addition, they participated in weekly chats (approximately 20 minutes a week) and completed online exercises, quizzes, and bulletin board postings. The researchers assessed learners’ language performance in various components of the final exam, such as listening and reading comprehension, grammatical and vocabulary knowledge, and written and oral production. Quantitative findings indicated that there were no significant differences among most of the components of the final exams, but when there were significant differences, they favored the learners in the conventional format.

At the same time, the Center for Academic Transformation’s PEW Grant Program in Course Redesign (http://www.thencat.org/PCR.htm) selected a handful of universities in various disciplines to research online learning effectiveness with cost-saving outcomes. Sanders (2005) redesigned a series of beginning Spanish courses at Portland State University to ascertain whether learning outcomes could be maintained by reducing face-to-face class time from 200 minutes to 130 minutes per week and moving most of the grammar, vocabulary, writing, and reading exercises to an online format and adding a CMC component. The experimental group had class sizes of 25 students that met 130 minutes per week. Instruction in these courses consisted of class discussions—both in class and online—of textbook topics, online compositions based on chat discussions, online readings on culture, and a “presentation of a group project in a poster session at [the] end of each quarter” (p. 527). The traditional group had class sizes of 30 students that met 200 minutes per week. Instruction for these courses consisted of discussions of textbook topics, homework compositions based on the in-class discussions, and group reading assignments on culture. Sanders compared learners’ proficiency and achievement scores/ratings between the two groups. Of 1,278 students in the project, 307 volunteers took the Brigham Young WebCAPE to measure FL achievement (in grammar, vocabulary, and reading comprehension,) and 57 students completed the American Council on the Teaching of Foreign Languages (ACTFL) Oral Proficiency Interviews (OPIs) and the writing proficiency test. Sanders found no significant differences in the WebCAPE scores nor in the OPI ratings between the groups. Students in the traditional group did perform significantly higher than those in the experimental group on writing.

The above studies begin to investigate the learning effects of mixed delivery systems in FL learning, but these researchers had to conduct research within specific program constraints and were unable to account for the multitude of variables that can affect program outcomes. To undertake research with the complexity inherent in assessing the learning outcomes of an
The entire program is an ambitious and costly venture. It stands to reason why Voos (2003, p. 2) argues that to date “rigorous (let alone standard) methodologies to assess outcomes remain elusive.”

Challenged to conduct the type of research Voos (2003) would characterize as “rigorous,” we collected data spanning 2 years in an effort to address the large number of variables related to empirically based research encompassing program evaluation. One purpose of the present study was to address the limitations of previous research. For example, we (a) ensured that the two groups were equal in terms of student performance from the start; (b) sought a sizeable number of subjects in the experimental and conventional groups; (c) used instruments with established reliability and validity and obtained pre- and posttreatment scores from them; (d) assessed a variety of language skills including speaking; and (e) maintained in-class pedagogical principles based on interactionist learning theory that reflected learner-centered, content-based, and task-based models and encouraged negotiation of meaning. In addition, after the first implementation phase of our project, we designed a follow-up study the following year to inform our initial findings.

THE STUDY

Similar to Portland State University, the University of Tennessee was awarded a Center for Transformation’s PEW Foundation grant to redesign an intensive Spanish transition course (Spanish 150) to provide data on the effectiveness of instructional approaches using technology and simultaneously generate cost savings for the university. Because of space limitations, the present article is limited to reporting the empirically based quantitative research component of the project.

The Intensive Spanish Transition Program

Spanish 150 in the Department of Modern Foreign Languages at the University of Tennessee is designed for students who studied Spanish but do not score high enough on the Spanish placement test to place into the second-year (third-semester) course. Spanish 150 reviews in one semester what is covered in two semesters for beginning-level students. Over 60% of entering freshmen consistently place into this course and contribute to the oversubscription of the course. The range in student preparedness for Spanish 150 covers students who took Spanish many years ago and remember very little or nothing to students who come directly out of high school having taken their second year in the 10th grade to the 12th grade.

The pedagogical framework of ¿Sabías que? (VanPatten, Lee, & Ballman, 2000), the textbook used in both instructional formats of the present study, is rooted in task-based, content-oriented instruction within an input-processing approach. Lecturers and GTAs often expressed preferences for teaching first- (111 or 112) or second-year courses (211 or 212) over Spanish 150 because of the fast-paced nature of the course. They often commented that the pedagogical framework of the textbook was compromised because of the lack of time for the communicative activities. They also struggled with staying in the target language claiming that using the native language “just saves time.”

Since we offer only four sections of beginning Spanish courses, most GTAs experience their teacher training in Spanish 150. The more pedagogically informed and experienced teaching personnel are often called on to teach second-year courses. Less experienced teaching staff usually do not teach second-year until they have more experience teaching...
and greater language proficiency. GTAs may teach independently only after accumulating 18 graduate hours. By the time graduate students complete their MA, they have had bare-bone limited experiences (teaching only two classes on their own—one in their third semester and one in their fourth semester). The Ph.D. students have more teaching experience, but the reading demands in their own courses often overwhelm many of them, thus limiting their ability to attend professional development functions related to language teaching.

