Delivery Systems and Student Achievement: A Case for Discrete, Immediate, Digital Delivery of Listening Comprehension Exercises

J. Scott Despain
North Carolina State University

Background

Each year foreign language departments and lab administrators are faced with the task of upgrading or replacing antiquated language laboratory facilities. Difficult questions are asked and answered: Do we upgrade, or replace the audiocassette student stations? Do we install digital media servers, replacing the entire analog system with a lab full of networked multimedia computers? Do we use some combination of the two? Do we create a virtual laboratory? (See Yang 2000.) Are there other options?

Authors and publishers of textbooks and ancillaries are faced with similar questions: Should we continue to develop traditional, paper-based textbooks, workbooks, and lab manuals with audiocassettes? Should we move towards the digitization of these learning resources? If we choose digitization, should audiocassettes simply be digitized as 45-60 minutes audio files, delivered using the World Wide Web or a LAN? Or should authoring systems be utilized to develop interactive exercises? For administrators, teachers and authors alike, the question then follows: “How will students and instructors react to these non-traditional language learning resources?” What are the advantages and disadvantages to implementing these technologies into the learning process?

Approximately 12 years ago I took a graduate phonetics course which included a listening laboratory component. Because I had acquired my principal L2 during a 16-month stay in Ecuador, this was my first experience as a language learner with
this type of instructional technology. To complete the lab component I took my lab manual to the laboratory, obtained the cassette tape that corresponded to the chapter I was working on and then completed my activities. The tape deck was full-featured, and including fast-forward, rewind, book marking of specific tape segments, and the ability to record my voice on a separate track. This linear, analog delivery system has been the mainstay of language labs for many decades, and continues to this day.

Not finding the traditional technology particularly endearing, my first attempt to improve on the standard approach was to create a set of HyperCard exercises and deliver the audio source via a Mac SE30 interfacing with a Tandberg audio cassette tape deck. As those who attempted the same will no doubt remember, one of the principal problems with that delivery system was the media—the HyperCard exercises would turn out to be tape-player and audio-cassette specific; the tape media would stretch and change over time, and accessing different segments on the audio cassette was slow. The interface cards were also rather expensive. My next attempt at improving on the traditional technology came in 1991, again with HyperCard exercises, but this time, using digitized audio. Hard drives were becoming larger and cheaper and the audio compression software had progressed to the point that digitized audio exercises were now possible. The exercises I produced were limited in scope, and limited to the Macintosh computer platform, but they proved much more successful than my first project.

In 1995 I began work on a project with Jennifer Despain to develop a full set of exercises to accompany the 1st edition of ¡Arriba!: comunicación y cultura, from Prentice Hall. We chose Macromedia's Authorware as the authoring tool because it would allow for the very complicated instructional requirements of the exercises that we would be digitally duplicating. Additionally, Authorware would allow us to develop the project on the Macintosh and then port it to Windows (and vice versa, today). We completed that project after a year and a half. Beginning in 1997 we started work on the second edition of the paper-based exercises as well as the LAN version of the digital exercises. Due to significant requests from students we produced a CD-ROM version of the software of the second edition in 1998. These digital exercises are discrete, meaning each individual item (not activity nor chapter) is an individual module with its accompanying, immediately accessible audio segment; the audio is a seamless component of the exercises, and no additional
media player or other software is necessary.

Recently, our Spanish section adopted Heinle & Heinle’s *Plazas* as the textbook package for our beginning Spanish language courses. Heinle & Heinle is piloting a new electronic lab manual format. They have outsourced the on-line lab manual to http://www.quia.com/ in order to deliver the lab manual exercises via the Internet. A paper lab manual continues to be available with audio CD’s or audiocassette sets. The exercises are delivered via the web browser and the audio component for each activity (not each individual item) is streamed to the user using a media player. Likewise, Prentice Hall has taken essentially the same approach with their third edition of *Arriba*. As I evaluated this latest trend in listening comprehension exercises, I wondered if it is not in fact a step backward. The exercises and the audio are separate entities, similar to the cassette/lab manual approach; the audio source is streamed from Quia’s servers and is not immediately accessible, especially on dial-up connections. The technology seems to once again be in the way of learning. This article reports on a research project I conducted which, although it does not use the web sites mentioned, does compare the analog, and discrete digital approaches, and applies directly to the central question of how to best delivery these listening comprehension exercises for more efficient learning.

Hammerly’s advice of 15 years ago continues to be timely and applicable, notwithstanding the changes we have experienced over the years in computer and instructional technology:

> We should not do with computers what we did with language laboratories—use them mindlessly.... If we do not allow technology to determine our methodology but, instead, control it so it serves a pedagogically sound philosophy of language teaching then, and only, then will technology play a clearly useful and constructive role.... (1987, 8-9)

In general, research related to computer-based instruction (CBI) versus traditional instruction (TI) suggests that 1) CBI has a small but significant positive effect on achievement; 2) CBI substantially reduces the amount of instruction time by up to 1/3 of that required by TI; 3) CBI students have more positive attitudes toward instruction; 4) CBI students have a less positive attitude toward the subject; and 5) there is a higher attrition rate for CBI (Kulik and Kulik 1986).1 (See Kulik, Kulik and Cohen 1980 and Kulik and Kulik 1986 for several meta-analytic
Since the early 1960's other researchers have questioned the fruitfulness of studies contrasting different instructional mediums. Most notably, Richard Clark (1983, 1985a, 1985b, and 1991) claims that:

