DOCUMENT RESUME

ED 427 515 FL 025 687

AUTHOR Carpenter, Mark, Ed.; Koike, Dale April, Ed.
TITLE Linguistic Perspectives from Spanish Second Language Acquisition.
INSTITUTION Texas Univ., Austin. Foreign Language Education Program.
ISSN ISSN-0898-8471
PUB DATE 1998-00-00
NOTE 177p.; For individual articles, see FL 025 688-696.
PUB TYPE Collected Works - Serials (022)
JOURNAL CIT Texas Papers in Foreign Language Education; v3 n3 Fall 1998
EDRS PRICE MF01/PC08 Plus Postage.
DESCRIPTORS Adult Students; Auditory Perception; Bilingual Education; Bilingual Teachers; Determiners (Languages); Discourse Analysis; *Grammar; Language Patterns; *Language Proficiency; Language Research; Language Tests; Language Usage; Multimedia Instruction; Oral Language; *Phonology; Second Language Learning; *Second Languages; *Spanish; Student Evaluation; Teacher Qualifications; Testing; *Time Factors (Learning); Verbs

ABSTRACT

Papers reporting research on Spanish second language learning include: "Discourse Features of Spanish Oral Production at the Novice Level" (Rebecca Jo Bearden); "A Discourse Approach to the Assessment of Foreign Language Oral Proficiency" (Dale April Koike, Fanny Hinojosa); "Acquisition of Spanish Definite Articles by English-Speaking Learners of Spanish" (Maria Ramirez-Mayberry); "The Spanish Psych Verb Construction: Beginning and Intermediate Learners' Patterns of Usage" (Christopher D. Gascon); "The Relationship between the Production and Perception of L2 Spanish Stops" (Mary Zampini); "An Acoustic Description of the Longitudinal Acquisition of Spanish Phonological Features by English Speaking Adult Learners" (Jeffrey Reeder); "Prochievement Testing: Matching Instructor Expectations, Student Level, and Task Levels" (Barbara Gonzalez Pino); "Current Issues in the Spanish Language Proficiency of Bilingual Education Teachers" (Michael Guerrero); and "Multiple Challenges of Multimedia: Development, Implementation, and Evaluation in Second Semester Spanish" (Margaret Ann Kassen). (MSE)

ISSN # 0898-8471

Texas Papers in Foreign Language Education

SPECIAL ISSUE: LINGUISTIC PERSPECTIVES FROM SPANISH SECOND LANGUAGE ACQUISITION

Volume 3 **Number 3** **Fall 1998**

Foreign Language Education Program
The University of Texas at Austin

Texas Papers in Foreign Language Education
(TPFLE)

TPFLE is the official journal of the Foreign Language Education Program of the University of Texas at Austin.

Publication of TPFLE has been made possible by a grant from the RGK Corporation

Foreign Language Education Program
The University of Texas at Austin

Texas Papers in
Foreign Language Education

Contemporary Issues in Foreign and Second Language Learning and Teaching

SPECIAL ISSUE: LINGUISTIC PERSPECTIVES FROM SPANISH SECOND LANGUAGE ACQUISITION

Volume 3 **Number 3** **Fall 1998**

Foreign Language Education Program
The University of Texas at Austin

Contributors:

REBECCA JO BEARDEN earned her M.A. in Hispanic Linguistics with a focus on Spanish second language acquisition from the University of Texas at Austin in 1997. She is currently working on her Ph.D., also at University of Texas, in the same areas.

CHRISTOPHER D. GASCÓN is a doctoral candidate in Spanish literature at the University of Texas at Austin. He is currently writing his dissertation on the role of desire in the Golden Age drama of Spain.

BARBARA GONZÁLEZ PINO, Associate Professor at the University of Texas at San Antonio, is Coordinator of Lower Division Spanish and Foreign Language Education. She has focused extensively on classroom and proficiency testing of speaking in her publications, presentations, and training activities.

MICHAEL GUERRERO is an Assistant Professor at the University of Texas at Austin in the Department of Curriculum & Instruction. His research interests center on the preparation of bilingual education teachers and related language development and testing issues.

FANNY HINOJOSA recently received her undergraduate degree at the Universidad Autónoma de Nuevo León in Monterrey, México. She is currently an M.A. exchange student at McGill University in Montreal, where she is studing language pedagogy.

MARGARET ANN KASSEN received her Ph.D. from the University of Texas at Austin and is now Associate Professor in the Department of Modern Languages at The Catholic University of America. With interests in technology-enhanced language teaching and foreign language writing, she coordinates the language program, works with graduate and undergraduate teacher preparation, and teaches language courses.

DALE APRIL KOIKE is Associate Professor of Spanish and Portuguese Linguistics at the University of Texas at Austin. Her research interests include pragmatics, discourse analysis, testing, and second language acquisition.

MARÍA RAMÍREZ-MAYBERRY is a Master's candidate in the Spanish and Portuguese Department at the University of Texas at Austin. Her interests are second language acquisition and interlanguage development processes in the classroom environment.

JEFFREY REEDER teaches Spanish language and foreign language teaching methodology courses in the Department of Foreign Languages, Sonoma State University, Rohnert Park, California. His research interests include the acquisition of L2 speech, acoustic phonetics, and technology-enhanced language learning.

MARY ZAMPINI is Assistant Professor of Spanish Linguistics at the University of Arizona. Her interests include second language Spanish speech production and perception.

FROM THE EDITORS:

Dear Reader,

We are proud to dedicate the pages of this issue to the outstanding scholarship presented at the First UT Spanish Second Language Acquisition Symposium (SSLAS) held on October 3 and 4, 1997, at the University of Texas at Austin. Sponsored by UT's College of Liberal Arts and Department of Spanish and Portuguese, with additional support from McGraw-Hill Publishers and the Centro Bilingüe-Multicultural of Cuernavaca, Mexico, the Symposium drew large audiences to hear key speakers Susan Gass (Michigan State University), Judith Liskin-Gasparro (University of Iowa), James Lee (Indiana University), and Bill VanPatten (University of Illinois). Other invited speakers were Barbara González-Pino (University of Texas at San Antonio), Margaret Ann Kassen (Catholic University of America), Michael Thomas (University of Mary Hardin Baylor), and Dolly Young (University of Tennessee at Knoxville). All of these presenters, with the exception of Susan Gass, are alumni of UT-Austin.

The articles included in this special issue exemplify the four major strands of research presented at the Symposium: discourse analysis of learner production, acquisition of syntax and phonological features, testing, and multimedia applications. In

the area of discourse analysis, Rebecca Jo Bearden ("Discourse Features of Oral Production at the Novice Level") investigates Novice-Level oral production, a proficiency level that has received little attention in the past and is insufficiently described in the ACTFL Guidelines. She examines the salient discourse features of this level and proposes oral production tasks that are most appropriate for eliciting samples that best reveal Novice learners' performance. The second article of this group is that of Dale Koike and Fanny Hinojosa ("A Discourse Approach to the Assessment of Foreign Language Oral Proficiency"), who propose a global discourse approach to classroom assessment of oral proficiency for the advanced levels. According to this approach, the assessor compares actual learner production for a given task to the discourse pattern of a model, which in turn reflects the top-down organization used by most Advanced-level learners and native speakers for that task. After describing what learners of the Intermediate and Advanced levels say for a particular stimulus and comparing their responses to the pattern in the model, the authors propose a set of descriptors for varying levels of production.

The next four articles address the acquisition of particular Spanish grammatical morphemes and phonological features. Those dealing with grammar include a study by María Ramírez-

Mayberry ("Acquisition of Spanish Definite Articles by Native English-Speaking Learners of Spanish"), who examines written samples from beginning Spanish learners and analyzes, both quantitatively and qualitatively, the stages by which the learners acquire Spanish definite articles. She accounts for the stages of acquisition in terms of simplification, communicative value, and frequency of input (VanPatten, 1987). The other article concerning morphemes is that of Christopher Gascón ("The Spanish Psych Verb Construction: Beginners' and Intermediate Learners' Patterns of Usage"), who examines the use of Spanish psychological verbs (for example, gustar *'to like') by learners of four semester levels of Spanish language study. He discerns apparent stages of psych verb acquisition and draws support for his analysis from the work on interlanguage by Corder (1978) and Selinker (1972).*

Two articles focus on the acquisition of phonological features. Mary Zampini ("The Relationship Between the Production and Perception of Second Language Spanish Stops") investigates the acquisition of Spanish voiced and voiceless stop consonants by advanced native English speakers. Her data suggest that there is not a strong correlation between learners' ability to perceive Spanish stops and their ability to produce them. The study by Jeffrey Reeder ("An Acoustic Description of the Longitudinal Acquisition of Spanish Phonological Features by English

Speaking Adult Learners") addresses the acquisition of the Spanish vowels, a number of consonants, and other phonemic features by four levels of learners. He compares the data from those learners with data from of a group of native Spanish speakers. The results reveal stages in which learners progress to a more Spanish-like production.

The third group of articles deals with aspects of testing. The paper by Barbara González-Pino ("Prochievement Testing: Matching Instructor Expectations, Student Level, and Task Levels") looks at the role of instructors' assumptions regarding oral tests and identifies areas of discrepancy between testing and rating practices. She confirms the need for instructors to delineate their specific expectations with respect to learner levels and the various features tested. Michael Guerrero's article ("Current Issues in the Spanish Language Proficiency of Bilingual Education Teachers") critically examines the context in which bilingual education teachers develop Spanish language proficiency, and he finds that the context insufficiently supports their efforts to attain the expected high level. He also finds problematic the current measures used to gauge proficiency.

The final article of the collection is that of Margaret Ann Kassen ("Multiple Challenges of Multimedia: Development, Implementation, and Evaluation in Second Semester Spanish"),

who gives an overview of the development, implementation, and evaluation of multimedia lessons to enhance second language listening comprehension. In addition to proposing guidelines for the use of multimedia, she describes the Libra authoring system.

All of these studies represent valuable advances in the field of Spanish second language acquisition, and all open new perspectives to language acquisition in general. It is our hope that, by bringing the work of these Spanish language practitioners and researchers together in this single issue, we can encourage you, our readers, to explore the questions they have raised and to add your vision to the search for answers.

References:

Corder, P. (1978). Language-learner language. In J. Richards (Ed.), *Understanding second and foreign language learning* (71-93). Rowley, MA, Newbury House.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics, 10* (3), 209-230.

VanPatten, B. (1987). Classroom learners' acquisition of *ser* and *estar*. In B. VanPatten, T. Dvorak, & J. Lee (Eds.) *Foreign language learning: A research perspective* (pp. 61-75). Cambridge: Newbury House.

Discourse Features of Spanish Oral Production at the Novice Level

REBECCA JO BEARDEN, The University of Texas at Austin

Despite the recent emphasis in second language acquisition research on communicative ability, little analysis or description of Novice-Level speech has been produced. This paper examines this topic and has two primary purposes: first, to describe some of the distinguishing discourse features of Spanish oral production at the Novice Level, and second, to discuss the methodological implications of these findings. The data consist of six Oral Proficiency Interviews all rated at the Novice Level by two separate interviewer/raters trained by the American Council on the Teaching of Foreign Languages (ACTFL). The data were analyzed for the presence of common discourse features, including cohesive devices, use of the first language, negotiation, and scaffolding. Methodological implications of the findings are discussed, and finally a revision of the Novice-Level ACTFL proficiency guidelines that incorporates the findings of the current study is suggested. It is hoped that this investigation will help establish a more complete understanding of oral production at the lower levels of proficiency and thereby encourage the development of teaching and testing procedures designed to fit the needs and skills of the students in our foreign language classrooms.

INTRODUCTION

Oral proficiency has recently become the subject of much research on second language acquisition. Development of oral proficiency testing, begun in the early 1950s by the Foreign Service Institute (FSI), was subsequently applied to the academic setting with the American Council on the Teaching of Foreign Languages (ACTFL) Oral Proficiency Interview (OPI). The result was a surge of interest in this innovative method that has brought about a profound change in the goals and attitudes towards the teaching of foreign languages. The current trend is to teach students to create with the language and to use the language to fulfill a variety of functions. This trend is a dramatic departure from the traditional methods of foreign language instruction, which were described in an editorial published in a major newspaper as "rote drill in quirky idioms and irregular verbs" plus "'cultural appreciation'—the sampling of tacos, quiche or sauerbraten" (cited in Heilenman and Kaplan, 1985, p. 71).

The reality, however, is that even with this newfound emphasis on communicative ability, the vast majority of students enrolled in college level foreign language classes will probably never progress beyond the ACTFL Novice-Level rating. Despite this fact, the majority of research and study involving oral proficiency is concerned with the more advanced levels; very little analysis or description of Novice-Level speech has been produced. Furthermore, studies investigating the discourse features of Novice-Level speech are

virtually nonexistent. Much research in recent years has shown the critical role that discourse plays in the acquisition of a second language. It is, therefore, the intention of this study to examine some discourse features of Spanish OPIs rated at the Novice Level. The study has two purposes: first, to describe some of the distinguishing discourse features of oral production at the Novice Level, and second, to discuss the methodological implications of these findings. It is hoped that the investigation of these two topics will help advance our understanding of oral production at the lower levels of proficiency and thereby encourage the development of teaching and testing procedures designed to fit the needs and skills of the majority of the students in our foreign language classrooms.