In the first implementation of this study, we selected five GTAs and lecturers that were capable of reaching the pedagogical objectives of the redesigned course. On the one hand, they continued their same pattern of teaching as in the traditional format, which meant including the input activities before any communicative activities and sometimes offering explicit grammatical instruction. On the other hand, these GTAs and lecturers also successfully implemented the pedagogical objectives of the redesigned course because they understood the in-class goals of the course, for example, understanding the purpose of jigsaw exercises, one- or two-way information-gap tasks, and signature search activities. They also encouraged negotiation of meaning in online communicative homework activities.

At the time we began the first phase of this study, there were no commercially available online workbooks, and the only noncommercial online workbook was one developed at another university that could be used only at that university. Consequently, personnel in the Instructional Technology Center at the University of Tennessee and the Spanish Language Program Coordinator in the Department of Modern Foreign Languages joined resources to create an online version of the workbook accompanying ¿Sabías que?. In addition to the workbook exercises, input and forced choice listening comprehension exercises were put online. We spent an entire fall semester transferring and adapting exercises from the commercial workbook and some textbook input activities into Blackboard and then implemented the research project the following spring.

In addition to measuring students’ FL performance on the basis of course formats, we gathered data on multiple variables (e.g., computer anxiety and use of learning strategies) that could potentially interact with their performance; we also collected qualitative data that could help inform blended-learning programs in general (e.g., student perceptions of their learning experience, instructor reactions to the changes in course format, in-class instructor practices, and online student behaviors). Moreover, we videotaped all classes to compare and contrast the two formats. In this article we focus exclusively on students’ FL performance.

**Research Question**

The research question framed by the null hypothesis is the following:

There is no significant difference in group means on three measures of learner performance in Spanish between the experimental group (redesigned course) and comparison group (conventional course) from which we drew our two samples.

**Subjects**

The subjects in this study were 209 university level students enrolled in an intensive one-semester review course, Spanish 150, which covered the curriculum of the entire first-year Spanish in one semester. A total of 107 females and 98 males participated in this study.
dent ages ranged from 17 to 42 with an average of 19 for both groups. While there were a few students who declared Spanish as their minor, there were no Spanish majors. Most had an average of 2-3 years of high school Spanish before enrolling in Spanish 150. No students indicated that Spanish was spoken in their home.

**Procedures**

Ten sections of Spanish 150 were divided into two groups, a comparison group and an experimental group. The same instructors taught each version of the course in order to control for instructor effects. Each class had from 22-26 students. All sections were set up so that a section of the comparison group and experimental group met at the same hour of the day, and each section had an equal number of students who had earned grades of A, B, C, and D in previous Spanish courses. Students who did not wish to use the online materials were placed into the corresponding section of the comparison group taught at the same hour; only 4 students took this option.

The conventional group met three days per week and used the hard copy of the workbook that accompanied ¿Sabías Que?. The experimental group met two days per week and completed the same workbook exercises online, with some modifications to meet the activity formats of Blackboard. In addition, we recorded the forced choice input-based exercises from the textbook and moved them online and adapted the communicative workbook writing activities into collaborative asynchronous writing assignments. These were called BlueBook assignments because students had to print out their responses and place them in a BlueBook. The information from these assignments, as well as information collected from interactive tasks in class was used to formulate the composition topics of the midterm and final exams. Lastly, in the experimental group we created collaborative homework assignments, referred to as CHAs, which required students to negotiate decisions online via email or the chat function of Blackboard. There were two parts to turn into the instructor from the CHA assignments: the online conversation, so to speak, in which students made joint decisions about the assignment, and the final product, such as a brochure, menu, poster, and so on. Thus, both process and product were equally important in these CHA assignments.

In both formats, students were asked to come to class prepared by examining the new material in the textbook, such as reviewing the vocabulary or reading the grammar. In both formats, in-class communicative approaches were encouraged.

In the first 2 weeks of the semester, participants in the study signed a consent form, completed a background information questionnaire, and took three pretests: a diagnostic measure—the University of Tennessee Spanish Placement Exam—and two proficiency measures—the reading and listening tests of the Minnesota Language Proficiency Assessments (MLPA). The midterm and final exam for Spanish 150 served as measures of achievement. In the last 2 weeks of the semester, participants took a different version of the Spanish Placement Exam (reversed order) and the same MLPA tests as posttests. Lastly, students completed a Simulated Oral Proficiency Interview (SOPI) in Spanish.