Consistent evidence is found for the generalization that there are no learning benefits to be gained from employing any specific medium to deliver instruction. Research showing performance or timesaving gains from one or another medium are shown to be vulnerable to compelling rival hypotheses concerning the uncontrolled effects of instructional method and novelty. (Clark 1983, 445)

Clark points to Mielke (1968) who suggested the same thing nearly sixteen years earlier when criticizing educational television versus live instruction. Mielke points out that in a rigidly controlled experiment, "mediation" would account for any variance only if all other aspects of the treatments, including the subject matter content and method of instruction, are identical (Mielke, 1968). According to Clark (1985, 1991), when instructional content is controlled in the research design, the positive effect for media more or less disappears. He suggests that the 30-50 percent times savings reported in some studies is plausibly due to a greater effort being put into the newer media (Clark 1983, 449).

In reference to these comparison studies, Clark (1991) does make positive reference to the idea that delivery technology is necessary to provide efficient and timely access to those methods and environments; that there are differences in the ability of one particular delivery technology over another. "Of course there are instructional problems other than learning that may be influenced by media (e.g., costs, distribution, the adequacy of delivery vehicles to carry different symbol systems, equity of access to instruction)" (Clark 1983, 454).

In recent years, as hardware prices have continued to decline, and local area network (LAN) technology has continually improved, computers and computer clusters have become commonplace and vital to the university campus, which includes foreign language departments and their language laboratories. High-speed access to the Internet and the World Wide Web both on and off campus is now become the expected norm. Wireless networks and notebook laboratories are also coming into
existence (Rowekamp 2000). This growth phenomenon is in many ways a duplication of the previously mentioned growth history of language laboratory facilities in the 1960s.

From a research study standpoint, Clark's suggestion, that as long as content and methodology are maintained, there will be little difference in learning achievement, would hold true in early studies comparing a paper workbook or an electronic workbook. However, instructional technology continues to experience dramatic evolutions that have obviously changed the education process in certain disciplines, and will create even more change in the future. (See Gates 1995 for one of many opinions.) One of the original foci of early computer applications was the area of foreign languages, and so it is not surprising to find a long list of technology-assisted language instruction projects. (See Morrison and Adams 1968; Holmes 1980; Colett 1982; Alemen-Centeno 1983; Wagers 1984; Kramsch, Morgenstern and Murray 1985; Jones 1989; Verano 1989; Madsen 1991; Nagata 1992; Despain 1993; Hernán 1994; and Monaghan 1995; Verano 2000; for representative projects and studies.)

There have been technology-assisted language instruction studies designed to demonstrate differences between traditional instruction and CAI. Niwa and Aoi (1990) reported on two studies, one of which appears to have controlled for content and methodology and which shows a 5% increase on a post-test score for the CAI group compared to the non-CAI group. Avent (1994) conducted a study that showed that the difference between means for grammar and vocabulary achievement for the computer-related group were significantly higher than for the traditional group. However, results from this study must be considered with caution for several reasons: 1) the study used volunteers; and 2) the lessons for the computer group were completely different from those used by the traditional group.

At least one study has been conducted that relates directly to the current study regarding student preference of delivery technology. Despain (1993) provides data (see Table 1, items 1-4) from the attitude surveys of a study involving beginning Spanish language students who completed the majority of the listening comprehension exercises using traditional delivery, as well as one chapter's exercises via computer delivery through a Local Area Network (LAN). Students who experienced both delivery technologies strongly preferred the exercises to be delivered through a LAN. They also believed that their time was
better spent when using the computerized exercises, felt that they were more inclined to review with the computerized exercises, and that they learned more quickly. These preliminary studies need to be expanded from one chapter's worth of exercises for a course to the curriculum for an entire semester, in order to determine if the novelty effect mentioned by Clark is influencing these results.

Table 1. Listening Comprehension Survey Results

<table>
<thead>
<tr>
<th>Question</th>
<th>N</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I prefer computerized exercises</td>
<td>44</td>
<td>2.25</td>
</tr>
<tr>
<td>2. The time spent on computerized exercises more efficient</td>
<td>43</td>
<td>2.49</td>
</tr>
<tr>
<td>3. I was more inclined to review w/ computerized exercises</td>
<td>44</td>
<td>2.55</td>
</tr>
<tr>
<td>4. I learned more quickly with computerized exercises</td>
<td>44</td>
<td>2.64</td>
</tr>
<tr>
<td>5. Language lab exercises help me with work in class</td>
<td>166</td>
<td>2.87</td>
</tr>
<tr>
<td>6. Time spent on language lab exercises is time well spent</td>
<td>168</td>
<td>3.14</td>
</tr>
<tr>
<td>7. I learn a lot from the exercises</td>
<td>168</td>
<td>3.27</td>
</tr>
<tr>
<td>8. Better success in course because of lab exercises</td>
<td>168</td>
<td>3.42</td>
</tr>
<tr>
<td>9. Language lab manual worth what I paid for it</td>
<td>168</td>
<td>3.65</td>
</tr>
<tr>
<td>10. I enjoy doing language exercises</td>
<td>167</td>
<td>3.83</td>
</tr>
<tr>
<td>11. Even if not collected, I would do them</td>
<td>167</td>
<td>3.86</td>
</tr>
<tr>
<td>12. I look forward to doing language lab exercises</td>
<td>166</td>
<td>4.11</td>
</tr>
</tbody>
</table>