BACKGROUND LITERATURE

The movement towards proficiency-based teaching and testing techniques can be considered to have originated in the early 1950s as a result of an initiative by the FSI to develop meaningful verbal descriptions of various skill levels in order to find a more accurate measure of the ability of foreign service officers to function in their overseas assignments. The FSI scale that was developed identified five proficiency levels, ranging from *Level 1, Elementary Proficiency*: Able to satisfy routine travel needs and minimum courtesy requirements, to *Level 5, Native or Bilingual Proficiency*: Speaking proficiency equivalent to that of an educated native speaker. In an effort to make these descriptions more applicable and useful in an academic setting, ACTFL subsequently revised the FSI scale to include more and finer-tuned levels at the lower end. Recognizing that most students of foreign language never reach proficiencies above a Level 3 on the FSI scale, the ACTFL scale acknowledged that more levels were needed to characterize and evaluate the oral proficiency of high school and college-level foreign language students. For this reason, the ACTFL scale does not include categories above the Level 3 on the FSI scale.

The implementation of proficiency-based learning has been gaining popularity in academic settings ever since the development of the ACTFL scale. Heilenman and Kaplan (1985) explain that, in the more traditional approaches, "the assumption seems to have been that learning a foreign language equals learning that language's structure, along with a generous amount of vocabulary, all carefully sequenced and spooned out in judicious doses. Enough of these doses, successfully swallowed, presumably result in functional language knowledge. It sounds nice. It sounds logical. It sounds as if it should work. But it doesn't" (p. 58). They contend that proficiency-based curriculum design and testing is an idea whose time has come. As they explain, "Proficiency is seen as a superordinate goal that represents more than the sum total of all discrete items learned and that attempts to balance accuracy with fluency and learning about a language with providing the

opportunity to learn a language by using it" (p. 59).

Despite the growing enthusiasm for proficiency-oriented testing and instruction in academic settings, there is a paucity of studies investigating the lowest proficiency levels. This lack of research is even more striking in light of Heilenman and Kaplan's statement that 80% of foreign language students currently enrolled in college-level foreign language courses will not likely progress beyond the Novice Level (p. 70). The most detailed information to be found about Novice-Level oral production is the generic descriptions of the 1986 ACTFL proficiency guidelines themselves, which are reproduced below.

Novice	**Novice level is characterized by an ability to communicate minimally with learned material.**
Novice-Low	Oral production consists of isolated words and perhaps a few high-frequency phrases. Essentially no functional communicative ability.
Novice-Mid	Oral production continues to consist of isolated words and learned phrases within very predictable areas of need, although quantity is increased. Vocabulary is sufficient only for handling simple, elementary needs and expressing basic courtesies. Utterances rarely consist of more than two or three words and show frequent long pauses and repetition of interlocutor's words. Speaker may have some difficulty producing even the simplest utterances. Some Novice-Mid speakers will be understood only with great difficulty.
Novice-High	Able to satisfy partially the requirements of basic communicative exchanges by relying heavily on learned utterances but occasionally expanding these through simple recombinations of their elements. Can ask questions or make statements involving learned material. Shows signs of spontaneity, although this falls short of real autonomy of expression. Speech continues to consist of learned utterances rather than of personalized, situationally adapted ones. Vocabulary centers on areas such as basic objects, places, and most common kinship terms. Pronunciation may still be strongly influenced by first language. Errors are frequent and, in spite of repetition, some Novice-High speakers will have difficulty being understood even by sympathetic interlocutors.

Other researchers have made very limited observations as to the nature of Novice-Level speech. These observations, like the ACTFL descriptions themselves, usually characterize oral production in terms of the components of Higgs and Clifford's (1982) Functional Trisection. This measurement identifies three components of language use that constitute the global rating at each proficiency level: *function*, which refers to the linguistic tasks the learner is able to complete successfully; *content*, which describes the setting in which those tasks are performed; and *accuracy*, which refers to the grammar, pronunciation, and syntax of the text.

With regard to function, Heilenman and Kaplan (1985) state that "learners [at the Novice Level] are more accurately characterized by what they are incapable of doing than by what they succeed in doing" (p. 58). Bragger (1985) concurs that "Novice-Level speakers have practically no functional ability, although they can communicate very simply with memorized material" (p. 97). Furthermore, the OPI training manual (Buck, 1989) explains that

> Novice Level speakers have acquired some of the building blocks necessary for creating their own utterances, but they cannot access and manipulate individual elements of learned material and thus can't adapt them to express unique, personalized, or situation-specific messages. Thus the Novice Level speaker is essentially limited to reacting to the conversational partner in highly predictable common daily settings and is only marginally able to initiate communication. (pp. 2-7)

Finally Valdman and Phillips (1977) describe the effect of the Novice-Level speaker's limited functional ability on the development of the interlanguage, stating,

> In initial stages of second language learning, the target language serves only a narrowly communicative function; as a consequence, it is acquired in a highly deviant form that exhibits all the characteristic features of pidginization: emphasis on content words, invariable word order, elimination of functors, etc. As the target language's range of function increases, the learner's interlanguage progressively expands and complexifies. (p. 22)

With regard to content, Omaggio (1986) identifies ten subject areas that Novice speakers generally control sufficiently to be able to list vocabulary associated with the subjects or ask and answer simple questions. These include names of basic objects, family members, colors, articles of clothing, weather expressions, days of the week, months, dates, and time. She states that these ten topics are known as the "ten desperate questions" because they "will often elicit some sample of speech when all other attempts at conversation fail" (p. 16). Omaggio explains that speakers at the Novice Level cannot create or paraphrase with the language, nor are they adept at handling conversation topics that they have not rehearsed before. She further states that they are not able to handle

a survival topic such as checking into a hotel or ordering a meal in a restaurant. She points out that control of these skills is the requirement for an Intermediate-Level rating.

Finally, with regard to accuracy, Bragger states that, "at the Novice Level, accuracy is defined primarily as *intelligibility* because few if any grammar structures exist in the speech to warrant discussion of the precision of the message conveyed" (pp. 97-98).

Omaggio (1986) explains that according to the Relative Contribution Model proposed by the Interagency Language Roundtable, different aspects of language use will play varying roles of importance in the global rating of oral interviews, depending on the proficiency level. Thus, at the Novice Level, vocabulary is the most important element in the speaker's attempts to communicate, while pronunciation, grammar, fluency, and sociolinguistic appropriateness are considered to a lesser degree in the global rating at this level.

Omaggio sums up the nature of Novice-Level speech by stating, "Essentially, the Novice-Level interview is like an oral achievement test, since the learner can produce little more than what he or she has learned in the course of his exposure to the target language" (p. 340).

DATA COLLECTION

The current study analyzed six OPIs conducted by an ACTFL-certified interviewer. The data consisted of two interviews rated at the Novice-Low (NL) level, two rated at Novice-Mid (NM), and two at Novice-High (NH). Each interview was evaluated by two raters, and agreement on the rating was reached in each case. The oral interviews ranged from 5 to 15 minutes and were taped and later transcribed by the researcher.

The NL-level interviews were the shortest. Both began with a series of personal questions such as "Where are you from?", "Where do you live now?", "Do you like it here?", and "What classes are you taking?" The second task in the NL-level interviews required the learners to name various objects indicated by the interviewer. At the NM-level the first task was the same, but the initial personal questions were often followed by additional questions requiring the learners to elaborate on their answers. In one of the NM interviews the second task was also a rather extended naming of objects, but in the other NM interview the second task required the learner to pose questions to the interviewer. The NH-level interviews again began with personal questions that were usually followed by additional questions requiring the learners to elaborate on their responses. The second task at this level was a role play situation, and the third task required the learners to pose questions to the interviewer.

The data were analyzed for common discourse features that could be used to characterize Novice-Level speech. The features that were considered were the use of cohesive devices, switching into the first language (L1), the use of scaffolding, and interactional devices used in negotiation. Finally, some methodological implications of these findings were derived.

RESULTS AND DISCUSSION

1. Cohesive Devices

Looking at the data, we find some interesting discourse patterns emerging already at the Novice Level. The utilization of cohesive devices serves as a good initial example. As Pavlou (1994) explains, "When we are dealing with the cohesion of a text we are investigating the linguistic means which enable a text to function as a single meaningful unit" (p. 4). He goes on to identify three major categories of cohesive devices: referential, conjunctional, and lexical. We find examples of all three types of cohesive device in the data. By "referential cohesive device," Pavlou refers to the use of pronouns. Observe Robin, for instance, as she talks about her roommate in Example 1. [Note: utterances by the learner are in the right column, while those by the interviewer are in the left.]

Example 1. Robin (NH): Referential Cohesion

1		Oh uh Nancy es muy *amidable. 'Nancy is very friendly'
2		*Amidable? 'Friendly?'
3	Um hum	
4		¿Sí? 'Yes?'
5	¿Y y cómo es físicamente cómo es? 'And what is she like physically, what is she like?'	
6		Oh um pequeña. Uh ella tiene um . . . blonde hair.
7		Uh ella es linda. 'Oh um small. Uh she has um . . . blonde hair.Uh she is pretty.'

In Line 1 Robin uses her roommate's name to establish the reference, but in Lines 6 and 7 she chooses the pronoun *ella* 'she.' This usage gives the discourse cohesion in that the interlocutor infers topic continuity from the discourse.

The occurrence of referential cohesion in the data is shown in Table 1. The data suggest that the use of referential cohesive devices greatly increases as proficiency increases. At the NL level the two learners used no pronouns at all, and at the NM level the two learners used a total of three pronouns. At the NH level we find a dramatic increase in the use of pronouns; the two learners use a total of 31 pronouns. The data, therefore, demonstrate that pronouns are used for greater cohesion in the discourse at the higher levels.

An interesting parallel finding concerns the use of the pro-drop feature across the three proficiency levels. It seems that, since Spanish is a pro-drop language, we must consider the absence of subject pronouns as another referential cohesive device. With the pro-drop feature, as with the use of pronouns, the interlocutor must make refer-

Table 1
Use of Pronouns and Pro-Drop by Learner and Level

Learner & Level	# of Pronouns	Instances of Pro-drop
Leslie - NL	0	1
Chin - NL	0	1
Frank - NM	1	5
Marissa - NM	2	6
Vince - NH	18	4
Robin - NH	13	3

ence to a previous utterance in order to understand fully the relationships between themes in the current message. Consider Example 2, in which Marissa talks about her brother:

Example 2. Marissa (NM): Use of Pro-drop

1	Ah Nueva York. ¿Y qué hace	
2	ahora Eric? ¿Está trabajando? 'Ah New York. And what does Eric do now? Is he working?'	
3		Sí um vive en Bangor, Maine uh
4		trabajando uh youth pastor. 'Yes um he lives in Bangor, Maine uh working uh youth pastor'
5	Oh ah hah. Y . . . y tu hermano	
6	Andy, ¿cómo es? 'And . . . and your brother Andy, what is he like?'	
7		Es um . . . um *mediocre 'He is um . . . um (invents word)'
8	Ah ha	
9		En height. Uh pelo es . . . (oh) brown 'In height. Uh hair is . . . '
10	Uh huh.	
11		Uh es veinte y uno años. 'Uh he is twenty-one years.'
12	Um hum.	
13		Um es um . . . uh (indo) uh going um a
14		la universidad en Bangor Maine 'Um he is um . . . uh (ing) uh going um to the university in Bangor Maine'
15	Uh huh.	
16		Uh vive con Eric. 'Uh he lives with Eric'

We see consistently throughout this example that Marissa provides no overt subject for her statements.

Table 1 shows that at the NL level each learner used the pro-drop feature once, while at the NH level the two learners used pro-drop a total of seven times. Thus, we also find greater use of this cohesive device at the higher levels of proficiency. A somewhat contradictory finding, however, is that the NM level is characterized by the greatest use of pro-drop. We find that the two learners used this feature a total of 11 times. The other NM learner demonstrates a similar grasp of the pro-drop aspect of Spanish. It is interesting to note that, in this regard, the learners do not show evidence of transfer of their L1 rules that require pronouns, but appear to have adopted quite early the second language (L2) norm and are able to exploit verbal morphology to a greater or lesser degree to maintain cohesive discourse across successive turns.

The question arises, then, why we see a reversal of this tendency at the NH level, where the learners begin supplying overwhelmingly more pronouns and use the pro-drop feature less frequently. There seem to be several plausible answers to this question. First, the NH level is the stage at which the learners start developing their syntactic systems and are producing longer strings of words. It could be that they are still uncomfortable with their command of word-order rules and their ability to exploit verbal morphology accurately; thus, they supply more pronouns in an attempt to ensure their meaning is understood. A second possible explanation is that, at the NL and NM levels, we see much more imitation of the interviewer's speech, which does not generally include pronouns. Perhaps at the NH level, as the learners start to rely less on memorized material and to construct their own sentences, they fall back on their L1 rules and include pronouns in their utterances. The result is that the discourse at this level often seems more redundant and less native-like than the discourse at the NM level with regard to the referential aspect of cohesion.