**Test Instruments**

**Proficiency measures**

Proficiency is defined as what a subject can actually do in a foreign language (how much
learners understand and can produce, irrespective of their number of hours, course work, etc.). The notion of proficiency stems from the ACTFL guidelines, which were designed to determine language ability using a nationwide common yardstick.

**Listening and reading proficiency**

The University of Minnesota’s Center for Advanced Research on Language Acquisition developed a battery of tests, the Minnesota Language Proficiency Assessments (MLPA), to assess reading, writing, speaking, and listening of both high school and college students via computer. We administered the Contextualized Reading Assessment (CoRA) and the Contextualized Listening Assessment (CoLA) to measure proficiency in reading and listening as pre- and posttests. The CoRA consists of a 35-item multiple-choice format with an average of 12 texts of authentic readings on a variety of topics from a range of sources united by a similar theme that attempts to engage the readers to interact with the readings. The reliability coefficients for the CoRA range from .89 to .92. The CoLA is comprised of minidialogues linked to one story line with characters that “engage in a variety of real-life interactions ... contextualized through the use of photographs and advance organizers” (MLPA, p. 16). The format is also multiple choice, and test takers can control when each segment is played and can listen to each one once or twice.

**Speaking proficiency**

The Center for Applied Linguistics developed the SOPI, a tape-mediated test of speaking proficiency based on the ACTFL guidelines. The test consists of a series of prompts, some referring to pictures and others offering situations to which the learner must respond within a specified time limit. This test offers a standardized approach to testing oral proficiency that is cost effective and not susceptible to subjectivity in its administration because the exam is executed via a tape and responses are recorded and rated by qualified raters. The SOPI is used by ACTFL and government agencies to measure general speaking proficiency in numerous languages. Research indicates that the SOPI is a valid and reliable test of speaking (Clark & Li, 1986; Stansfield, Kenyon, Paiva, Doyle, Ulsh, & Cowles, 1990; Shohamy, Gordon, Kenyon, & Stansfield, 1989; Stansfield & Kenyon, 1992, 1993).

**Spanish placement exam**

The departmental Spanish Placement Exam was designed and tested for interrater and intrarater reliability by the previous FL Coordinator and has been used to place students at the appropriate course level for over 15 years. It consists of 5 parts: listening, reading, language functions, vocabulary, and grammar.

**Achievement tests**

The Departmental midterm and final exams were used to assess how well the students learned class material in the textbook and workbook. The format of the exams was modeled after textbook, workbook, and in-class/online exercises and tasks. All students in the project received a description of the exam format 2 days before the exams. The exams assessed listening comprehension, vocabulary, grammar, culture, and writing.
RESULTS

Measures of Spanish Performance

For each of the three measures of language performance—proficiency, diagnostic, and achievement, t tests were used to determine whether there was a significant difference between the means of the two groups.

Proficiency Assessments

Pretest results

The mean for the listening component of the MLPA was 13.8 for the comparison group and 13.9 for the experimental group, and a t test showed no significant differences in mean scores between the two groups. The comparison group mean for the reading component of the MLPA was 19.9 and 18.8 for the experimental group. A t test on these mean scores again showed no significant differences between the two groups. The mean for the Spanish Placement Exam was 29 for the comparison group and 30 for the experimental group, but, once again, a t test showed no significant differences between the groups. Thus, for both proficiency and diagnostic measures, the pretest results indicated that the two groups were not statistically different from each other at the beginning of the study.

Posttests results: Proficiency

While participants in both groups had higher posttest scores on the listening and reading components of the MLPA (2-point gain score for listening and 3-point gain score for reading), t test results indicated no significant differences between the groups (see Table 1).

<table>
<thead>
<tr>
<th>Test</th>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening</td>
<td>Comparison group</td>
<td>83</td>
<td>15.74</td>
<td>4.9</td>
<td>0.26</td>
<td>.780</td>
</tr>
<tr>
<td></td>
<td>Experimental group</td>
<td>77</td>
<td>15.96</td>
<td>5.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>Comparison group</td>
<td>71</td>
<td>21.50</td>
<td>6.0</td>
<td>-0.28</td>
<td>.770</td>
</tr>
<tr>
<td></td>
<td>Experimental group</td>
<td>70</td>
<td>21.30</td>
<td>5.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Posttest results: Spanish placement exam (diagnostic test)

Again, while participants in both groups had higher posttest scores on the Spanish Placement Exam (6-point gain score for the comparison group and 4.2-point gain score for the experimental group), t-test results indicated no significant differences between the groups (see Table 2).
Baseline data accumulated over the past 5 years suggest that students scored no differently on the Spanish Placement Exam at the end of this course than at the end of the course in previous semesters. The mean score for the Spanish Placement Exam for Spanish 150 in previous years was in the range of 34-35.