Note: Items have been ordered starting with questions earning the highest positive attitudinal score to those with the highest negative attitudinal score. The mean is based on a scale of 1 (Strongly Agree) to 5 (Strongly Disagree)

Carroll's Model of School Learning, published in 1963, provided the genesis for several mathematical representations of learning. "It should be understood that 'spending time' means actually spending time on the act of learning. 'Time' is therefore not 'elapsed time' but the time during which the person is oriented to the learning task and actively engaged in learning" (Carroll 1963, 725). Critical studies such as those conducted by Wiley and Harnischfeger built on Carroll's work and helped to refute a predominant belief at the time which suggested that schooling has minimal effects on learning (Wiley and Harnischfeger 1974; Harnischfeger and Wiley 1976). They even went so far as to suggest that quantity of time in school could be reduced if the quality were improved. Based on “exposure-to-instruction”
data, they "concluded that in schools where students receive 24 percent more schooling, they will increase their average gain in reading comprehension by two-thirds and their gains in mathematics and verbal skills by more than one-third (Wiley and Harrischfeger 1974, 9).

Relating Academic Learning Time (ALT) to computer-assisted instruction (CAI), Vockell recommends using ALT "as the major factor in deciding whether and how to introduce the computer into the curriculum at any grade level or in any subject area" (1987, 72).

Variation in seat-time at the primary/secondary level and time spent on non-academic activities are similar to the variation possible in the foreign language lab at all levels using the traditional delivery technology. On the other hand, because discrete digital delivery technology allows for student exploration and self-directed activities, it not only has the potential to increase ALT, but also to resolve many issues of "seat-time" variation through its ability to measure elapsed time, and track student performance and the actual learning process followed by any given student.

This study attempted to examine several questions related to how, when, and where technology is integrated into foreign language instruction. Patrikis' question, "Where is computer technology taking us?" (1995, 36) would suggest a passive approach to a volatile, pervasive, and at times, almost overwhelming aspect of education and modern society. Working towards answering the active and encompassing question: "Should foreign language professionals put forth the necessary time and expense to provide digital (vs. analog cassette) delivery of the listening comprehension exercises for university-level beginning language courses?" this study addressed two essential questions: 1) Will students who use computer-delivered listening comprehension exercises learn the language more effectively than students using a cassette tape and lab manual? 2) Will these students learn the language more efficiently?

This study focused exclusively on listening comprehension materials and the language laboratory. The main purpose was to gather more specific data as to how students use the two separate delivery systems and what effect they have on student achievement. Its purpose was not to propose new teaching methodologies based on new technologies; rather it is to study the "cognitive effects with" computers, "constituting improved
performance while an intellectual tool is available”, versus the “cognitive effects of computer tools, meaning the subsequent cognitive residue as result” (Salomon 1990, 521). The strengths and weaknesses of the two delivery systems that were utilized in this study are summarized in Table 2.

Table 2. Comparison of Delivery System Characteristics

<table>
<thead>
<tr>
<th>Item</th>
<th>Traditional</th>
<th>Discrete Digital</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reinforcement</strong></td>
<td>Slow Linear Partially controlled by the learner, but timing is mostly controlled by periods of silence on the audio tape.</td>
<td>Immediate Randomly accessible Controlled almost exclusively by the learner depending on each individual student’s learning strategies.</td>
</tr>
<tr>
<td><strong>RecordKeeping</strong></td>
<td>Time and date, if a lab is so equipped. Lab manual turned in by students.</td>
<td>Total session time, date, % correct of judgeable answers, exact order in which students complete each activity, number of minutes/activity, number of repetitions/activity, etc.</td>
</tr>
<tr>
<td><strong>Setup</strong></td>
<td>Slow Access is limited to exercises for one-half of a chapter at any time.</td>
<td>Almost instantaneous setup and access to any chapter.</td>
</tr>
<tr>
<td><strong>Feasibility to “chunk” the learning</strong></td>
<td>Very problematic</td>
<td>Each activity for any chapter is immediately accessible at any time during any session.</td>
</tr>
<tr>
<td><strong>Number of sessions per chapter of exercises</strong></td>
<td>Realistically, a maximum of two.</td>
<td>Minimal time is wasted in accessing the instruction. Therefore, multiple shorter sessions are practical and beneficial.</td>
</tr>
</tbody>
</table>

In essence, the current medium of instruction used to deliver language lab exercises is significantly limited in its instructional attributes. As Heterick states, “the plethora of digital
technologies offers the opportunity to break the industrial age model of teaching and learning and offer a customized service directly to the learner" (1993, 4). Computerized interactive exercises, such as those utilized in this study, allow for instantaneous review, immediate feedback, immediate presentation of any audio segment, and individual learner-guided and/or computer-assisted instruction deliverable to multiple learning sites (libraries, residence halls, computer clusters) simultaneously. By using a computer delivery system, the technology no longer constrains the student into completing, for example, activity one and then activity two by allowing only point A, to point B to point C access. With discrete digital technology the student can access any section of any activity of any chapter at any time, or follow the order of learning proposed by the textbook author.