Pavlou identifies the use of conjunctions as another feature of cohesive discourse. In Table 2 we see a gradual increase in the frequency of conjunctions as proficiency increases. Table 2 shows that the learners at the NL level use no conjunctions. This finding is not surprising given that their utterances are rarely longer than a single word. The only conjunction used at the NM level is *y* 'and.' Again, however, the learners at this level do not produce utterances longer than a single clause and therefore their speech rarely requires the use of conjunctions. At the NH level, on the other hand, we see a much wider variety of conjunctions, and they are used more frequently. Vince in particular uses conjunctions productively and, in addition to *y* 'and,' his repertoire includes *o* 'or,' *pero* 'but,' and *porque* 'because.' He produces utterances of much more complexity than we see at the NL and NM levels, and he achieves a considerable degree of cohesion by the inclusion of these conjunctions, as evidenced in Example 3, where he explained why he studies Spanish.

Table 2
Use of Conjunctions

Learner & Level	No. of Conjunctions
Leslie - NL	0
Chin - NL	0
Frank - NM	2
Marissa - NM	3
Vince - NH	16
Robin - NH	3

Example 3. Vince (NH): Use of Conjunctions

1		Oh, es necesario tener dos años de español.
		'It is necessary to have two years of Spanish.'
2	¿Ah sí? 'Oh yes?'	
3		Sí, sí. Muy importante porque uh en
4		Jacksonville uh ellos no hablan español uh
5		muchas pero uh um va a sur de Jacksonville,
6		uh en Orlando, Miami habla español muy
7		bie . . . uh muchas.
		'Yes, yes. Very important because uh in Jacksonville uh they don't speak Spanish uh many but uh um go to south of Jacksonville, uh in Orlando, Miami speak Spanish very we . . . uh many.'

Robin (NH) used only two conjunctions during her interview, *y* and *pero*, and she was not able to create the same degree of cohesion in her discourse as Vince demonstrated in Example 3.

The third feature of cohesive discourse identified by Pavlou is lexical devices. The only examples of these in the current data are provided by Vince (NH) when he employs the word *también* 'also' on several occasions to link successive utterances. Again, although he displays a lack of variation in his use of cohesive devices, the learner's use of these lexical elements gives his discourse a fluidity that is not seen in the less proficient samples.

2. Use of L1

At all three proficiency levels in the data, the learners often revert back to the use of the L1 while attempting to converse in the L2. In these instances, the use of the L1 appears to serve a variety of purposes. The learners often use the L1 to supply a

vocabulary word that is unknown in the L2. Sometimes the learners solicit a translation of the L1 word, while in other cases they use a word in the L1 to fulfill a communicative need. Example 4 contrasts these two uses of words in the L1.

Example 4. Vince (NH): Contrasting Uses of L1

1	¿Es, cómo, es similar a Middlebury	
2	o es diferente?	
	'Is it, how, is it similar to Middlebury or is it different?'	
3		Uh es no muy diferente, sí uh. . . .
		'Uh it is not very different, yes uh . . . .'
4	¿Cómo, qué diferencias hay?	
	'How? What differences are there?'	
5		Um es un um universidad de um
6		community colegio uh que no es un
7		um . . . es un uh . . . como se dice upper
8		level uh . . . .
		'Um it is a um community college university um that is not a um . . . it is a uh . . . how do you say "upper level" uh . . . .'
9	Sí, es de dos años.	
	'Yes, it is for two years.'	
10		Uh, sí sí sí. Dos años.
		'Uh, yes,yes,yes. Two years.'

In Line 6, Vince uses the word *community* without a rise in intonation or explicitly asking for a translation, and he continues speaking without hesitation. In Lines 7 and 8, however, he requests a translation of the L1 word.

Throughout the data the learners appear to use the L1 in times of L2 linguistic breakdown when the communicative pressure of the situation is great and they do not want to interrupt their message by having to solicit translations, as in Example 5.

Example 5. Marissa (NM): Use of L1 When L2 Breakdown Is Experienced

1	¿Estudias mucho? ¿Estudias?	
	'Do you study a lot? Do you study?'	
2		¿Aquí? 'Here?'
3	Uh huh.	
4		Uh um no as mucho 'much' um. . . as I should.
5	(Laughs) Ok. ¿Y dónde vives aquí	
6	en Middlebury?	

	'And where do you live here in Middlebury?'	
7		En Batel North dormitorio. 'In Batel North dormitory'
8	Um hum, ¿Te gusta Batel? 'Um hum, Do you like Batel?'	
9		Uh sí es muy um . . . uh energy 'Uh yes it is very um . . . uh energy'

In Example 5, Marissa uses the L1 in Lines 4 and 9 when she lacks the language structures to communicate her thoughts, but she also chooses not to ask for a translation of the word in Line 9, even though she knows the phrase *Cómo se dice* 'How do you say,' as evidenced later in the interview. This avoidance, too, could be a function of the relationship that has been established with the interviewer. In several previous instances during the interview, Marissa used a word in English with rising intonation to request a translation and the interviewer did not oblige. It could be that, by this point in the interview, Marissa had given up trying to negotiate at times of linguistic breakdown and just used the L1 in these situations.

Another instance in which we find frequent use of the L1 in the data is when the learner is attempting to clarify or repair a previous utterance in the L2. Take the following from Frank (Example 6):

Example 6. Frank (NM): Use of L1 to Clarify or Repair

1	Um hum. Muy bien, muy bien. ¿Y e en South	
2	Dakota, hace mucho calor? 'Um hum. Very good, very good. And in South Dakota, is it very hot?'	
3		Sí mucho calor. Muy, muy. 'Yes very hot. Very, very.'
4	¿Sí? ¿Siempre? 'Yes? Always?'	
5		Sí uh en . . . v . . . *vierno es . . . 'Yes uh in . . . w . . . winter is...'
6		that's Spring right?
7	En invierrno hace frío, ¿no? 'In winter it's cold, right?'	
8		Ok that's what I thought.
9		En invierno es muy frío,
10		muy . . . ' In winter is very cold, very . . .'
11	¿Pero te gusta South Dakota? 'But do you like South Dakota?'	
12		Sí, me gusta South Dakota. 'Yes, I like South Dakota.'
13		I don't know why.
14	(Laughs)	
15		But I do.

Frank uses the L1 in Line 6 to verify that he has correctly expressed what he intended to say in Line 5, and again in Line 13 he uses the L1 to clarify and elaborate on his statement in Line 12. This tendency to use the L1 to clarify and elaborate previous utterances in the L2 is common in the data at all three levels of proficiency.

3. Whispering to Self

Yet another use of the L1 in the data serves a very different function. Platt and Brooks (1994) found in their analysis of two beginning learners of English that their subjects frequently whispered comments to themselves in the L1. These comments appeared to be directed principally at the learners themselves and not at the interlocutor. They explain that this whispering to self "serves important mediational purposes for individuals for completing cognitively demanding tasks and for orienting themselves and their interlocutors to the task and to the language that is used to construct the task" (p. 507). We find examples of such behavior in the present data as well (Example 7).

Additionally, there are many examples of the learners whispering to themselves all or part of a statement that the interviewer just directed at them (Example 8).

Example 7. Robin (NH): Whispering to Self in L1

1	¿Qué hacen Uds. uh cuando cuando no	
2	van a Burlington y cuando no estudian,	
3	qué hacen Uds.?	
	'What do you do uh when when you don't go to Burlington and when you don't study, what do you do?'	
4		Um . . . (let's see) vamos a la 'we go to the'
5		Fitness Center?
6	Uh huh ¿y está cerca? 'and is it nearby?'	
7		Sí. 'Yes.'

Example 8. Marissa (NH): Whispered Repetition of Interviewer

1	Uh, ¿y qué trabajo hace tu padre?	
	'Uh, and what work does your father do?'	
2		Real estate.
3	Uh huh, no sabes. 'Uh huh, you don't know.'	
4		(No sabes)
5	No sabes. Um pues mira, um ¿dónde	
6	vives aquí en Middlebury?	
	'You don't know. Um well look, um where do you live here in Middlebury?'	
7		Yo vivo en uh Stuart.
		'I live in uh Stuart.'
8	¿Es una residencia? 'Is it a dormitory?'	

9		Sí. 'Yes'
10	Sí. ¿Y tienes compañera de cuarto? 'Yes. And do you have a roommate?'	
11		(Compañera?)
12	Compañera de cuarto. Otra persona en	
13	el cuarto 'Roommate. Another person in the room'	
14		Oh compañera sí, sí. 'Oh roommate yes, yes'

Examples 7 and 8 appear to be instances in which the learners try to grapple with utterances that were not understood and to access information stored cognitively. The fact that the learners whisper makes it doubtful that they are addressing the interlocutor; thus, these examples should not be considered a request for clarification or a comprehension check, even though the interviewer sometimes responds to it as such.

4. Discourse Strategies

Perdue and Klein (1992), in their study of two beginning learners of English, report that their subjects relied heavily on scaffolding with their interlocutor to express themselves (p. 268). The data gathered in the present study support this finding. The learners at the NL and NM levels rarely give more than a one-clause answer before pausing and waiting for a response from the instructor. At all three levels of proficiency the learners make few attempts to initiate topics. At the NL level the learners never take control of the conversation. Frank does initiate an interaction when he is instructed to ask the interviewer a question, but he is uncomfortable in this role and he quickly throws the responsibility back to the interviewer (Example 9):

Example 9. Frank (NM): Avoiding Topic Initiation

1	¿Tienes tú alguna pregunta para mí? 'Do you have a question for me?'	
2		Uh, ¿un profesor? ¿Tú un profesor? 'Uh, a professor? You a professor?'
3	Aquí en Middlebury no. 'Here at Middlebury, no.'	
4		¿No? 'No?'
5	Estoy tres días aquí con las entrevistas. 'I am here three days with the interviews.'	
6		No ah otra preguntas. 'No ah another questions.'

At the NH level there are examples of several hesitant attempts to take control of the interaction, especially in the role-play situation.

The success of these exchanges, however, is still dependent on help from the interlocutor to supply key vocabulary words, as in Example 10 from Vince.

Further evidence for the importance of scaffolding is found in the frequent occurrence of several interactional features identified by Ellis (1985). The features that are most common in the current data include clarification requests, comprehension checks, repetitions, and expansions. The learners rely on these fea-

Example 10. Vince (NH): Interlocutor to Supply Key Vocabulary

1	¿Y cúal es tu trabajo? 'And what is your work?'	
2		Ah yo trabajo con la compan
3		compañía de telefono AT&T. 'Ah I work with the telephone compan company AT&T.'
4	Ah sí? 'Oh yeah?'	
5		Sí, uh tenemos muchos um . . . um . . .
6		¿cómo se dice customers? 'Yes, uh we have many um . . . how do you say customers?'
7	Clientes. 'Customers'	
8		Ah, ok. Tenemos muchos *quientes . . . 'Ah, ok. We have many (mispronounces)'
9	Clientes. 'Customers'	
10		Clientes que hablan español. 'Customers who speak Spanish.'

tures to build vertical constructions that enable an exchange of meaning.

A clarification request occurs when the learner does not understand part of the interlocutor's utterance and solicits help. The learners in the data used a wide variety of phrases to request clarification, including *no entiendo* 'I don't understand,' *¿perdón?* 'pardon?' *repita* 'repeat,' *¿cómo?* 'what?' and *¿qué?* 'what?' An even more common technique, however, was the repetition of the misunderstood part of the utterance with rising intonation, as in Example 11, where Frank has just responded to the question, "Where are you from?":

Example 11. Frank (NM): Clarification Requests

1	¿En South Dakota? Ah muy lejos. 'In South Dakota? Oh very far.'	
2		¿Muy lejos? Ah . . . ¿qué lejos? 'Very far? Ah . . . what far?'
3	Muchas millas. 'Many miles.'	
4		¿Muchas millas? No, no um . . . .

'Many miles? No, no um . . . .'

5 ¿Y tienes familia en South Dakota?
'And do you have family in South Dakota?'

In this example, Frank indicates in Line 2 that he does not understand the word *lejos* and he asks for clarification. The interviewer responds with circumlocution, but in Line 4 we see that understanding is still not achieved, because Frank again repeats the interviewer's words with rising intonation. So in Line 5, the interviewer switches the topic.

We find additional examples of repetition of part of the interviewer's utterance with rising intonation from Frank later in the interview (Example 12):

Example 12. Frank (NM): Comprehension Checks

1	Ah ha. Um, ¿Cuándo cuándo	
2	empiezan las clases?	
	'Ah ha. Um, When do classes begin?'	
3		¿Cuándo empiezan? 'When do they begin?'
4	Sí 'Yes'	
5		Uh mi clases uh um es . . . a la ocho
		'Uh my classes uh um is . . . at eight'
6	¿A las ocho? 'At eight?'	
7		Uh to um son las uno. 'Uh to um it is one.'
8	¿La una? 'One?'	
9		La una 'One'
10	Um hum. ¿Todos los días?	
	'Un hum. Every day?'	
11		¿Todos los días? Sí uh all cinco días.
		'Every day? Yes uh all five days.'
12	¿Qué días tienes clase?	
	'What days do you have class?'	
13		¿Qué días? Uh lunes, martes, miercoles, jueves,
14		viernes. 'What days? Uh Monday, Tuesday, Wednesday, Thursday, Friday.'