**Posttest results: Achievement tests (midterm and final exams)**

The results of $t$ tests on the midterm exam (see Table 3) indicated a significant difference in mean scores between the comparison and experimental groups with the comparison group scoring higher (4.8 points, $p = .003$).

| Table 3 |
|-----------------|--------|--------|--------|--------|--------|
| Results of $t$ Test for Midterm Exam |
| $N$ | $M$ | $SD$ | $t$ | $p$ |
| Comparison group | 70 | 85.8 | 8.7 | -2.94 | .003 |
| Experimental group | 73 | 81.0 | 10.3 | |

A $t$ test on the final exam (see Table 4), however, indicated no significant difference in scores between the two groups, although the mean for the experimental group was 2.7 points higher than that of the comparison group.

| Table 4 |
|-----------------|--------|--------|--------|--------|--------|
| Results of $t$ Test for Final Exam |
| $N$ | $M$ | $SD$ | $t$ | $p$ |
| Comparison group | 69 | 77.4 | 9.7 | 1.41 | .158 |
| Experimental group | 68 | 80.1 | 12.0 | |

**Simulated Oral Proficiency Interview (SOPI)**

The lack of significant differences between the groups on the MLPA reading and listening exams suggested that we would probably not see any differences between the two groups on the SOPI. The SOPI measures speaking performance globally and can include a wide range of performance-based tasks on a nonincremental scale. Consequently, and in consultation with the project’s external evaluation consultant, we decided to analyze the SOPI data using a finer grained approach that “would stand a better chance of identifying differences between the groups that could logically be connected to the instruction provided to the two groups” (Liskin-Gasparro, 2001, p. 2).

To ensure that the results of the analysis in this finer grained approach were not skewed, we identified approximately 20% of the total number of participants comprised of students in three categories—high, mid, and low—on the basis of their scores on the Span-
lish Placement Exam, MLPA in reading and listening tests, and the midterm and final exam. Participants who scored consistently high on all language measures formed the high group, students who maintained consistent midrange scores formed the middle group, and students with low scores the low group.

We then transcribed the SOPI tapes for these students in the aggregate to investigate their completion of the task (i.e., answer the questions) in spite of inaccuracies. Table 5 shows the success/lack of success in the SOPI for both the comparison and experimental groups.

Table 5
Task Completion

<table>
<thead>
<tr>
<th>Sect.</th>
<th>Question</th>
<th>Comparison group (n = 25)</th>
<th>Experimental group (n = 29)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>% completing task</td>
<td>% not completing task</td>
</tr>
<tr>
<td>1</td>
<td>1. ¿Cómo te llamas?</td>
<td>80</td>
<td>20</td>
</tr>
<tr>
<td>1</td>
<td>2. ¿De dónde eres?</td>
<td>72</td>
<td>28</td>
</tr>
<tr>
<td>1</td>
<td>3. ¿Estudias o trabajas?</td>
<td>96</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>4. ¿Qué haces en tu tiempo libre?</td>
<td>16</td>
<td>84</td>
</tr>
<tr>
<td>1</td>
<td>5. ¿Hace cuánto tiempo estudias español?</td>
<td>44</td>
<td>56</td>
</tr>
<tr>
<td>1</td>
<td>6. ¿Por qué estudias español?</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>1</td>
<td>7. ¿Con quién hablas español?</td>
<td>16</td>
<td>84</td>
</tr>
<tr>
<td>2</td>
<td>1. Pregunta a Alberto algo sobre Carlos</td>
<td>68</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>2. Pregunta a Alberto sobre su escuela</td>
<td>80</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>3. Pregunta a Alberto sobre sus actividades</td>
<td>64</td>
<td>36</td>
</tr>
<tr>
<td>2</td>
<td>4. Pregunta a Alberto sobre su novia Susana</td>
<td>72</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>1. ¿Qué hace la gente típicamente en el mar en EU?</td>
<td>84</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>1. Describe un día típico de un estudiante en EU</td>
<td>52</td>
<td>48</td>
</tr>
<tr>
<td>5</td>
<td>1. Una mujer te pregunta cómo llegar al museo</td>
<td>92</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>1. Describe la situación de Carmen (compras)</td>
<td>4</td>
<td>96</td>
</tr>
</tbody>
</table>

The questions on the SOPI are organized by levels and within those levels increasing degrees of difficulty. For the most basic questions (questions 1-6 in Table 5), the experimental group outperformed the comparison group. In level two, the experimental group outper-
formed the comparison group on two questions, and the comparison group outperformed the experimental group on the other two questions. For the remaining levels (3-6), the same pattern of successful completion as that found in level two was repeated.