As discussed previously, research in many disciplines shows that as successful time-on-task increases, learning increases. Research also shows that the more concise and succinct the instructional unit, the more learning that takes place. Students typically confine their completion of the language lab exercises to one session of approximately one hour per chapter. Using the digitized format students can go straight to the instruction without having to passively wait for English directions and publisher-determined pauses. Students can immediately access and/or review entire sections of activities or chapters and jump between chapters according to their individual needs and learning strategies. Students are in control of the reinforcement of their responses. The present study attempted to determine if students would take advantage of these media attributes.

Methodology

An experiment was designed in which participants were randomly assigned to one of two treatment groups: Group A received normal classroom instruction plus they completed learning exercises in a standard language-lab setting with cassette audio listening stations. Group B received normal classroom instruction but completed their learning exercises using computer-based exercises that were a virtual duplicate of the same content. The experiment was designed to test the following research hypotheses:

1. Group B will perform significantly higher in learning achievement, compared to Group A, as measured by a test of oral listening comprehension.
2. Within each group there will be a significant, positive correlation between the amount of successful practice
with the exercises and achievement.
3. Within each group there will be a significant, positive correlation between the number of repeated exercises and achievement.

Participants

The participants for this study were 80 students enrolled in five sections of a first-year, first-semester, college-level Spanish course at Institution X during a five-week summer session. The three instructors assigned to teach the five sections were either permanent faculty members or visiting lecturers, each having taught the course previously; the researcher was not one of the instructors. The participants ranged from first-year through graduate student status (see Table 3), and represented 32 distinct majors. Previous Spanish language experience ranged from zero to six years with 79% having two years or less of previous Spanish experience (see Table 4).

Table 3. Year in School Demographics

<table>
<thead>
<tr>
<th>Year</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshman</td>
<td>6</td>
<td>7.50</td>
<td>7.50</td>
</tr>
<tr>
<td>Sophomore</td>
<td>21</td>
<td>26.25</td>
<td>33.75</td>
</tr>
<tr>
<td>Junior</td>
<td>13</td>
<td>16.25</td>
<td>50.00</td>
</tr>
<tr>
<td>Senior</td>
<td>17</td>
<td>21.25</td>
<td>71.25</td>
</tr>
<tr>
<td>Other *</td>
<td>23</td>
<td>28.75</td>
<td>100.00</td>
</tr>
<tr>
<td>Totals</td>
<td>80</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>

*Life-long students, graduate students, etc.

Table 4. Previous Spanish Language Experience

<table>
<thead>
<tr>
<th>#of years</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>30</td>
<td>37.50</td>
<td>37.50</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>13.75</td>
<td>51.25</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>27.50</td>
<td>78.75</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>11.25</td>
<td>90.00</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1.25</td>
<td>91.25</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1.25</td>
<td>92.50</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1.25</td>
<td>93.75</td>
</tr>
<tr>
<td>Other *</td>
<td>2</td>
<td>2.50</td>
<td>96.25</td>
</tr>
<tr>
<td>Missing</td>
<td>3</td>
<td>3.75</td>
<td>100.00</td>
</tr>
<tr>
<td>Totals</td>
<td>80</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>

*Study or experience abroad, lived with Hispanic family, etc.
Instruction

The course was a typical beginning Spanish language course intended to provide opportunities for students to develop the five language skills: listening, speaking, reading, writing, and cultural understanding. Students were assigned material to read and activities to prepare outside of class, including exercises in the workbook and the lab manual. Students then attended class to practice their skills in small groups, to participate in choral work with the instructor, listen to brief lectures by the instructor, participate in interview situations, in-class writing activities, etc.

Listening Comprehension Materials

1. Group A (Traditional) - ¡Arriba!: Comunicación y cultura, Lab Manual. The listening comprehension exercises were delivered via the language lab manual and cassette tape that accompany ¡Arriba!, Comunicación y cultura. Cassette tapes were made available in the language lab and were played at Tandberg student stations. Students were not allowed to dub the tapes for use in completing the exercises somewhere other than in the language lab. Students wrote their answers in the lab manual. The lab manual gave part of the answers at the end of the lab manual, while the audiocassette provided the remainder.

2. Group B (Computer) - ¡Arriba Audio! 1.1 The program ¡Arriba Audio! 1.1 was used to deliver the listening comprehension exercises through a Local Area Network (LAN). These exercises are a digital duplication of the exercises in the lab manual, but were not simply an electronic audio file player. They were developed by the researcher and an asset programmer using Authorware Professional by Macromedia. The master audio cassettes were digitized with an 8 bit, 11.127 kHz sampling rate, using a cassette deck with Dolby C noise reduction, the MacRecorder from Farallon, and Sound Edit 16 software from Macromedia. The digitized audio was then separated into individual item segments, or even single words as needed. All graphics that could not be replicated using the authoring software were scanned.

The greatest concern was placed on ensuring complete content and pedagogical consistency in the development of the computer version in order to eliminate as many of the confounding variables as possible that have been cited in most CAI-versus-traditional-method research studies. Therefore, when using ¡Arriba Audio! what a student sees on the computer screen and hears is exactly what a student sees on the lab manual page and hears on the tape, activity by activity. The principal difference between the two delivery systems is that ¡Arriba Audio! provides
the means by which a student can immediately access any item, for any activity, from any section of any chapter. Additionally, all dead time has been eliminated, meaning the student does not need to wait a specific amount of time for a reinforcement response.