In Example 12, however, the repetition with rising intonation serves a different purpose. Here we have instances of comprehension checks. In Lines 3, 6, 8, 11, and 13, the speaker wants to ensure that he has correctly understood the interviewer's meaning; thus, he repeats all or part of the previous utterance in an attempt to confirm what he has heard.

Repetition is an interactional device that serves several purposes in the data. In addition to the examples of repetition described above, there are also frequent examples of other repetitions where the speaker repeats all or part of the interlocu-

tor's previous utterance without the rising intonation that would signal a comprehension check or a clarification request and thus require a response from the interlocutor. In Line 8 of Example 13 we find an occurrence of this feature on the part of the interviewer:

Example 13. Leslie (NL): Self-Repetition

1	Uh huh, ¿Y tienes una compañera de	
2	cuarto? Una amiga que vive contigo?	
	'Uh huh, And do you have a roommate? A friend that lives with you?'	
3		Sí. 'Yes.'
4	¿Y cómo se llama? 'And what is her name?'	
5		Leslie.
6	No tú te llamas Leslie, ¿pero tu amiga? 'No your name is Leslie, but your friend?'	
7		Oh, Amanda.
8	Amanda. ¿Y cómo es Amanda? Quiero	
9	información sobre Amanda. 'Amanda. And what is Amanda like? I want information about Amanda.'	

Also evident in Example 13 is the use of self-repetition. We find many instances of self-repetition in the data where the interviewer repeats or rephrases all or part of the previous utterance, as in Lines 1 and 2, and Lines 8 and 9 above.

A fourth interactional device prevalent throughout the data is expansion. This device appears when the learner rephrases the interlocutor's previous utterance and adds grammatical and/or semantic information, thus creating a *more complete* and complex thought. Observe Example 14.

Example 14. Frank (NM): Expansion

1	¿Tienes hermanos? 'Do you have siblings?'	
2		Sí un hermano uh es Benjamin uh
3		en Germany. 'Yes one brother uh he is Benjamin uh in Germany.'
4	Oh, está en Alemania. ¿Y cuántos años	
5	tiene Benjamin? 'Oh, he is in Germany. And how old is Benjamin?'	

Here the interviewer uses the information provided by the learner in Line 3 to form a complete and grammatically accurate sentence in Line 4. Again, the restatement of the information provided in Line 3 cannot be understood as a confirmation check because there is no rising intonation. Hatch (1978) also argues that such examples should not be considered repairs, because the rules of polite conversation would allow for a repair only if the learner had solicited one by using rising intonation in the difficult part of his utterance, as in Example 15. There we see that in Line 2 Vince is unsure of the word *bilote* so he solicits a repair with the use of rising intonation. When the difficulty is still unresolved, he tries again to get help by repeating the problematic word, and

Example 15. Vince (NH): Learner Solicits a Repair

1		Te gusta um la biblio uh bi um
2		*bilote? Um um. . . biblio. . . 'Do you like um the libra uh li um (invents word)? Um um . . . libra . . . '
3	¿Biblioteca? 'Library?'	
4		No, no biblioteca. (Laughs) Uh um
5		*bilota? Um bicycle? 'No, not library. (Laughs) Uh um (invents word)? Um bicycle?'
6	Bicicleta. 'Bicycle.'	
7	Sí. 'Yes.'	Bicicleta, bicicleta. Sí, sí,sí, sí. 'Bicycle, Bicycle. Yes, yes, yes, yes.'
8	Para mi la bicicleta es transporte,	
9	no es uh diversión.	
	'For me the bicycle is transportation, not diversion.'	

finally gives up and uses the word in English. Hatch proposes that the type of expansion found in the previous example from Frank, therefore, is not a repair, but rather serves to reassure the learner that he has been understood (p. 427).

Ellis (1985) found that the patterns of occurrence of several interactional features vary at different levels of proficiency and that this variance is often the result of the speaker's attempts to achieve comprehensible input. As he explains, "Because comprehensible input is a negotiated rather than an absolute phenomenon, dependent on the learner's developing communicative proficiency. . . different features may aid development at different times" (p. 82). The data in the present study support this finding. Tables 3 and 4 show the occurrence of interactional features at each level of proficiency and for each learner. Table 3 shows that self-repetitions and other repetitions are common across all three levels of proficiency, but, as in the Ellis study, self-repetitions decrease as proficiency increases. At the

Table 3
Interactional Features by Proficiency Level

Interactional Feature	NL	NM	NH
Clarification requests	0	10	6
Comprehension Checks	0	9	6
Self-repetitions	16	8	8
Other repetitions	5	7	5
Expansions	6	2	5

Table 4
Use of Interactional Features by Learner

Learner & Level	Clar. Requests	Comp. Checks	Self-repetition	Other repetition	Expansion
Leslie - NL	0	0	10	2	3
Chin - NL	0	0	6	3	3
Frank - NM	4	7	3	5	2
Marissa - NM	6	2	5	2	0
Vince - NH	5	5	4	2	3
Robin - NH	1	1	4	3	2

NL level there are a total of sixteen self-repetitions, while at the NH level there are only eight. Unlike the Ellis study, the current data show no significant change in the number of expansions from the NL to the NH level, but the increase noted by Ellis in comprehension checks as proficiency increases is found in the current data as well. Comprehension checks, however, were most common at the NM level, with a total occurrence of nine for the two NM learners, while at the NH level there was a total of six. The occurrence of clarification requests similarly was greatest at the NM level, with a total of ten occurrences versus six occurrences at the NH level. In the present data there were no instances of clarification requests or comprehension checks at the NL level.

Due to the limited scope of both the present and the Ellis studies, it is difficult to draw any firm conclusions about the relationship between proficiency and the frequency of occurrence of these interactional features. Ellis cautions that the interactional styles of the participants will probably be equally if not more influential than the proficiency levels in determining use of these features (p. 76). Thus, at the NH level, for example, five clarification requests and five comprehension checks were used in Vince's interview, while only one instance of each feature was found in the data from Robin.

What is confirmed in the Ellis study, however, is the role these interactional features play in second language acquisition. In a seminal article discussing the role of discourse analysis in second language research, Hatch (1978) questioned the well entrenched position that the ability to converse was a direct result of having acquired a sufficient amount of language. She proposed conversely that language acquisition grew out of conversation (pp. 406-7). She explains, citing Scollon, that interactions called "vertical structures" form the basis for syntactic constructions that Scollon calls "horizontal structures" later on. We find an example in the current data from Chin (Example 16):

Example 16. Chin (NL): Vertical Structures That Lead to Horizontal Structures

1	¿Cómo se llama tu profesor de matemáticas? 'What is your math professor's name?'	
2		Um Peterson.
3	Uh huh. ¿Es hombre o es mujer? ¿Hombre o mujer? 'Uh huh. Is that a man or a woman? Man or woman?'	
4		Hombre. 'Man.'
5	Uh huh. Es señor Peterson. 'Uh huh. He is Mr. Peterson.'	
6		Señor Peterson. 'Mr. Peterson.'
7	Uh huh. ¿Y tu profesor um . . . tu profesor de español,	
8	es señor o señora? 'Uh huh. And your professor um. . . your Spanish professor is Mr. or Mrs.?'	
9		Señor um Gustavo Mía 'Mr. um Gustavo Mía'
10	Mejilla	
11		Mejilla
12	Uh huh	
13		Señor Mejilla 'Mr. Mejilla'

Here we find the interviewer uses expansion and repetition to provide the building blocks with which Chin is able to produce a proper title in Spanish in Line 13.

5. Negotiation

Pica (1987) also suggests that negotiated interactions are crucial to help the learner reach higher levels of proficiency. As she explains,

> What enables learners to move beyond their current interlanguage receptive and expressive capacities when they need to understand unfamiliar linguistic input or when required to produce a comprehensible message are opportunities to modify and restructure their interaction with their interlocutor until mutual comprehension is reached. (p. 8)

Pica (1994) points out, however, that this type of negotiation does not appear to help the learners develop accuracy. A particularly striking example to support this conclusion can be found in data from a previous study by Bearden (1997) (see Example 17).

Example 17. Learner #1: Failure to Improve Accuracy

1		Tengo un esposo. ' I have a husband.'
2	¿Tú tienes esposo? 'You have a husband?'	
3		No, ella tenga. . . 'No, she have . . .'
4	Tiene 'Has'	
5		tiene un esposo . . . uh se llamo Rick
6		y tengo un . . . uh . . . niño. 'has a husband . . . uh his name is Rick and I have a . . . uh . . . boy.'
7	Um hum...¿tú tienes? 'you have?'	
8		No ella . . . ella tienes un niño. 'No she . . . she have a boy.'
9	Uh huh	
10		Se llama . . . uh . . . Nick . . . uh . . . tengo dos
11		años. 'His name is . . . uh . . . Nick . . . uh . . . I am two years old.'
12	No, ¿tú tienes dos años? 'No, you are two years old?'	
13		*Ello tienes '(invents pronoun) have'
14	Él 'He'	
15		Él ten . . . tengo ' He ha . . . I have'
16	Tiene 'Has'	
17		Tiene dos años ' He has two years = He is two years old'
18	Ok	
19		Um . . . soy . . . ella mi hermana soy
20		muy simpático. 'Um . . . I am . . . she my sister I am very nice.'
21	Ok	

In Line 1 the learner intends to refer to his sister, but states *Tengo un esposo* 'I have a husband.' The instructor initiates a repair in Line 2

that prompts the learner to attempt to correct his statement in Line 3. His new statement is still incorrect with regard to mood, but is now inflected properly for person. In Line 4, the instructor supplies the correct form that the learner repeats in Line 5, but later in the same utterance, he reverts back to the incorrect form of the verb he originally used in Line 1. The whole process is repeated two more times in Lines 7 through 17. This example makes a clear case for the learner's inability to process the form of his utterance or the corrections that the instructor is making while under the immediate pressure of trying to communicate specific information.

A study by Gass and Varonis (1994), however, suggests that these negotiated interactions may improve accuracy in subsequent conversations by focusing the learners' attention on the problematic part of the utterance and thus enabling them to correct the problem. The authors state,

> What we claim is that interactional input provides a forum for learners to readily detect a discrepancy between their learner language and the target language and that the awareness of the mismatch serves the function of triggering a modification of existing second language knowledge, the results of which may show up at some later point in time. (p. 299)

Another explanation for the lack of improvement with regard to accuracy may be that grammar is rarely the topic of negotiation in a natural conversational setting. In fact, the OPI Training Manual (Buck, 1989) specifically prohibits interviewers from correcting grammar during the interview. There are no instances of negotiation concerning accuracy in the data from the present study. Hatch (1978) argues that the biggest challenge facing adult learners of a second language is identifying the topic and that topic nomination, rather than grammatical accuracy, is the purpose of most negotiation involving language learners. She explains that, with children, there is a greatly reduced set of topics the learner can be expected to comment on, usually concrete things or on-going actions, so the topics of conversation are naturally limited. With adults, however, there is a much greater variety of possible topics and time references, which makes topic identification a major difficulty. Hatch explains, "Since the topics in adult discourse are much more varied and more abstract, it is much more difficult for the adult to identify conversation topics unless he knows the necessary vocabulary for that topic" (p. 431). According to Hatch, the purpose of negotiation at the beginning levels is soliciting and clarifying vocabulary purely in an attempt to establish the topic.

There is ample evidence to support this position in the current data. At the NL, where self-repetition is the greatest, we find that the interviewer uses the technique of stating and then immediately paraphrasing a question as a means of providing a wide range of vocabulary clues to signal the topic. A good example is found in the data from Leslie (Example 18):

Example 18 Leslie (NL): Interviewer States and Paraphrases

1	Um, ¿dónde vives aquí en Middlebury?	
2	¿En qué residencia vives?	
	'Um, where do you live here in Middlebury?	
	In what dormitory do you live?'	
3		(Silence)
4	¿Vives en en Painter o en Star?	
	'Do you live in Painter or in Star?'	
5		Oh, um Stuart?
6	En Stuart uh huh. Y tu cuarto,	
7	¿es bueno o es malo?	
	'In Stuart uh huh. And your room,	
	is it good or is it bad?'	
8		Bueno. 'Good.'
9	Uh huh, ¿Y tienes una compañera de	
10	cuarto? Una amiga que vive contigo?	
	'Uh huh, And do you have a roommate?	
	A friend that lives with you?'	
11		Sí. 'Yes.'

The interviewer asks Leslie a question in Line 1 and immediately paraphrases in Line 2. When Leslie does not respond, the interviewer gives more vocabulary clues by naming several possible dorms in which the learner might live. The learner is thus able to make the necessary associations and answer the question. Again, in Lines 9 and 10 we find another instance of self-repetition. This technique of self-repetition by the interviewer is used 10 out of 18 times when a new topic is introduced in the interview with Leslie.