We also analyzed the SOPI transcriptions to ascertain whether there were any differences between the groups with respect to the number of words students produced, the range of students’ vocabulary (number of language chunks, nouns, verbs, modifiers, and prepositions), the number of students’ language repairs, and number of successful task completions. Table 6 summarizes the results of this analysis.9

Table 6
Paired-sample t Tests of Student Performance on the SOPI (N = 51)

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of words</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison group</td>
<td>120.88</td>
<td>5.2</td>
<td>-1.88</td>
<td>.06</td>
</tr>
<tr>
<td>Experimental group</td>
<td>148.34</td>
<td>5.4</td>
<td>-1.88</td>
<td>.06</td>
</tr>
<tr>
<td>Range of vocabulary</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language chunks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison group</td>
<td>2.46</td>
<td>4.8</td>
<td>-1.96</td>
<td>.05</td>
</tr>
<tr>
<td>Experimental group</td>
<td>3.61</td>
<td>7.7</td>
<td>-1.96</td>
<td>.05</td>
</tr>
<tr>
<td>Nouns</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison group</td>
<td>3.00</td>
<td>6.9</td>
<td>-5.40</td>
<td>.00</td>
</tr>
<tr>
<td>Experimental group</td>
<td>4.60</td>
<td>6.2</td>
<td>-5.40</td>
<td>.00</td>
</tr>
<tr>
<td>Verbs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison group</td>
<td>2.21</td>
<td>6.7</td>
<td>-2.92</td>
<td>.00</td>
</tr>
<tr>
<td>Experimental group</td>
<td>2.78</td>
<td>6.6</td>
<td>-2.92</td>
<td>.00</td>
</tr>
<tr>
<td>Modifiers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison group</td>
<td>2.09</td>
<td>4.9</td>
<td>-1.16</td>
<td>.24</td>
</tr>
<tr>
<td>Experimental group</td>
<td>2.38</td>
<td>4.6</td>
<td>-1.16</td>
<td>.24</td>
</tr>
<tr>
<td>Prepositions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison group</td>
<td>3.50</td>
<td>7.1</td>
<td>.90</td>
<td>.40</td>
</tr>
<tr>
<td>Experimental group</td>
<td>1.17</td>
<td>1.1</td>
<td>.90</td>
<td>.40</td>
</tr>
<tr>
<td>Successful Repairs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison group</td>
<td>1.44</td>
<td>1.1</td>
<td>-0.38</td>
<td>.04</td>
</tr>
<tr>
<td>Experimental group</td>
<td>1.59</td>
<td>1.5</td>
<td>-0.38</td>
<td>.04</td>
</tr>
<tr>
<td>Task completion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison group</td>
<td>14.60</td>
<td>7.2</td>
<td>-3.63</td>
<td>.00</td>
</tr>
<tr>
<td>Experimental group</td>
<td>17.60</td>
<td>8.6</td>
<td>-3.63</td>
<td>.00</td>
</tr>
</tbody>
</table>
Students in the experimental group outperformed the students in the comparison group in all categories except modifiers and prepositions. In other words, the language of the students in the redesigned course format with online activities demonstrated broader and more varied language use and completed more of the oral tasks than the students in the traditional instructional format.

DISCUSSION
In this study, the achievement tests (midterm and final exams) detected more differences than the proficiency tests and the diagnostic (placement) test and showed that students in the experimental group outperformed those in the comparison group on most measures in the achievement tests. However, results also indicated that the students in the comparison group outperformed those in the experimental group on the midterm exam. One explanation for this result rests with the technological problems we encountered at the beginning of the semester when we implemented the online workbook (related to the capacity of the server). This interpretation is supported by analysis of both student and instructor qualitative data. At the beginning of the semester, students and instructors alike complained about technology-related problems. The server was extremely slow; the program often crashed, and students who had begun an exercise could not return to finish it because the computer reported it as having been completed. Once students had submitted a completed exercise, they had to wait as long as 15-20 minutes for the computer to make the workbook available again for them to use. Instructors were frustrated at having to reset exercises to enable students to finish or redo an exercise. These problems lasted approximately for the first 3 to 4 weeks of the spring semester. We also suspect that the higher mean on the final exam by the experimental group may indicate that once the issues with Blackboard had been resolved, the positive effects of online learning began to emerge.

The unexpected results that surfaced from the analysis of the SOPI data indicate that the instructors successfully followed the pedagogical guidelines for the 2-day per week redesigned course in which communicative and interactive tasks became the hallmark of in-class activities. Instructors were able to get to the open-ended activities in the textbook much sooner in class because the input-based listening exercises were to be done online before coming to class. When asked about the strengths of in-class practices, students in the experimental group consistently confirmed that they were doing interactive activities in class. Instructors also observed that students in the experimental group came to class more prepared than the students in the comparison group. Instructors explained that a lack of preparation from a single student could influence another student or even a group of students. In other words, peer pressure from other students to complete the online work could have motivated them to come to class better prepared.