**Measurement Instruments**

The researcher created a series of five achievement tests to be used in determining the listening comprehension proficiency of the participants. Each test required between 10 and 20 minutes to complete. All items were variations of the listening comprehension exercises that the participants completed as part of the study. Item types included multiple choice and true/false. Each test was computer-administered, and used segments from the listening comprehension exercises as the sole audio source. The participants normally took the practice test for each chapter the day after completing the respective exercises. The practice test for Chapter Five was also used as a pretest to baseline each group and verify that the two groups were not significantly different at the beginning of the study.

**Scale Formation**

The reliability of each of the practice tests was measured, and corresponding scales to be used in testing the achievement hypotheses, were formed using a principal component analysis limited to two factors. Data for these factor analyses came exclusively from the study, except for the pretest/chapter five test which included data from a pilot study conducted using Spanish language students registered in courses at Institution X the semester prior to the study. Items with zero variance and all negative loading items were removed, plus additional items with a loading near zero, until a maximum Cronbach alpha coefficient was approached. The alpha for each of the tests was as follows: Practice test one, .7836; practice test two, .7926; practice test three, .7439; practice test 4, .7761; and pretest/practice test 5, .8428. Full reliability results for all scales used in the study are presented in Table 5.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>N of Cases</th>
<th>N of Items</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest/</td>
<td>221</td>
<td>37</td>
<td>.8428</td>
</tr>
<tr>
<td>Post-test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practice Test 1</td>
<td>85</td>
<td>31</td>
<td>.7836</td>
</tr>
<tr>
<td>Practice Test 2</td>
<td>89</td>
<td>42</td>
<td>.7926</td>
</tr>
<tr>
<td>Practice Test 3</td>
<td>69</td>
<td>40</td>
<td>.7439</td>
</tr>
<tr>
<td>Practice Test 4</td>
<td>63</td>
<td>34</td>
<td>.7761</td>
</tr>
</tbody>
</table>

Table 5. Instrument Reliability: Statistics and Cronbach Alpha Coefficients
Procedures

Several weeks prior to the beginning of the study three experienced Spanish instructors were contacted and then committed to participate in the study. The instructors were provided with a written description of how the study should proceed. During the first day of classes the participants were given an introduction to the course. They were also asked to fill out a form regarding their previous language experience and their placement score results. They were then given a brief oral description of the study, followed by a randomly distributed written information/consent form which included each subject’s group assignment within his or her own section.

The students, accompanied by the instructor, spent the remainder of the first class period in the language laboratory where they would be doing the majority of the course work related to the study. The software programs were described and all questions were answered regarding lab policy, etc. Members of the traditional group were also given a brief tour of the other lab facility where they would be completing the listening comprehension exercises using the lab manual/cassette. Participants returned over the next few days to complete the listening comprehension proficiency pretest prior to completing any of the listening comprehension exercises.

The participants used the following schedule to complete the remaining elements of the study: 1) Complete the first half of the exercises for Chapter One; 2) Wait a day or two, then complete the second half of the exercises for Chapter One; 3) Complete the chapter “practice exam” (chapter post-test) one day following completion of the exercises and one day prior to the chapter exam. They then completed steps 1, 2, and 3 for Chapters Two to Four. For Chapter Five, participants completed only the first half of the listening comprehension exercises (to coincide with the curriculum of the course). They then took the practice test for Chapter Five (the listening comprehension pretest)2.

Record Keeping

Data for the proficiency pretest, the chapter practice tests, as well as the information for the successful practice variable were automatically recorded into computer databases for the computer group8. For the traditional group, the pretest and practice test data were collected by the computer, whereas the remaining data was collected and entered by hand9.

Variables

The “oral listening comprehension/learning achievement” variable, used to test the three hypotheses was calculated by converting raw scores for each of the five practice tests to T-
scores, then averaging the top three T-scores for each subject. The "successful practice" variable, used in testing hypothesis two, was calculated by adding together the products of session time (reported in minutes) and self-reported success (reported as a percentage) for each individual session for each subject. The "number of repeated exercises" variable, used in testing hypothesis three, was calculated by summing the number of individual activities completed by each individual subject during each session.