As already noted, as proficiency increases, self-repetitions decrease, but even at the higher levels negotiation still seems to deal primarily with vocabulary. Thus, at the NM and NH levels, the comprehension checks and clarification requests are principally concerned with specific words, as we see in the following example (Example 19):

Example 19. Marissa (NM): Negotiation of Vocabulary

1	Sí. ¿Cómo es tu cuarto? ¿Qué tienes en	
2	tu cuarto?	
	'Yes. What is your room like? What do	
	you have in your room?'	
3		¿Perdón? 'Pardon?'
4	¿En tu cuarto, en tu habitación?	
	'In your room, in your room?'	
5		Um hum.

6	¿Cómo es? 'What is it like?'	
7		Uh es pequeño. 'Uh it's small.'
8	Ah ha.	

In Line 3 Marissa asks for clarification of the interviewer's question. We cannot know for certain what part of the utterance Marissa did not understand, but the interviewer appears to assume that it was the vocabulary word for "room," since the interviewer repeats and then supplies a synonym for this word. Thus, Marissa's clarification request enabled her to get the important content word "room," which consequently allows her to correctly identify the topic.

Hatch summarizes as follows:

> It is tempting . . . to hypothesize that the adult focus in second language learning is on vocabulary . . . but, instead, it appears that the learner is only asking for enough vocabulary to allow him to nominate topics and participate in conversational discourse. Perhaps once the pressure of needing a particular word in order to take his turn in the conversation is gone, the vocabulary is also "forgotten." (p. 430)

The interview with Vince gives a striking example to support Hatch's suggestion that vocabulary is only retained long enough to establish the topic (Example 20):

Example 20. Vince (NH): Learner Does Not Retain Negotiated Vocabulary

1	Sí, y el niño, ¿le gusta leer, le gustan	
2	los deportes? 'Yes, and the boy, does he like to read, does he like sports?'	
3		Sí uh tiene un uh . . . uh uh bi
4		*bilote? *Bilota? 'Yes uh he has a uh . . . uh uh (invents words)'
5	¿Una pelota? ¿De béisbol? 'A ball? For baseball?'	
6		Sí, sí. 'Yes, yes.'
7	Sí, sí sí sí. Sí. 'Yes, yes,yes,yes. Yes.'	
8		Um . . . y um . . . cuántos uh . . . no um
9		qué es la uh *pricio? 'Um . . . and um . . . how many uh..no um what is the uh price?'
10	Ah, bueno, tenemos perdón, de dos	
11	precios; tenemos una pelota grande	
12	que se usa para el softbol.... 'Oh well, we have pardon two prices; we have a large ball that is used for softball'	
13		Sí, sí 'Yes, yes'
14	que cuesta ocho dólares y tenemos una	
15	pelota más pequeña para el béisbol que	
16	cuesta diez. 'that costs eight dollars and we have a	

	smaller ball for baseball that costs ten.'	
17		Ah...uh muy bien. Uh yo *quiera um uh
18		*pilote uh diez dólares pero puede . . . puede . . .
19		um puede . . . tiene un uh *caje para um uh el
20		*pilote?
		'Ah . . . uh very good. Uh I want um uh
		(invents word) uh ten dollars but can
		you . . . can you . . . um can you . . . do you have a
		uh box for um uh the (invents word)?'
21	Sí, sí sí sí. 'Yes, yes yes yes.'	

In Lines 3 and 4 Vince makes several attempts at producing the correct word for "ball" and requests a repair using rising intonation. The interviewer responds by supplying the correct form of the word that enables the conversation to continue, but as we see in Lines 18 and 20, Vince has modified his initial attempt at the word by adopting the correct initial consonant. He still does not use the correct word, however. An interesting parallel to this example comes later in the same interview when he is asking the interviewer a question (Example 21):

Example 21. Vince (NH): Use of Invented Word to Solicit Key Vocabulary

1		Te gusta um la biblio uh bi um
2		* bilote? Um um . . . biblio. . .l
		'Do you like um the libra uh li um
		(invents word)? Um um . . . libra . . .'
3	¿Biblioteca? 'Library?'	
4		No, no biblioteca. (Laughs) Uh um
5		*bilota? Um bicycle?
		'No, not library. (Laughs) Uh um
		(invents word)? Um bicycle?'
6	Bicicleta. 'Bicycle.'	
7	Sí. 'Yes.'	Bicicleta, bicicleta. Sí, sí,sí, sí.
		'Bicycle, Bicycle. Yes, yes, yes, yes.'

We find in this example that Vince uses the same word in an attempt to express "bicycle" as he used in Example 20 when he was attempting to express "ball." This usage is noteworthy because "bicycle" in English and Spanish are cognates, but instead of trying a cognate, which is often a natural reaction when negotiating for vocabulary between the two languages, Vince uses the word *bilote*. Except for the initial consonant, this word bears no resemblance to *bicicleta*. Thus, it seems, at least in these two examples, that Vince employs the word *bilote* on occasions when he is unsure of the correct vocabulary word in an attempt to solicit the proper word from his interlocutor, a tactic that was successful in both of the previous examples.

Hatch underscores the importance of negotiation for topic identification, stating, "Once the topic is recognized, [the learner] can further use his knowledge of the world and of discourse in his own language to predict the possible questions, the concerns which might be expressed in questions about that topic" (p. 423). Thus, without the ability to successfully identify the topic through negotiation, no conversation can take place. We find many instances where this happens in the data from all three levels of proficiency. At the NL level the learner frequently opts out of the conversation by remaining silent. We see this strategy in an example from Leslie (Example 22).

In Line 5, Leslie's silence causes the interlocutor to abandon the topic and to pose a new question. In Line 9, however, Leslie's silence does not impede the conversation to the same degree, because she has a sympathetic interlocutor who continues the process of negotiation even when Leslie tries to opt out.

Example 22. Leslie (NL): Learner Remains Silent

1	En comparación. Ah ha, ¿y um te gusta	
2	Middlebury? ¿Estás contenta en Middlebury?	
	'In comparison. Ah ha, and do you like Middlebury? Are you content at Middlebury?	
3		Sí. 'Yes'
4	¿Y por qué te gusta? 'And why do you like it?'	
5		(Silence)
6	Um, ¿cómo son las clases para ti?	
	'Um, how are the classes for you?'	
7		Muy bien? 'Very good?'
8	Um hum. ¿Y qué clases tienes?	
	'Um hum. And what classes do you have?	
9		(Silence)
10	Español 'Spanish'	
11		Um, geography

At the higher levels of proficiency, however, it is often the interviewer who does not participate in the negotiation, thus bringing the topic to an abrupt end as in Example 23:

Example 23. Marissa (NM): Interviewer Does Not Negotiate

1	¿Qué, dime qué llevas. ¿Qué ropa llevas?
2	¿Qué es esto?
	'What, tell me what you are wearing. What clothes are you wearing? What is this?'

3		Uh blue jeans, uh camisas 'Uh blue jeans, uh shirts'
4	Um hum. ¿Y eso? 'And that?'	
5		Um . . . sneakers
6	Pero esos no son sneakers. 'But those aren't sneakers.'	
7		Um pero como se dice uh
8		shoes? 'Um but how do you say shoes?'
9	Ok, sabes los colores, ¿Qué color es éste? 'Ok, you know the colors, What color is this?'	
10		Sue, suete 'Swea, sweate'
11	Um hum, ¿y el color? 'And the color?'	
12		Sí . . . uh es éste? 'Yes . . . uh is it this?'
13	Uh huh	
14		Es . . . rojo? 'It is . . . red?'

In Lines 7 and 8 Marissa attempts to negotiate for the word "shoe," which she does not know, and thus tries to keep the conversation going. The interviewer does not respond to her request, however, and instead changes the topic in Line 9. The learner, who is expecting a response to her request, is unprepared for the topic change and, in Line 10, she continues to respond to the interviewer's original topic of identifying the items of clothing.

6. Interviewer's Role

Examples such as the previous raise the question of the interviewer's role in negotiation during the interview. The OPI Training Manual (Buck, 1989) states that the OPI is "intended to replicate natural conversational behavior" (p. 5-15) and instructs interviewers to "perform negotiative moves when communication is not totally successful" (p. 5-2). These two stipulations would obligate the interviewer in Example 23 to respond to the learner's request for an unknown vocabulary word. Yet the manual also instructs interviewers not to furnish vocabulary and to assume the role of a monolingual speaker of the target language during the interview. These directions would prohibit the interviewer from responding to the learner's request in the above example. Thus, the interviewers find themselves in a somewhat contradictory position, which partially explains the wide variety of responses to this situation found in the data. There is variation even within the same interview as to how the interviewer responds when confronted with this type of request, and across the data it is obvious that certain interviewers are much more willing to supply the vocabulary that is being solicited. It is noteworthy that the interviews in which such requests are fulfilled resulted in much more authentic conversation and much smoother topic transitions, as can be seen in Example 24:

Example 24. Vince (NH): Interviewer Provides Requested Vocabulary

1	Uh, ¿Es es por eso que estudias	
2	español?	
	'Uh, Is is that why you study Spanish?	
3		Um, sí, sí. Uh en en Florida es um
4		uh . . . ¿cómo se dice mandatory?
		'Um, yes,yes. Uh in in Florida it is um uh . . . how do you say mandatory?'
5	Necesario. 'Necessary.'	
6		Oh, es necesario tener dos años de
7		español. 'Oh, it is necessary to have two years of Spanish.'
8	¿Ah sí? 'Oh yeah?'	
9		Sí, sí. Muy importante porque uh en
10		Jacksonville uh ellos no hablan español uh
11		muchas pero uh um va a sur de Jacksonville, uh
12		en Orlando, Miami habla español muy bie..uh
13		muchas.
		'Yes, yes. Very important because uh in Jacksonville uh they don't speak Spanish uh many but uh um go to south of Jacksonville, uh in Orlando, Miami speak Spanish very we . . . uh many.'
14	Um hum. Sí, sí. Bueno eso es interesante.	
	'Um hum. Yes, yes. Well that is interesting.'	

We see in this example that the interviewer's response to Vince's request for the word "mandatory" in Line 4 enables him to continue the conversation and express, although somewhat awkwardly, a long and complex utterance in Lines 9 through 13. On the contrary, if the interviewer had denied Vince's request and switched the topic as in Example 23, the conversation would have come to a halt and Vince would not have had the opportunity to demonstrate the knowledge he does in Example 24.

It seems, therefore, that if the intention of the OPI is to replicate genuine conversation as nearly as possible, interviewers are obliged by rules of polite conversation to respond to such requests. In the OPI setting, the learners are aware that the interviewer speaks both the target language and the L1. Thus, as the pressure to communicate increases, the learners can be expected to take recourse to any and all tools at their disposal to keep the conversation moving, and they will expect the interviewer to do the same. Any reluctance on the part of the interviewer to cooperate in achieving mutual understanding will be seen as a violation of the rules of interaction and will thus detract from the sense of the interview as a genuine conversational exchange. Byrnes (1987) argues that this cooperation is crucial if the

interview is to be an accurate measure of the learner's pragmatic competence. As she explains,

> With a keen awareness on the part of the tester of what constitutes natural conversational behavior, including natural listening behavior, curbing it only in so far as it would preclude an optimal and efficient sampling of speech on the part of the candidate, I have observed over and over again how the dynamics of message transmission become so powerful in themselves that the candidate can present an amazingly multi-faceted performance, even at the lower level of language use. . . . Conversely, when the tester does not bring to the task that level of interpersonal involvement and willingness to become a true conversational partner within the constraints of the event, then little of the kinds of evidence for pragmatic competence I have mentioned can surface. (pp. 174-75)

Furthermore, even in instances in which the learner is truly interacting with a monolingual speaker of the target language, offering key vocabulary in the L1 if the learner does not know the necessary word in the target language is a valid strategy that can often lead to a successful interaction if the interlocutor can understand the word from the context or recognize it as a cognate.

CONCLUSION

This study has examined several aspects of Novice-Level discourse in an attempt to describe some of the distinguishing patterns and to gain a better understanding of the language being produced by the majority of students in foreign language classrooms today. A look at the cohesive devices used by these learners shows that, despite limited linguistic resources, Novice-Level speakers do attempt to produce cohesive discourse, and, as proficiency increases, so does the use of cohesive devices. At the NL level, for example, referential cohesive devices were not used at all, while at the NH level there were 32 instances of pronoun usage. The data also showed that learners did not readily transfer their L1 rules regarding pronoun usage, but, from an early stage, they were able to exploit the pro-drop aspect of Spanish to create cohesive discourse through topic continuity. The NM level was characterized by the greatest use of pro-drop. This characteristic is seen perhaps because, at the NH level, learners are beginning to construct and implement their L2 syntactic systems and are more likely to supply pronouns for clarification, or because they are using less imitative or memorized speech, and they fall back on L1 rules to help them construct original utterances in the L2.

The data also show that, as proficiency increases, learners achieve greater cohesion in their discourse by attempting to use more conjunctions and other lexical items such as *también* 'also.' Again, the use of these devices was most evident at the NH level because learners at the lower levels of proficiency rarely produced utterances of sufficient complexity to include conjunctions or other lexical devices. Even at the NH level, however, the discourse was often redundant because beginning learners have limited linguistic

tools at their disposal to create cohesion. Well-timed instructional emphasis at the NH level on a wider variety of features that learners might incorporate into their oral production could be effective in helping learners achieve greater cohesion in their discourse.