LIMITATIONS
Even though this research sought a more rigorous design than previous research in that we attempted to account for variables that had been overlooked in previous designs, this research, similar to Sanders (2005) and Chenoweth and Murday (2003), was constrained by staffing issues. At the University of Tennessee, the GTA stipend allows GTAs to teach up to two classes. Since we wanted to control for the instructor variable by having the same instructor teach sections in both groups and given that we focused on the redesign of the traditional 150 course, we limited our research design to two groups. By doing so, we could not differentiate the effects of three variables in the project. We could not ascertain whether the differences between the conventional course and the redesigned course could be explained by (a) contact
time, (b) the successful application of interactive in-class pedagogy, or (c) the effectiveness of the online component in the redesigned course. We should have included a third group, one that met 2 days per week, used the hard copy of the textbook’s workbook, and encouraged the communicative/interactive pedagogy of the redesigned course. Consequently, we set out to inform the results of this first implementation of the redesigned course by a follow-up study the next semester (see synopses of the two implementations in the appendix to this article).

FOLLOW-UP STUDY
In the follow-up implementation of the redesigned course, 34 sections of Spanish 150 were divided into two groups. One group of 19 sections taught by instructors, met 3 days per week, and used the online workbook. The other group of 15 sections were taught by GTAs and met 2 days per week and also used the online workbook. All personnel in the Spanish 150 program were instructed to emphasize interactive and communicative activities. Training toward this objective (in addition to Blackboard training) took place the week before classes began in our "preservice" workshops.

In the second week of the fall semester, we administered the listening and reading components of MLPA as a pretest to ascertain that the two groups were not significantly different from each other at the outset. We again administered the same midterm and final exams to both groups to measure achievement gains. To reduce the amount of time we were taking from an already intensive program, we did not administer the Spanish Placement Exam or assess speaking in this second implementation.

PHASE II IMPLEMENTATION

Proficiency Measures
After doing a t test on the MLPA listening test, we found no significant difference in scores between the groups. A significant difference did surface, however, in the reading comprehension pretest favoring the 3-day per week group (See Table 7). That difference leveled out by the end of the semester; no significant differences were found in the reading or the listening posttests between the groups.

Table 7
\begin{tabular}{lrrrr}
 & N & M & SD & t & p \\
2-day group & 234 & 18.1 & 5.0 & -2.43 & .01 \\
3-day group & 422 & 19.2 & 4.7 & & \\
\end{tabular}

Achievement Tests
As in the first implementation, we examined students’ scores on the midterm and final exams to ascertain any difference in achievement.

Midterm exam
A t test indicated that there was no significant difference in midterm exam scores between the 2-day and 3-day per week groups (See Table 8).
Table 8
\(t\)-test Results on Midterm Exam by Group

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-day group</td>
<td>287</td>
<td>81.7</td>
<td>12.0</td>
<td>-1.23</td>
<td>.21</td>
</tr>
<tr>
<td>3-day group</td>
<td>316</td>
<td>82.9</td>
<td>12.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Final exam**

A \(t\) test on the final exam scores by group did indicate a significant difference in final exam scores (See Table 9). The three-day per week group scored significantly higher (by almost seven points) than the two-day per week students.

Table 9
\(t\)-test Results on Final Exam by Group

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-day group</td>
<td>282</td>
<td>73.3</td>
<td>12.5</td>
<td>-6.70</td>
<td>.00</td>
</tr>
<tr>
<td>3-day group</td>
<td>306</td>
<td>80.4</td>
<td>12.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**DISCUSSION OF FOLLOW-UP STUDY**

During the first few weeks of the follow-up study, we experienced similar technological problems (again related to the capacity of the server). To explain the results of this follow-up study, we suspect that the material covered in the first part of the Spanish 150 curriculum was basic enough for both groups to master, thus neutralizing the effects of face-to-face contact. The second part of the Spanish 150 curriculum, however, is traditionally more complex and challenging for students. This could explain why contact time became more significant, as manifested in the significant differences in the final exam scores.

The results in the follow-up study controlled for the pedagogical variable (both groups were to emphasize collaborative and communicative, task- and content-based learning) and the workbook format (all students used the online version of the workbook). However, a variable we could not address was that of the instructor, a limitation that Sanders (2005) also expresses in his research. Recall that lecturers taught the 3-day per week courses and the GTAs taught the 2-day per week classes. Hints that the instructor variable needed to be examined surfaced when we compared the mean scores on the tests in the first and second implementation. A comparison of these means revealed that the means were higher on all measures for both groups in the first implementation. As we commented earlier, in the first implementation, we selected only the most pedagogically informed instructors to teach the classes. In the second phase of implementation, the background of GTAs and instructors was more varied. The fact that the groups were assigned according to whether a GTA or instructor taught the course meant we could not control for this variable. In other words, we cannot attribute any differences in performance based on contact time because of the uncontrolled instructor variable.