**Results**

**Hypothesis 1:** Group B will perform significantly higher in learning achievement, compared to Group A, as measured by a test of oral listening comprehension. Means and standard deviations were computed on the listening comprehension pretest and the learning achievement score from the practice tests. A t-test for equality of means was then conducted to compare the means of the traditional and computer groups. Levene’s test for equality of variances was also conducted for each test. Results of the t-tests and variance tests are presented in Table 6.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>S.D</th>
<th>t</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pretest</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traditional Group</td>
<td>35</td>
<td>17.94</td>
<td>4.51</td>
<td>-.545</td>
<td>.015</td>
</tr>
<tr>
<td>Computer Group</td>
<td>32</td>
<td>18.53</td>
<td>4.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Learning Achievement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traditional Group</td>
<td>33</td>
<td>50.52</td>
<td>10.22</td>
<td>-1.399*</td>
<td>.695</td>
</tr>
<tr>
<td>Computer Group</td>
<td>34</td>
<td>53.56</td>
<td>7.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Learning Achievement (w/o outlier)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traditional Group</td>
<td>33</td>
<td>50.52</td>
<td>10.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer Group</td>
<td>33</td>
<td>53.07</td>
<td>6.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Successful Practice</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traditional Group</td>
<td>42</td>
<td>205.6</td>
<td>83.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer Group</td>
<td>38</td>
<td>166.8</td>
<td>131.57</td>
<td>1.557*</td>
<td>3.3381</td>
</tr>
<tr>
<td><strong>Successful Practice (w/o outlier)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traditional Group</td>
<td>42</td>
<td>205.6</td>
<td>83.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer Group</td>
<td>37</td>
<td>151.7</td>
<td>93.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Completed Activities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traditional Group</td>
<td>33</td>
<td>92.96</td>
<td>34.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer Group</td>
<td>34</td>
<td>91.26</td>
<td>40.50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant at the p< .10 level (1-tailed)
Means of the raw scores for the listening comprehension pretest were 17.94 for the traditional group, and 18.53 for the computer group. The difference between means of the two groups was not significant at the p < .05 levels, suggesting that the two groups were similar in initial listening comprehension ability. Means for learning achievement (averaged T-scores) were 50.52 for the traditional group, and 53.56 for the computer group. Levene's test for equality of variance was not significant for either of the t-tests, indicating adequate equality of variance within groups. The difference between the means of the two groups was not significant at the p < .05 level, but it did approach significance (p = .085, 1-tailed), with the computer group scoring a third of a standard deviation higher than the traditional group. However, because the significance level for the study was p < .05, the null hypothesis was not rejected for hypothesis one. (See hypothesis two testing for a further analysis.)

**Hypothesis 2:** Within Group A and Group B there will be a significant, positive correlation between the amount of successful practice with the exercises and achievement on the practice tests. Pearson product-moment correlations were computed and were then tested for significance within both groups to determine if there was a correlation between the amount of successful practice time participants spent completing the listening comprehension exercises and their learning achievement. There was a significant positive correlation between successful practice and achievement of .516 for the traditional group, and .597 for the computer group. Both correlations were significant at the p <.01 levels Therefore, the null hypothesis for hypothesis two of the study was rejected.

In order to compare the two groups, means and standard deviations were computed for successful practice time (see Table 6). Group B had a standard deviation 63% higher than Group A. A line graph (see Figure 1) comparing the two distributions together, and two histograms (see Figures 2 and 3) showing the distributions separately, were plotted to determine possible reasons for the substantial difference in variance between the two groups, and to determine their respective distributions. Based on the resulting graphs, Group A appeared to have a normal distribution. However, Group B had a positively skewed distribution, due solely to a single subject with a successful practice score 4.37 times higher than the mean for the group. This outlier was a non-traditional, female student enrolled in the Continuing Education program.
at institution X. She spent 23 sessions in the lab during the 25-day study!

Figure 1

Figure 2

Figure 3
An analysis of variance (ANOVA) was performed on the means for the two groups, both including and excluding the previously discussed outlier. Results of the ANOVA are presented in Table 7. When including the outlier, the difference between means (206 for Group A, and 167 minutes for Group B) was not significant at the a priori alpha level of \( p < .05 \), but the difference did approach significance (\( p = .058 \), one-tail). After excluding the outlier from the analysis, the difference between means (206 minutes for Group A, and 152 minutes for Group B) was highly significant at the \( p < .01 \) alpha level.

![Table 7. Successful Practice Time, ANOVA](image)

Based on the previously mentioned significant correlation between Learning Achievement and Successful Practice Time, as well as the ANOVA results for Successful Practice Time, an analysis of covariance was used to test for significant differences between groups for Learning Achievement (the dependent variable), taking into consideration Successful Practice Time (the covariate). This analysis was performed to verify the rejection of hypothesis one. The results of the ANCOVA are presented in Table 8. Scatter diagrams were created to plot the regression of Learning Achievement on Successful Practice for each group in order to ensure linearity of regression between groups (see Figures 4 and 5). (One outlier was removed from each group to more accurately reflect the regression line.) In addition to plotting the slopes for each group, the assumption of equal slopes was checked using the PROC GLM of SAS\textsuperscript{TM} release 6.11 and was found to be tenable (\( F = 3.09 \), ns at the \( p < .05 \) level).
Table 8. Achievement with Successful Practice, ANCOVA

<table>
<thead>
<tr>
<th>Covariates/Achievement/Successful Practice</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Effects</td>
<td>363.5</td>
<td>1</td>
<td>363.5</td>
<td>6.13</td>
<td>.008</td>
</tr>
<tr>
<td>Model</td>
<td>1511.8</td>
<td>2</td>
<td>755.9</td>
<td>12.75</td>
<td>.000</td>
</tr>
<tr>
<td>Residual</td>
<td>3794.6</td>
<td>64</td>
<td>59.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5306.4</td>
<td>66</td>
<td>80.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Based on the adjusted means for learning achievement of 49.65 for the traditional group and 54.39 for the computer group, and a significant F of 6.13 at the p < .01 level, participants in the computer group did have significantly higher learning achievement scores than did those in the traditional group.
Therefore, based on the findings of this additional analysis, the null hypothesis was ultimately rejected for hypothesis one.