Scaffolding and negotiation were found to be two important aspects of Novice-Level discourse. Only at the NM and NH levels did the learners make any attempts to initiate the topic or take control of the dialogue, and on the few occasions when this did occur, the learners were hesitant and uncomfortable in this role. It would be valuable to implement tasks in the foreign language classroom that would force learners to direct the conversation so that they could gain practice in this important discourse skill as well.

The four interactional features that were examined in the study were clarification requests, comprehension checks, repetition, and expansions. The data support Ellis' finding that the occurrence of these features varies with level of proficiency. Thus, self-repetitions decreased as proficiency increased, while clarification requests and comprehension checks increased with proficiency. The data show that the use of these features at the Novice level is crucial to building the vertical constructions that allow for exchanges of meaning. Consequently, negotiation at the Novice level is principally concerned not with issues of grammar, but rather with the vocabulary necessary to establish the topic.

There are times in the data, however, where negotiation fails to occur. Often this happens because the learner uses the L1 or opts out of the conversation by remaining silent. Classroom discussion concerning the importance of negotiation and explicit instruction in various negotiation techniques could help learners become more adept at managing situations in which they encounter L2 breakdown and improve their chances for successful communication. The interviewer's willingness to direct the dialogue and help out when problems arise was also found to be crucial to successful interaction. When the interviewer is not willing to cooperate in the negotiation, the possibility of creating genuine conversation is limited, and the learners are less able to demonstrate their full potential in the target language.

From these findings, it is now possible to revise the Novice-Level ACTFL proficiency guidelines to include discourse features. I would propose that the following descriptions be added to the ACTFL guidelines:

Proposed Additions to ACTFL Novice Oral Guidelines

Novice-Low: Learner does not attempt to initiate topic or control conversation and frequently opts out of conversation by remaining silent. Discourse rarely includes cohesive devices. Interaction at this level is characterized by much

self-repetition and repetition of the other and the rare use of clarification requests or comprehension checks. The L1 is used frequently to repair and elaborate and in times of L2 breakdown.

Novice-Mid: Learner makes occasional hesitant attempts to initiate topic. Negotiation is primarily concerned with vocabulary. Clarification requests and comprehension checks are common. Few referential and conjunctional cohesive devices are used, but learner demonstrates a good understanding of pro-drop.

Novice-High: Learner makes more attempts to control dialogue and nominate topic, but negotiation for key vocabulary is critical for success. Interaction is characterized by fewer repetitions and many clarification requests and comprehension checks. Referential, conjunctional, and a few lexical cohesive devices are used more frequently, but the learner's repertoire is limited. Pronouns are frequently included.

This study has focused primarily on discourse aspects of Novice-Level oral production. While descriptions of these discourse features are an important first step, there is still a need for additional studies that examine other aspects of Novice-Level speech, such as grammar, phonology, and syntax. Such studies are crucial if we are to develop level-appropriate testing and teaching techniques. This endeavor should be a priority in the research on oral proficiency so that we can better serve the needs and abilities of all foreign language learners.

REFERENCES

American Council on the Teaching of Foreign Languages. (1986). ACTFL proficiency guidelines. In H. Byrnes & M. Canale (Eds.), *Defining and developing proficiency: Guidelines, implementations and concepts* (pp. 123-124). The ACTFL Foreign Language Education Series. Lincolnwood, IL: National Textbook Co.

Bearden, R. (1997). *Discourse features of beginning learners' oral production: What can we expect?* Unpublished manuscript, University of Texas at Austin.

Buck, K. (Ed.). (1989). *ACTFL oral proficiency tester training manual.* Yonkers, NY: American Council on the Teaching of Foreign Languages.

Bragger, J. (1985). Materials development for the proficiency-oriented classroom. In C. James (Ed.), *Foreign language proficiency in the classroom and beyond* (pp. 79-122). The ACTFL Foreign Language Education Series. Lincolnwood, IL: National Textbook Co.

Byrnes, H. (1987). Features of pragmatic and sociolinguistic competence in the oral proficiency interview. In A. Valdman (Ed.), *Proceedings of the Symposium on the Evaluation of Foreign Language Proficiency* (pp. 167-177). Bloomington, IN: University of Indiana.

Ellis, R. (1985). Teacher-pupil interaction in second language development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 69-85). Cambridge, MA: Newbury House.

Gass, S., & Varonis, E. (1994). Input, interaction and second language production. *Studies in Second Language Acquisition, 16* (3), 283-302.

Hatch, E. (1978). Discourse analysis and second language acquisition. In E. Marcussen Hatch (Ed.), *Second language acquisition: A book of readings* (pp. 401-436). Rowley, MA: Newbury House.

Heilenman, L., & Kaplan, I. (1985). Proficiency in practice: The foreign language curriculum. In C. James (Ed.), *Foreign language proficiency in the classroom and beyond* (pp. 55-78). The ACTFL Foreign Language Education Series. Lincolnwood, IL: National Textbook Co.

Higgs, T., & Clifford, R. (1982). The push toward communication. In T. Higgs (Ed.), *Curriculum, competence, and the foreign language teacher* (pp. 57-80). Skokie, IL: National Textbook Co.

Omaggio, A. (1986). *Teaching languages in context*. Boston: Heinle & Heinle.

Pavlou, P. (1994). *The role of cohesion in communicative competence as exemplified in oral proficiency testing*. Paper presented at the Language Testing Research Colloquium, Washington, D.C.

Perdue, C., & Klein, W. (1992). Why does the production of some learners not grammaticalize? *Studies in Second Language Acquisition, 14* (3), 259-272.

Pica, T. (1987). Second language acquisition, social interaction, and the classroom, *Applied Linguistics, 8* (1), 3-21.

Pica, T. (1994). Research on negotiation: What does it reveal about second language learning conditions, processes, and outcomes? *Language Learning, 44* (3), 493-527.

Platt, E., & Brooks, F. (1994). The "acquisition-rich environment" revisited. *Modern Language Journal, 78* (4), 497-511.

Valdman, A., & Phillips, J. (1977). Pidginization, creolization, and elaboration of learner systems. *Studies in Second Language Acquisition, 1* (1), 21-40.

A Discourse Approach to the Assessment of Foreign Language Oral Proficiency

DALE APRIL KOIKE, The University of Texas at Austin
FANNY HINOJOSA, Universidad Autónoma de Nuevo León, Monterrey

This study examines the varying degrees to which learners of Spanish at the Intermediate and Advanced levels use paragraph-like structure in their answers during a simulated oral proficiency examination. We compare the structure of the data to a model template that reflects the organization of most Advanced-level learners for this task. An analysis reveals that the Advanced learners in the study provide more propositions and supporting statements to complete the paradigm, closely following the proposed model. The Intermediate testees provide fewer propositions and supporting statements and either do not complete the paradigm or begin to diverge from their original intent. The less proficient speakers usually work at the level of the single proposition with single supporting statement, and they use other strategies to attempt to compensate for their deficiencies. We propose that the template approach be used in the assessment of foreign language oral proficiency in the classroom, since it can reveal more of the global expression of ideas, as opposed to the evaluation of discrete items such as verb conjugation and agreement errors.

INTRODUCTION

The focus on communicative activities in the foreign language classroom that allow learners to use the target language in meaningful, interactive, and engaging ways has made a profound impact on the way classes are taught, on expectations on the part of teachers and learners, on the content and format of textbooks, and on the organization of class time and environment. It has inspired a change in the way we assess language proficiency in a global sense, seen for example in the American Council on the Teaching of Foreign Languages (ACTFL) Oral Proficiency Interview. What is still lacking is a reliable, valid, and time- and cost-efficient form of testing that can be used in the foreign language classroom and that truly reflects the communicative goals that are set out for the learners.

A proficiency test is not particularly useful for course assessment because it is not bound to the curriculum and does not "cover a pre-established body of content assigned beforehand" (Higgs, 1987, p. 284). Its purpose is to measure a testee's performance against what a native speaker might be expected to say in the same situation. Higgs states that the proficiency test is not an appropriate tool for rank ordering a group of students, since the ratings represent *ranges* of performance and cannot be used productively to make fine distinctions seen in grading systems in this country. The interviewer notes *patterns* of strengths and weaknesses instead of individual right or wrong answers.

This type of test format is not intended to examine discrete points, but rather integrative efforts to express a meaningful whole message.

Teachers can also use an achievement test, which covers an explicit, finite amount of content and is intended to rank the students who take the exam. The typical achievement test tends to target discrete points of grammar and vocabulary, but the problem, as Higgs states, is that "the emphasis in achievement tests is almost always exclusively on form rather than function, on being correct rather than having something interesting, important, or true to say" (p. 283). Certainly it is the goal of the classroom test to cover what is taught. In a skill-focused class such as an advanced conversation course, what is taught is a number of ways in which learners can improve their fluency in the speaking skill, whether the improvement be in terms of vocabulary, grammar, organization, or simply increased opportunity to speak in the target language. The ultimate goal is an increase in overall proficiency, which is difficult to pinpoint in terms of an achievement test.

Another option is the "prochievement test," a hybrid test that elicits discrete points of grammar and vocabulary within the context of some kind of oral interview format. It is our belief that such a test, as currently defined, is not the answer for the conversation course. The prochievement test, at least as described by Hendrickson (1992), is still achievement oriented in that it is not aimed at assessing a body of discourse in a holistic way, since the focus is still on whether or not a testee uses a given set of items in responding to a stimulus.

We will describe in this study a method of assessment of ongoing development for the upper-level Spanish conversation course, a course in which the focus is the organization of learner's discourse. We will show that, by looking at the structure of the discourse, one can see the learner's production in a more constructive, top-down, performance-based way, and the more discrete points of grammar are contextualized in meaningful communication.

To this end, we present our research questions, data from various testees at the Intermediate and Advanced levels, and an analysis of these data. We conclude with some suggestions for oral proficiency assessment in the advanced conversation class.

RESEARCH QUESTIONS

The questions addressed in this study are the following:

1. What differences are seen in discourse structure produced by various learners at Intermediate and Advanced levels of Spanish proficiency on a simulated oral proficiency exam?

2. What is the applicability of a structure-focused oral proficiency exam for classroom purposes?

METHODOLOGY OF DATA COLLECTION

To answer these questions, data were collected from an oral proficiency test used in an advanced Spanish conversation course. The

following is a description of the test instrument, the testees, and the data collection and analysis procedure.

The Test Instrument

The data for this study are drawn from a practice simulated oral proficiency exam administered in a language laboratory at the end of a semester-long advanced Spanish conversation course aimed at preparing potential Spanish language teachers of the elementary and high school levels to take the Spanish Texas Oral Proficiency Test (TOPT). This test is required of all prospective elementary and secondary Spanish and bilingual education teachers in the state of Texas, and they must be rated according to their performance on the test at the Advanced level to pass and receive certification to teach. The practice tests, given in the Spanish-American variety of Spanish, were created to simulate as closely as possible the content, functions, and format of the TOPT (Stansfield, 1993; Stansfield & Kenyon, 1992). A difference between the practice and live TOPT that should be noted is that, on the practice tests, testees are usually given 2 minutes to answer each question, approximately 30 seconds more than on the typical Advanced-level question.

This study focuses on one stimulus that presumably requires more linguistic skill than others and is considered to be of the Advanced level: specifically, that of comparing advantages and disadvantages. Successful performance on this task is considered to be indicative of the Advanced level of proficiency according to the ACTFL Guidelines. The actual stimulus used is given below:

Stimulus:

> While talking with some of your friends in Mexico, one of them says that people should use bicycles as their major means of transportation, while another friend says that motorcycles would solve many problems. Then the others present turn to you. *Present your position logically by comparing advantages and disadvantages* of the bicycle/motorcycle issue. (20-second pause) (Prompt: *¿Y tú? ¿Qué crees?* 'And what do you think?')

As in the original version of the TOPT, testees heard the test stimuli in English over a headset in individual language laboratory booths and recorded their answers in Spanish onto individual audio cassette tapes. The tapes were graded by the instructor for features such as fluency, grammar, vocabulary, and appropriateness of language, but mainly for organization, as part of the final grade for the course.

Testees

Since the 24 testees were students enrolled in a course designed to prepare them to take the TOPT, one can assume that, at the point in the course at which data were collected, these testees were more skilled at taking this kind of test than the average testee. All the testees in this study were estimated to be at least at the Intermediate level of oral proficiency, according to the ACTFL Guidelines (1997). Each

learner was screened for an approximate proficiency level and other background factors (for example, true intent to take the TOPT) by the instructor before being allowed to register for the course. The group included six Hispanic Americans, two of whom were estimated to be at the Advanced level.

The course these learners were enrolled in focused on helping them prepare for the TOPT by working on rhetorical structure and organizational skills, and, most of all, by practicing the language functions and the format of the test. The learners typically were very motivated to work hard in the class and took the testing procedures seriously since they needed to pass the TOPT in order to qualify for a teaching job.

Data Collection and Analysis

In order to look at the organization of the discourse produced by the testees, we compared their responses to two models of how a logical, top-down organizational structure of propositions and supporting statements might appear in the task of comparing advantages and disadvantages. These models were compiled from our overall impression of what many native speakers of Spanish do to complete this task, and what most Advanced and Advanced High speakers appear to be attempting to do when they answer under these test conditions.