While we were unable to conduct a formal study after the second implementation, we remained curious about whether classes taught by a GTA or instructor would make a difference in students’ performance. Consequently, in the semester after the second implementation we examined students’ midterm and final exam scores on the basis of the instructor variable. In that semester all 18 sections of Spanish 150 met 3 days per week using the redesigned (blended-learning) format. All students used the online workbook and followed
the pedagogical guidelines for in-class activities. Advanced GTAs taught seven sections of Spanish 150 and lecturers taught 11 sections. The midterm mean for the GTA-taught classes was 79 and for the lecturer-taught classes 80.9. A t test on midterm scores by group (GTAs versus lecturers) indicated that there were no significant differences between the two groups (t = -.978, p = .344). We found, however, a significant difference between the two groups on the final exam scores (t = -4.88, p < .001). On the final exam, the GTA-taught classes had a mean of 71.5 and the lecturer-taught classes 80.5, suggesting that the instructor variable could be a significant variable and should not be factored out of any study on learning gains.

GENERAL DISCUSSION
The collective findings of the two implementation phases are summarized in the appendix. Under the redesigned course format, a 2-day per week course could produce student outcomes equivalent to a 3-day per week format, particularly if talented or experienced personnel were to teach the redesigned course and the pedagogy of the course shifted to an approach featuring learner-centered, interactive, communicative, collaborative, task-based, content-oriented learning. In the second implementation, we suspect that the pedagogical changes effected in both the in-class and online format of the redesigned course probably worked in concert with the instructors’ pedagogical effectiveness to help compensate for any loss in contact time.

Without doubt, the best case and more desirable language-learning scenario would be to maintain the maximum amount of contact time, adopt an online workbook format, and train teaching personnel to rethink their in-class practices and adopt practices suggested by current research findings in second language acquisition. For those institutions that need to reduce costs, a blended-learning course might be an option, especially if the institution has well informed teaching personnel. The blended-learning format is particularly attractive for those programs with many students who do not intend to continue their language learning beyond meeting their institution’s FL requirements.

Well informed and experienced lecturers and GTAs capable of applying research-based shifts in teaching pedagogy may well be necessary to reduce face-to-face instruction. Unless there is strong faculty and administrative support to ensure that teaching personnel, whether GTAs or lecturers, obtain continuous professional development and training in teaching and technology, learner outcomes could be seriously compromised.

If this sounds simplistic, a reality check may be in order. The most common graduate school scenario is the GTA who teaches as a means to achieve a M.A. or a Ph.D. in literature. Many of these graduates go on to teach at 2-year or 4-year colleges and private or public universities. Thus, the majority of lecturers with degrees in literature who teach the bulk of first- and second-year courses at postsecondary institutions have had at most only one course in second language acquisition or FL teaching. Consequently, and all too often, a communicatively oriented textbook is transformed into a series of conventional practices supported by a limited number of teacher-prepared grammar-oriented mechanical worksheets.

The wave of applied linguists who were hired in the early 1980s to train GTAs and teaching personnel, especially in large state institutions, are confronted with this scenario and faced with effecting change via a single methods class, a handful of observations, and sparsely attended workshops. There is a tremendous disparity between what graduate programs require of their graduate students and the reality that those students will face in the profession. In a report on the “Job Crisis and Graduate Studies,” The Modern Language Association (2005) reports:
disparity between the expectations and assumptions about college teaching that most graduate programs inculcate in their Ph.D. candidates and the actual work most of those candidates will do once they leave the research-oriented Ph.D.-granting institutions where most of them have studied.

Unless we change this scenario and offer a more realistic balance and diversity of course offerings at the graduate level so that teaching personnel are more informed about second language acquisition theory, research, and practice, advances in technology and blended learning will fall short of their full potential.

CONCLUSION
In the introduction of this article, we made reference to how quickly new advances in technology surface. In a relatively short period of time (approximately 5 years), the demand for online workbooks spread rapidly based on apparent benefits of this format, for example, reduced grading for the instructor, anytime-anywhere access, instant feedback, increased attraction in multimedia components, interactive online exchanges, and potential cost savings (Dede, 2000). Today, commercially available online workbooks have become a standard option in the first- and second-year textbooks and facilitate the implementation of blended-learning programs. Nevertheless, a body of sound research on blended learning has yet to develop, and with good reason. The amount of work involved in programmatic evaluations of learner outcomes is complex; it requires “a village,” so to speak, is costly, and can be tremendously time consuming for those involved, particularly program supervisors and coordinators.

Understandably, the research on blended learning will continue to lag behind consumption, but the “no significant difference” (Twigg, 2001) that characterizes the sparse research that does exist will be cited as a stamp of approval for “no harm done.” While we strive to move beyond this, until educational funding is given a higher priority by state and federal agencies, it is incumbent upon academics to find ways to reduce costs without hindering learning.