**Hypothesis 3:** Within Group A and Group B there will be a significant, positive correlation between the number of repeated exercises and achievement on the practice tests. Pearson product-moment correlations were computed and were then tested for significance within each group to determine if there was a correlation between the number of completed/repeated exercises and subject learning achievement. For the traditional group, the resulting correlation of .280 was not significant at the $p < .05$ levels. However, for the computer group, there was a significant positive correlation of .560 at the $p < .01$ level between the number of completed/repeated exercises and learning. Therefore, the null hypothesis for hypothesis three of the study was rejected for Group B, the computer group, but not for Group A, the traditional group.

**Discussion**

The first hypothesis—*students that complete the listening comprehension exercises will do better on the practice listening comprehension exams*—was initially not validated by the study results, using the a priori $p < .05$ level of confidence for significance and t-tests. The differences between the computer group and the traditional group means for learning achievement only approached significance. However, after finding a significant positive correlation between successful practice time and learning achievement, and using an analysis of covariance to adjust the achievement scores using successful practice time, a significant difference was found between group means for learning achievement.

Conservatively, what *can* be concluded from the results is that when content and methodology remain the same, and an attempt is made to keep other variables controlled, the computer seems to be at least an equal delivery system, compared to the cassette tape/lab manual. When considered in conjunction with differences in successful practice time, students appear to learn more effectively using the computer-based listening comprehension exercises. One of the most reasonable explanations for this improvement would be that the computer-based system engages the learner to a greater degree than the traditional delivery system—the student using the computer must do something in order to access the instruction. The student using the lab manual and cassette passively sits at a workstation while the instruction flows by.
The second hypothesis—students that spend more successful practice time on the exercises will learn more—was validated by the study for both groups. These significant, positive correlations for both groups suggest that the participants that spent more time on the lessons, regardless of delivery medium, performed better on the practice tests. This is an important message for students to hear, that the engaged time that they spend in the lab will improve their performance.

To determine if one of the delivery systems was more efficient than the other, the difference between the means for the two groups for successful practice was tested and found to only approach significance, as mentioned previously. Therefore, at the least, the discrete-digital is equal to the traditional delivery system regarding efficiency. After plotting and analyzing the distributions of successful practice time for the two groups, an even stronger case is made for this conclusion; the successful practice time for one subject in the computer group created a positive skewness in the distribution, the score being 4.37 times the mean for the group as a whole. After excluding the outlier and recalculating the ANOVA the difference between groups was found to be significant. Therefore, the data suggest the possibility that the computer group participants in fact required a significantly less amount of time to achieve the same (or higher) level of learning.

Combining the findings of the testing of hypothesis one and two together, this researcher is inclined to say that the computer group participants learned more effectively and efficiently. However, that conclusion is drawn with some hesitation, principally because the “successful practice” variable included a self-reported “success rate” by the students. Participants in the computer group were provided with a percentage score of how well they had performed on all judgeable responses immediately prior to their self-reporting of how well they did for all activities during each session, whereas the traditional group received no such report. It is possible that the computer had an effect on how participants in the computer group perceived their success. When comparing means of the total time participants spent in the lab (265 minutes for Group A and 248 minutes for Group B) they were found to not be significant. However, removing the outlier produced means of 265 minutes for the traditional group and 228 minutes for the computer group. The resulting t-score of 1.312 did approach significance (p = .097). Therefore, taken as a whole, the data from hypotheses one and two would suggest that the computer is a more time-efficient
and effective delivery system.

The third hypothesis—students that complete/repeat more exercises will do better on the practice tests—was validated by the study, but only for participants in the computer group. This finding also appears to support the idea that the computer may be more efficient for the listening comprehension exercises task, because means for the two groups were found to not be significantly different. Although the data cannot be quantified at this time, after compiling all of the data for the study, this researcher has the sense that the vast majority of students in the traditional group simply completed each activity once, if that, and did not make an attempt to modify the instructional flow as originally designed by the exercise authors. Very few participants repeated exercises or skipped exercises. However, those in the computer group tended to skip some activities and repeat others, appearing to modify the instructional experience.

Subject/Data Attrition: This study was conducted during a five-week summer school session. The students had to complete many questionnaires and practice tests during those five weeks. Several questionnaires and proficiency tests had to be completed during one or two days at the beginning and the end of the study in order to have valid pre- and post-test measurements.

The number of participating students for the pretest measures was better than for the post-test measures. From beginning to end, the number of participants for any one instrument in the study tended to decrease. This attrition is partly due to natural attrition in the course (approximately 10%), but also some decrease (between 10% and 20%) can probably be attributed to participants simply growing tired of the process. Additionally, the instructors that taught the course required and rewarded their students' completion of the various aspects of the study to varying degrees. Therefore, because not all participants in the study completed each practice test or questionnaire (see Table 5), the power of the data analyses was decreased.

Testing: Both groups took the achievement tests on the computer, which might have given the computer group an advantage in testing.

Participants: Students in the study were somewhat atypical of the traditional beginning language student: they were taking a summer language course, and a full two-thirds were in their third year or more of school, which may be linked to the fact that
proficiency in a foreign language at the second semester level is a general education requirement of the institution (i.e. they had to be there).

Other Variables: This study did not take into account differences regarding gender, major curriculum, age, year in school, nor the possible masking effects of the classroom experience of even more significant differences between the groups. Future projects should also allow for increased flexibility regarding where (and when) students complete the exercises.