Table 1 illustrates both of the two models. Table 1A shows a listing strategy, by which the testee lists several advantages and disadvantages of a given referent before doing the same for the second referent. The supportive statements in parentheses indicate possible examples and other elaboration for those points. Table 1B provides a contrasting strategy, by which the testee gives the advantages of a referent and follows with a contrastive statement of the disadvantages of the other referent. Then the testee gives the advantages of the second referent and follows with another contrastive statement of the disadvantages of the first referent. In both models the paradigm of giving both advantages and disadvantages is completed, followed by a statement of the testee's opinion as to which referent is perceived as better. One can assume that the more the testees complete the paradigm in a logical progression, give points of comparison or contrast, and support those points with examples and other elaboration, the more complete and well presented the discourse will be. One can also assume that the ability to do this is linked to proficiency in the target language.

The transcription of each testee's response was analyzed with respect to the presentation of a topic sentence, advantages and disadvantages with supporting statements, and other elaboration, as well as any concluding remarks. Each of the investigators completed her own analysis of the data, which was then compared to the other's. Discrepancies were discussed after all data were analyzed.

RESULTS

We begin our presentation of the results with sample replies from three testees who illustrate different levels of responses, followed by a brief discussion of what we found in

Table 1
Models of Propositional Organization for Comparing
Advantages and Disadvantages

Model 1A - Listing Strategy

Topic sentence: Being a woman has advantages and disadvantages over being a man

Advantage 1 of X: Women are more intelligent
Advantage 2 of X: Women are more sensitive

Disadvantage 1 of X: Women are not as physically strong
(Support: You don't see as many women in professions requiring physical strength)
Disadvantage 2 of X:

Advantage 1 of Y: Men are strong and tend to be more assertive.
Advantage 2 of Y: Men are more independent.
(Support: They get to do more because they don't feel so dependent on others to act)

Disadvantage 1 of Y: Men are not as intelligent.
Disadvantage 2 of Y:

Conclusion: It is more advantageous to be a woman.

Model 1B - Contrastive Strategy

Topic sentence: Being a woman has advantages and disadvantages over being a man

Advantage 1 of X: Women are more intelligent
Advantage 2 of X: Women are more sensitive

Disadvantage 1 of Y: Men are insensitive
Disadvantage 2 of Y: Men think they know everything

Advantage 1 of Y: Men are physically stronger
(Support: Men can lift heavy things)
Advantage 2 of Y: Men are more independent

Disadvantage 1 of X: Women are not as physically strong
Disadvantage 2 of X: Women take too long to get ready in the morning

Conclusion: Still, it's more advantageous to be a woman than a man.

them as compared to our models. We then discuss categories seen in the corpus as a whole and conclude with some implications for classroom use.

According to the structural quality of responses to the rather difficult test item, the testees were divided into three general groups: very complete answers, fairly complete, and skeletal, with discrepancies noted within each of the groups. An example of a very complete answer is that of Grace, as shown in Example 1:

Example 1. Grace's Reply: Comparing Advantages and Disadvantages [2]:

1 Grace: Pienso que, usando una bicicleta o una motocicleta los dos tienen, aventajas y desventajas.
2 Una mayor, aventaja para la bicicleta pienso es que, la bicicleta es muy bien para el ambiente.
3 El bicicleta no, no sale, no: saca humo ni nada y es, eso es muy bien. Y pienso que también la bici-
4 cleta es bien para, para hacer ejer≠ciosØ. (Uh, una: aventaja para la motocicleta, será, sería
5 que, ((sniff)) es muy pequeña, y no, nece—no: (0.5) tiene, no es tan grande como un carro. Pero una
6 desventaja de: la motocicleta es que no es bien para el ambiente. Y las motocicletas hacen mucho
7 ruido. Pienso yo que, la mejor idea sería usar una bicicleta porque, la bicicleta >pienso que
8 tiene< más, aventajas, que la, que la motocicleta.>La bicicleta te puede dar< ejercicios, (0.5) te
9 puede llevar a muchos lugares una, una desventaja de la bicicleta a lo mejor sería que, el tiempo.
10 Que en una moticicleta llegarías en un lugar como en diez minutos, pero por bicicleta a lo mejor,
11 (0.5) neces—vas a necesitar más tiempo, para llegar adonde necesitas llegar y, a lo mejor si hace
12 mucho calor afuera, vas a sudar y, a lo mejor eso no es, no es una manera de, de, transportación
13 para Ud. So: pienso que a lo mejor la motocicleta será, una: (0.3) teng—tenga una aventaja para
14 las personas que necesitan llegar a lugares, con aprisa o no, o no necesita-o no tienen que, no
15 tienen el tiempo para, para gastar en una bicicleta.

'I think that, using a bicycle or a motorcycle the two have, advantages and disadvantages. A greater, advantage for the bicycle I think is that, the bicycle is very fine for the environment. The bicycle doesn't, doesn't leave, doesn't let out smoke or anything and it is, that is very fine. I think that also the bicycle is fine for, for doing exercise. Uh, an, advantage for the motorcycle, will be, would be, is it's very small, and doesn't nee-doesn't have, isn't as big as a car. But a disadvantage of, the motorcycle is that it isn't fine for the environment.

And motorcycles make a lot of noise. I think that, the best idea would be to use a bicycle because, the bicycle I think that it has more, advantages, than the, than the motorcycle. The bicycle can give you exercises, can take you to many places a, a disadvantage of the bicycle probably would be that, the time. That on a motorcycle you would arrive to a place like in ten minutes, but by bicycle probably, you nee—you are going to need more time, to arrive where you need to arrive and, probably if it's very hot outside, you're going to sweat and, probably that isn't, isn't a means of of transportation for you. So, I think that probably the motorcycle will be, a, may—may have an advantage for the people who need to get to places, hurriedly or not, or don't nee—or don't have to, don't have the time to, to waste on a bicycle.' [3]

Table 2 shows a structural analysis of Grace's response. Numbers represent the order in which the main propositions appeared, while the words in parentheses represent elaborations or supporting statements. Grace gave a general, organizing statement at the beginning of her reply, stated several advantages and disadvantages for both bicycles and motorcycles, contrasting positive and negative points, and provided several supporting statements. She was also able to make a summarizing statement in her final sentence (Lines 13-15). Grace's discourse approximates the model given in Table 1B (Contrastive Strategy) for comparing advantages and disadvantages.

Table 2
Grace's Propositional Organization for Comparing Advantages and Disadvantages

Topic sentence: *Los dos tienen aventajas y desventajas* 'Both have advantages and disadvantages'

Advantage 1 of bicycles: *No saca humo* 'It doesn't let off smoke'

Advantage 2 of bicycles: *Es bien para hacer ejercicios* 'It's fine for doing exercise'

Advantage 1 of motorcycles: *Es muy pequeña (No es tan grande como un carro)* 'It's very small (It isn't as big as a car)'

Disadvantage 1 of motorcycles: *No es bien para el ambiente* 'It is not fine for the environment'

Disadvantage 2 of motorcycles: *Hacen mucho ruido.* 'They make a lot of noise'

(Continued)

Table 2 *(Continued)*

Topic sentence 2: *La bicicleta tiene más aventajas que la motocicleta* 'The bicycle has more advantages than the motorcycle'

Advantage 3 of bicycles: *Te puede llevar a muchos lugares* 'It can take you to many places'

Disadvantage 1 of bicycles: *El tiempo. (Vas a necesitar más tiempo)* 'The time (You are going to need more time)'
Disadvantage 2 of bicycles: *Vas a sudar.* 'You're going to sweat'

Conclusion: *Motocicleta tenga una aventaja para las personas que necesitan llegar a lugares con aprisa* 'Motorcycle has an advantage for the people who need to get to places in a hurry'

Let's look now at Example 2, the response by a student named Mary who provided a good reply, reflective of the "fairly complete" group, but one not as tightly structured as that of Grace's:

Example 2. Mary's Reply for Comparing Advantages and Disadvantages:

Mary: Ento:nces, uh, >mis amigos<, creo que, uh, la bicicleta es, es la solución, uh, mejor, para el problema de viajar. Porque, uh, creo que la bicicleta? Uh, no da, uh más palus–, polución, a la ciudad, en México, y, um el, la, motocicleta, sí, um da, la, polución, um, el. Aunque la bi—la motocicleta es, es más rápido, um, para viajar, la (bicicleta (no requiere, uh, no requiere tiempo o dinero, o demasiado tiempo o dinero para arreglar y para segurar que, uh, está funcionada, uh, todo el tiempo. Um (0.5) uh, (0.5) m-me gusta la bicicleta más, uh, >creo que< la bicicleta uh, es, la, uh, es la selección, uh, mejor para una ciudad como ésa? Uh. Y, um, (2 sec.) Y uh, necesitamos pensar en que, uh, en que más razones para eso. Um, ((laugh)) uh, la polución y la arreglación, um, y uh. También la bicicleta da más ejercicio, um, a la persona que monta? Que lo monta? La monta. Um, y la motocicleta no da, uh, ejercicio porque en la motocicleta, uh, solamente uh, solamente: una persona se sienta, en eso.

'So, uh, my friends, I think that, uh, the bicycle is, is the, uh, better, solution, for the problem of traveling. Because, uh, I think that the bicycle? Uh, doesn't give, uh, more pollu—pollution, to the city, in Mexico, and um the, the motorcycle, does, um, give, the, pollution, um, the. Although the bi—the motorcycle is, is faster, um, for traveling, the bicycle, doesn't require, uh, doesn't require time or money, or too much time or money to fix and to be sure that, uh, it is working, uh, all the time. Um, uh, I—I like the bicycle more, uh, I think that, the bicycle uh, is, the, uh, is the best, uh, option for a city like this one? Uh. And um. And uh, we need to think about, uh, about more reasons for this. Um, (laughter), uh, the pollution and the fixing, um, and uh. Also the bicycle gives more exercise, um, to the person who rides? Who rides it? Rides it. Um, and the motorcycle doesn't give, uh, exercise because on the motorcycle, uh, just, uh, a person just sits, on it.'

Mary's sample, formulated more in terms of stating and supporting an opinion than listing advantages and disadvantages, shows much repetition, very little elaboration, no disadvantages for bicycles, and no concluding statement. The lack of a concluding statement, however, was not uncommon in the responses. Because the testees had only 2 minutes to speak and were cut off when the time limit was reached, and because most were not watching a clock while speaking, but instead were concentrating on their responses, few actually produced a complete "paragraph" in terms of having a beginning and an end.

If one looks at the way in which Mary changed the task to that of giving and supporting an opinion, however, it is evident that she was aware that she needed to support her opinion with reasons (Lines 7-11). In other words, she demonstrated her knowledge of an organizational structure, albeit for a different task, but had trouble filling it in. Thus, Mary's paradigm is not that of our proposed model for comparing advantages and disadvantages. She performed well in the organizational structure of her points, but according to a different model with a different function, that of stating and supporting an opinion, as shown in Table 3.

Table 3
Mary's Propositional Organization for Comparing Advantages and Disadvantages

Topic sentence: *La bicicleta es la solución mejor para el problema de viajar*
'The bicycle is the best solution for the problem of traveling'

(Continued)

Table 3 *(Continued)*

Advantage 1 of bicycle: *La bicicleta no da más polución* 'The bicycle doesn't give more pollution'

Disadvantage 1 of motorcycle: *La motocicleta sí da la polución* 'The motorcycle does give pollution'

Advantage 1 of motorcycle: *La motocicleta es más rápido* 'The motorcycle is faster'

Advantage 2 of bicycle: *La bicicleta no requiere tiempo o dinero* 'The bicycle doesn't require time or money'

Conclusion (?): *Me gusta la bicicleta más* 'I like the bicycle more'

--*Necesitamos más razones* 'We need more reasons'

Advantage 3 of bicycle: *La bicicleta da más ejercicio* 'The bicycle gives more exercise'

Disadvantage 2 of motorcycle: *La motocicleta no da ejercicio* 'The motorcycle doesn't give exercise'*(Una persona solamente se sienta)* ('A person just sits')

Conclusion: ---

The response from Deb (Example 3), who is a much less proficient speaker than the other two, illustrates a much more fragmented and less cohesive structure, typical of the "skeletal group," as seen in Table 4.

Deb's reply for comparing advantages and disadvantages begins with two general statements that were probably intended to contrast with each other, but do not do so (Lines 1-3). They seemingly organize

Example 3. Deb's Reply for Comparing Advantages and Disadvantages

1 Deb: Bueno (1.0). Hay dos, dos cosas. Uno es que, que la gente debe montar, a una: bicicleta, para:

2 para: trans—la transporta≠ción,Ø uh, y: otro es, los motocicletas son mejores, o mejor manera,

3 una me—jor, una mejora manera, de, de, de (0.5) transpor—te, de transportación. Y, y creo que,

4 que, los, las bicicletas son más sanas para el ambiente. Y, y: no hacen, humo, ni nada en en el
5 aire. Y, eso es una, buen característica porque: estamos en Mej—en la ciudad de México, donde: el
6 humo y todo de: de la industria es, (0.5) uh, malísima. Y: los motocicletas, uh, tienen, tienen,
7 ventajas? Y desventajas también. Las ventajas son que: que es una: (0.2) un, un tipo, una tipa de
8 transportación más rápida, y toda la gente puede: puede uh, (0.5) movar o, o, (0.5) sí, más
9 rápida. Más rápido. Y, la, la otra cosa es que: la gente no suda, cuando, cuando monta una moto-
10 cicleta. En, en una bicicleta sí, todos sudan. Y ésa es una, una, buen, buena aventaja, para todos
11 los que trabajen.