To move beyond the “no significant difference” phenomenon may mean taking into account variables that affect learning, such as the effectiveness of the instructor and in-class pedagogy. We suspect that both were crucial in compensating for a reduction in face-to-face learning in the present study. Moreover, learner variables, such as students’ cognitive styles, comfort level for independent learning, self-discipline, and motivation to learn are important factors, just as they are in a conventional class, but probably even more in a blended-learning format because most learners find it difficult to learn independently. Future research in these areas is warranted, and we hope will continue to be funded by centers, foundations, and government agencies.

ACKNOWLEDGMENTS
This research was made possible through The Center for Academic Transformation’s PEW Foundation Grant. Dr. Julie Little, Dr. Jean Derco and Alec Riedl from the Instructional Technology Center contributed in significant ways to acquiring, managing, and developing the online material. I also wish to acknowledge the many graduate students and lecturers, too many to enumerate, who participated in the various phases of this project.
NOTES

1 The conventional classes often reviewed grammar explicitly, but the pedagogical framework was determined by the textbook, ¿Sabías que? which has a content-based, task-oriented pedagogical framework with an emphasis on interactive pair/group work.

2 Unlike CALL where the use of technology is treated as an “add-on” component, TELL fully integrates technology into the curriculum.

3 The research from this project has not been published. The author acquired a copy of the report through personal correspondence with Dr. Musumeci.

4 Mallard was originally designed by the University of Illinois at Urbana-Champaign as a computer-based testing instrument.

5 Sanders’ research was funded by The Center for Academic Transformation’s PEW Foundation. His 2005 article includes an account of how the redesigned course resulted in an increase in student enrollment of 85%, a 29% cost saving per student, and a 9% pay increase for course instructors. A substantial cost savings for the research described in the current article occurred in the fourth implementation phase and will be published in another article.

6 Articles that are more qualitative in nature are in the process of being written.

7 The personnel in Instructional Technology Center were Dr. Julie Little, Dr. Jean Derco, and Mr. Alec Reidl.

8 Dr. William Heflin, Coordinator of the lower division program at the time, wrote the test and collected reliability and validity information for the test in 1989 and revised it in 1993.

9 Interater reliability levels of .92 were established between the graduate student who scored the majority of the SOPI data and the researcher.

10 Analysis of qualitative data confirms this. We administered a questionnaire around the 13th week of class to ascertain students’ perceptions of their experiences with the online workbook and in-class lessons.

11 The administration and Spanish faculty would not have supported lecturers teaching 2 days a week and some GTAs teaching 3 days per week. Lecturers’ salary are not part of the department’s operating budget. The head of the department must request additional section funds or soft money from the College of Liberal Arts and Sciences every semester to cover courses that lecturers teach. GTAs are paid with 1-year stipends.

REFERENCES


**APPENDIX**

First Implementation (spring semester)

Research Question: Is there a significant difference in student outcomes by group: a redesigned blended-learning course and the conventional course?

Intervening Variables: contact time, online versus hard copy, in-class pedagogy

<table>
<thead>
<tr>
<th>Groups</th>
<th>Number of sections</th>
<th>Instructors</th>
<th>Number of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional</td>
<td>5</td>
<td>3 GTAs/2 lecturers</td>
<td>209</td>
</tr>
<tr>
<td>Experimental</td>
<td>5</td>
<td>teach both groups</td>
<td></td>
</tr>
</tbody>
</table>

Findings: significant differences on midterm exam favoring conventional group and significant differences on SOPI favoring blended-learning group; no significant differences by group for proficiency and diagnostic tests.
Follow-up Study (following fall semester)

Research Question: Is there a significant difference in student outcomes by group: a 2-day per week redesigned course taught by GTAs and a 3-day per week course taught by lecturers?

Intervening variables: contact time, instructor’s professional background

<table>
<thead>
<tr>
<th>Groups</th>
<th>Number of sections</th>
<th>Instructors</th>
<th>Number of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional</td>
<td>15</td>
<td>lecturers</td>
<td>656</td>
</tr>
<tr>
<td>Experimental</td>
<td>19</td>
<td>GTAs</td>
<td></td>
</tr>
</tbody>
</table>

Results: no significant difference on midterm; significant difference on final exam favoring 3-day course.

AUTHOR’S BIODATA

Dolly J. Young received her Ph.D. in Foreign Language Education at the University of Texas and is currently Professor at the University of Tennessee where she coordinates the lower division Spanish program. She has published numerous quantitative and qualitative studies on language anxiety, foreign-language/second-language reading, and foreign-language textbook analysis. She has edited two volumes on language anxiety and published supplementary readers in Spanish, French, and German. She is coauthor of ¿Qué te parece? a second-year textbook and is currently writing a first-year textbook.

AUTHOR’S ADDRESS

Dolly Jesusita Young
The Department of Modern Foreign Language & Literatures
701 McClung Tower
University of Tennessee
Knoxville, TN 37996-0470
Phone: 865 974 2312
Fax: 865 974 7096
Email: djyoung@utk.edu