Conclusion

This study attempted to answer several fundamental questions related to the effects on learning when delivering listening comprehension exercises using a digital, discrete-item approach. The study suggested that students learn more effectively and efficiently when using this approach. The study showed that students who complete more exercises learn more, at least for the computer group participants. The study results indicated that there is a time advantage to using this delivery type, but this needs to be studied further. Although nearly all of the hypotheses were validated by the results of the study, the researcher is not completely confident of some of the variables used in the study, especially "engaged time". Therefore, the results should be treated with some caution.

The research project demonstrated that there was no significant difference when comparing students who passively listening to a tape and those who completed the exercises using the digital version. This would not be far removed from the current approach taken by Prentice Hall and Heinle & Heinle in their e-lab manual projects. The problem is in the lag time between click and listen. The other problem is the dependency on an Internet connection.

The Plazas and ¡Arriba! approaches do provide relatively seamless digital access to individual exercises, making it a step beyond the audio cassette/lab manual technology. However, there continues to be a lag time with this approach to delivery. The findings of this study would suggest that this most recent trend, that of delivering the exercises by activity, and not individual item, and streaming the content using distant servers may very well be a step backward towards the lab manual/cassette approach. The findings suggest that the academic learning time savings achieved through the discrete digital delivery approach would be eliminated if the Internet delivery in its current state continues.
Content authors should feel comfortable moving away from the lab manual/cassette tape and toward interactive discrete digital exercises. Also, the dynamic attribute of a digital medium allows for the editing, and enhancement of exercises in a much more timely manner. Administrators should feel comfortable replacing traditional language labs with a computer lab, as long as the necessary software for the computer lab is available to meet their needs. Publishers should consider moving away from the traditional lab manual/cassette and/or streaming real-audio files or simply providing Audio CDs, and instead invest in interactive discrete digital exercises.

Suggestions for Future Research

This study should be replicated on several different campuses, during the regular academic year, with a broader sample of the population. More effort should be made to ensure that all study participants complete each aspect of the study. Formal studies should be conducted that allow students within sections of a course to sample both delivery systems, completing a chapter's worth of exercises on each, and then be given the option to choose which system they would like to use to complete the remaining exercises. Studies could also be conducted to determine possible differences between the two groups based on gender, year in school, school major, and age. Additionally, studies should be conducted that have the same content but that take more advantage of the media attributes of the multimedia computer. It would appear to be fruitful to compare academic learning time (or at least seat time) between Internet delivery and CD delivery of the same listening comprehension exercises.

Notes

1. Effects on attitude were also investigated as part of the current study but are not included herein.


3. Although other students participated in different aspects of the study, typical course-enrollment attrition of approximately 10% occurred. Therefore, only data from the
students enrolled at the conclusion of the course were included in the study. Also excluded from the study were 5 students that unintentionally, or intentionally, self-selected into one of the two study groups.

4. Some images were reduced or enlarged to better fit in the viewing area of the computer monitor.

5. Some of the English instructions that provide background for the activities were minimally altered in order to reduce confusion as to what the student should do in the digitized format in order to check an answer, view an image, etc. (i.e. “Click on the correct answer” instead of “Circle the correct answer.”)

6. The answers, or reinforcement, are provided following the lab manual/cassette’s design and schedule: If the answer is in the back of the lab manual, an answer button is provided on the screen for the student to click on in order to access the answers to a particular activity; if the answer is provided orally on the cassette, ¡Arriba Audio! 1.1 provides oral reinforcement, again accessed when the student clicks on various buttons on the screen.

7. It will be apparent from the data analysis in chapter 4 that not all participants completed every single component of the study, nor was the suggested calendar followed exactly.

8. For the computer group, the following scores were recorded by the software and were automatically written to a data file: Session date, session time, lesson number, and the number of repetitions of each activity. After the participants completed a session with ¡Arriba Audio! they were provided with a “percent correct” score of the work completed. The participants were then asked to provide an estimated percentage (either “100%,” “90%,” “80%,” “70%,” “60%,” or “50% or below”) of: 1) the amount of audio listened to; 2) answers consulted via audio and/or answer buttons; and 3) items answered correctly.

9. For the traditional group, the following scores were self-reported on forms provided to them by lab assistants: Session date, session time in minutes (often recorded by the lab assistant) and the number of times each activity was repeated. Next, given the following choices: “100%,” “90%,” “80%,” “70%,” “60%,” or “50% or below,” participants were asked to...
provide an estimated percentage of: 1) the number of answers they consulted via the tape or the back of the lab manual; 2) the amount of audio they listened to; and 3) the number of items they answered correctly.

10. Raw scores were converted to T-scores because the composite score to be calculated was to be an average of five practice tests that were not equal in difficulty, and only the top three scores were to be included. Only the top three scores were used because not all participants took all of the tests; the vast majority did take at least three of the tests.

References


Avent, J.H. 1994. A study of language learning achievement differences between students using the traditional language laboratory and students using computer-assisted language learning courseware. Diss., Univ. of Georgia.


——. 1991. When researchers swim upstream: reflections on


Monaghan, P. 1995. Logging on to languages: multimedia programs help Brigham young students learn foreign languages. *Chronicle of Higher Education* 42. 8 (October 20,


J. Scott Despain is Assistant Professor of Spanish at North Carolina State University in Raleigh.