'Well. There are two, two things. One is that, that people should ride, a bicycle, for, for, trans—transportation, uh, and, the other is, motorcycles are better, or a better means, a better, a better means, of, of, of, transport, of transportation. And, and I think that, that, the, the bicycles are healthier for the environment. And, and, they don't make, smoke, or anything in the air. And, that is a, good characteristic because, we are in Mex-, in Mexico City, where, the smoke and all of, of the industry is, uh, really bad. And, the motorcycles, uh, have, have, advantages? And disadvantages too. The advantages are that, that it is a, a type, a type of faster transportation, and all the people can, can uh, move or, or, yes, more quickly. More quickly. And the, the other thing is that, people don't sweat, when, when they ride a motorcycle. On, on a bicycle yes, everyone sweats. And that is a, a good, good advantage, for all of those who work.'

Table 4
Deb's Propositional Organization for Comparing Advantages and Disadvantages

Topic sentence: *Hay dos cosas: (a) la gente debe montar una bicicleta para transportación; (b) los motocicletas son mejor manera de transportación* 'There are two things: (a) people should ride a bicycle for transportation; (b) motorcycles are a better means of transportation'

Advantage 1 of bicycles: *Son más sanas para el ambiente* 'They are healthier for the environment'
(No hacen humo) ('They don't make smoke')
(Es una buen característica porque estamos en la ciudad de México) ('It is a good characteristic because we are in Mexico City')

(Continued)

Table 4 *(Continued)*

Topic sentence 2: *Los motocicletas tienen ventajas y desventajas también.* 'Motorcycles have advantages and disadvantages also'

Advantage 1 of motorcycles: *Una tipa de transportación más rápida.* 'A kind of faster transportation'
(La gente puede movar más rápida) ('People can move more quickly')
(La gente no suda) ('People don't sweat')

Disadvantage 1 of bicycles: *Todos sudan* 'Everyone sweats'

[Referring to Advantage 1 of motorcyles above] *Esa una buena aventaja para todos los que trabajen.* 'It is a good advantage for all those who work'

Conclusion: ---

the answer by stating the two points she wants to elaborate. She does give details on why people should ride bicycles (Lines 3-6), but instead of elaborating on why motorcycles are a better means of transportation, she goes on to say that they have advantages and disadvantages (Lines 6-7). She only states the advantages, however. She shows a notion of structure in her reply, but the content of her answer, what she actually says in terms of propositions, interferes with the unity of the sample, making it difficult to follow. In addition to her structural problems, she reveals a definite lack of key vocabulary and shows problems with agreement, which she self-corrects. All of these problems contribute to her difficulty in expressing many propositions and supporting statements (see Table 4).

Thus, the three examples reflect distinct levels of structural quality of answers in our data: complete, fairly complete, and skeletal. The groups are especially distinguishable when matched against an ideal model that represents a complete and coherent reply and a top-down organization, as seen in Tables 1A and 1B.

After analyzing each response for the various elements of a topic sentence, propositions of advantages and disadvantages, and supporting elaboration, as well as checking for a change of the task from that of comparing advantages and disadvantages to that of giving and supporting an opinion, we grouped samples according to patterns that emerged. What we found was a continuum of structural differences that could be characterized as shown in Table 5. These descriptions, still in various stages of development, are based on the models 1A and 1B we proposed earlier, and show where learners diverge from them. We should note that these categories should not be seen as tied to the ACTFL Guide-

lines. We have grouped them into three classes, labeled High, Mid, and Low, with three subcategories within each. On one end of the continuum we see a very complete response, including topic sentence, advantages and disadvantages of both referents, and a conclusion, all presented in a logical manner and approximating one of our two models. Then, at the High- level we see a topic sentence but fewer propositions than in the High+ and High-Mid levels. The testee may also have changed the topic sentence from a statement that there are advantages and disadvantages to a statement of an opinion about which one is better (for example, *Hay muchas ventajas y desventajas, pero yo creo que la bicicleta es mejor por varias razones* 'There are many advantages and disadvantages, but I think that the bicycle is better for various reasons'). The response, however, includes advantages and disadvantages of both referents.

In the next two levels the testees commonly change the task from that of comparing advantages and disadvantages to that of giving and supporting an opinion. This change in task is understandable if one considers that, by stating a preference for one referent, such as bicycles, the testee can avoid having to account for the advantages and disadvantages of both referents, thus simplifying the task. In the Mid+ level, the task is changed and the testees state propositions actually supporting only one side of an opinion. The Mid-Mid testees give fewer advantages and disadvantages, or only present either advantages or disadvantages. There is a notable lack of development in the ideas, so that the testee seems to

Table 5
Characteristics of the Three Proficiency Levels and Their Subcategories for Comparing Advantages and Disadvantages

	<u>High (More Complete)</u>
High+	Has topic sentence, both advantages and disadvantages of each referent, and a conclusion, in a coherent and cohesive organization, closely following and completing an organizational model (as in 1A or 1B).
High-Mid	Has topic sentence and several advantages and disadvantages, in a coherent presentation of ideas, following a partially complete organizational model.
High-	Has topic sentence and some advantages and disadvantages are given. One or two propositions may not fit tightly in a coherent

(Continued)

Table 5 *(Continued)*

	presentation of ideas. Possible change of task from comparison of advantages and disadvantages to stating and supporting an opinion.
	Mid (Fairly Complete)
Mid+	Has topic sentence, but changes task to that of stating and supporting an opinion. Several propositions are given that really support only one side of the argument.
Mid-Mid	Has topic sentence, but changes task to that of supporting an opinion. Response not complete, either with too few propositions expressing advantages and disadvantages or only a presenting a few advantages or disadvantages. Lacks development of response either because of too much elaboration on only one or two points or extraneous material inserted that does not directly support the reply.
Mid-	Has topic sentence but changes task to an opinion, with some advantages and disadvantages that are not logical or consistent in their support of the opinion.
	Low (More Skeletal)
Low+	Has topic sentence, but there are some incoherencies in the sample, either in the statements of the advantages and disadvantages or in the development of the response.
Low-Mid	Has topic sentence, but answer is incomplete, with few advantages and disadvantages. Some statements are incoherent and difficult to follow in logic. Language difficulties in vocabulary and grammar notably affect the ability to answer.
Low-	May or may not have topic sentence. Answer is incomplete due to difficulties in vocabulary and grammar, with perhaps only one advantage and/or disadvantage.

be simply listing. The Mid- level gives some advantages and/or disadvantages that do not support the topic sentence well, so one sees here the beginning of some problems of logic and consistency.

Problems of logic and consistency are seen notably in the Low level of testees. Difficulties with the

vocabulary and grammar definitely affect the testees' ability to answer the stimuli at all. Some statements are incoherent, and the logic of the propositions is difficult to follow. At the low end of the continuum, the response may or may not have a topic sentence, and perhaps only one or two propositions are stated in the 2 minutes allotted.

Of the 24 included in this study, the testees were divided as shown in Table 6. Most of the testees placed in the Mid level, and relatively few in the Low level.

IMPLICATIONS FOR TEACHING

Grading oral production is a problem for all levels of foreign language teachers, especially at the more advanced levels, where we often encounter learners who have different backgrounds and prior experiences with the target language. In a conversation course, the teacher faces the problem of grading the learners' achievement for that course and not their level of proficiency, although the ultimate goal, which is not always attainable in a single semester, is to see some gain in proficiency. If one grades oral proficiency alone, then those students who have a relatively high level of proficiency prior to taking the course—from having traveled abroad or by heritage, for example—will almost always demonstrate a higher level of proficiency than those without such a background, and they may not have to study for the course at all. On the other hand, if one grades only on the attainment of particular knowledge of the course content, then the teacher cannot grade the test results in a holistic manner. Often, this focus on particular knowledge means that the teacher downplays the importance of what the learners are actually saying in their responses.

Table 6
Divisions of the 24 Testees According to the Grading Criteria

High		Mid		Low	
Level	No. of Testees	Level	No. of Testees	Level	No. of Testees
High+	2	Mid+	3	Low+	2
High-Mid	4	Mid-Mid	6	Low-Mid	1
High-	2	Mid-	3	Low-	1

What we propose is that the advanced conversation teacher grade for both proficiency and achievement. In some respects, we are proposing a new kind of prochievement test, but not in the sense that the learner must show knowledge of discrete items. Instead, the learner must demonstrate proficiency in performing given functions, such as comparing advantages and disadvantages, in a cohesive, coherent, and organized manner. It follows, then, that, since course tests must reflect what is taught, a large part of the learner's achievement should be mastery of organizational structure. This focus then implies that an important component of the advanced conversation course be teaching good organizational templates for the tasks targeted at the Advanced and higher levels, such as those seen in the models of Tables 1A and 1B. If learners can be given a set of discourse templates that they can rely on to fill in propositions and if teachers listen for their skill in applying those templates, then we believe the learners will have better guidance in producing top-down and, hopefully, fairly complete answers.

These templates can be easily learned. In our course there was a de-emphasis of vocabulary specific to a given topic, since the TOPT confronts the learner with any topic related in any general way to what Spanish teachers in the U.S. would need to know or say in their teaching. In a typical (non-TOPT) advanced conversation course, however, the teacher could also teach and test for specific vocabulary, such as connective adverbials. We would hope, nevertheless, that discourse structure be considered one of the most, if not *the* most, important element to be mastered in the course.

One may argue that these templates are the same as those used in English and, therefore, we are not teaching the learners anything new. Experience, as well as our data, however, has shown that learners do not follow these frameworks under pressure in the target language. Learning how to express a topic sentence in a foreign language is a challenge, especially one that is clear and to the point of the question and that presents a good proposition for development.

Finally, we are proposing to move the basis of testing for the conversation class to the level of the discourse itself, thereby contextualizing the vocabulary and discrete grammar points. Learners can produce more coherent speech, and those who already have a fairly good oral proficiency upon entering the course can sound even more polished and articulate. Our recommendation for a top-down structured approach to discourse assessment is based on our encouragement of a top-down approach to the four skills of listening comprehension, reading, writing, and, as we have discussed here, speaking.

NOTES

1 One could argue that the logical conclusion to a response comparing advantages and disadvantages is an opinion and, in fact, that is how the model answers conclude in 1A and 1B. The question in the stimulus itself ("What

do you think?") may also seem to elicit an opinion, despite the emphasis on comparing advantages and disadvantages. In such a perspective, the organization would seem to be bottom-up, rather than top-down, with the opinion stated at the end. We point out, however, that an opinion is not the objective of the task, and that this task is answered most successfully with a topic sentence at the very beginning that guides the construction of the rest of the response. Moreover, testees rarely have time to give a conclusion at the end of their response, a fact that would also argue against a bottom-up analysis.

2 The notations used in the transcriptions are those found in Atkinson and Heritage (1984).

3 These translations do not show errors such as vocabulary or agreement errors.

REFERENCES

American Council on the Teaching of Foreign Languages. (1997). *ACTFL provisional revised speaking guidelines.* Yonkers, NY: ACTFL.

Atkinson, M., & Heritage, J. (1989). *Structures of social action.* Cambridge: Cambridge University Press.

Higgs, T. (1987). Oral proficiency testing and its significance for practice. *Theory into Practice, 26* (4), 282-287.

Hendrickson, J. (1992). Creating listening and speaking prochievement tests. *Hispania, 75* (5), 1326-1331.

Stansfield, C. (1993). An approach to performance testing of oral language proficiency for bilingual education teacher certification. *Proceedings of the Third National Research Symposium on Limited English Proficiency Student Issues: Focus on middle and high school issues.* Vol. 1. 187-215.

Stansfield, C., & Kenyon, D. (1992). Research on the comparability of the SOPI and the OPI. *System, 20,* 347-364.

Acquisition of Spanish Definite Articles by English-Speaking Learners of Spanish

MARÍA RAMÍREZ-MAYBERRY, The University of Texas at Austin

This study examines the stages of acquisition of Spanish definite articles by two groups of English-speaking learners enrolled in their first semester of Spanish. The approach taken was to analyze writing samples produced by the groups at different times in the semester in order to acquire the basis for a pseudo-longitudinal analysis of article usage. The main goal was to expose the variability in their interlanguage system and to propose early developmental stages of acquisition of the articles by learners of an introductory course. The results supported studies that postulate a natural order of acquisition. The results also showed a relatively significant incidence of first-language (L1) transfer consistent with the weak form of the Contrastive Analysis Hypothesis. The L1 interference, therefore, seems to be only one of many factors in the acquisition. VanPatten's ideas on a natural order of acquisition and Ellis's Interlanguage Theory provided a frame for analyzing the results. The interaction of several factors (simplification, communicative value, frequency of input, and first language transfer) proposed by VanPatten were useful in accounting for some of the stages of acquisition suggested in this study.

INTRODUCTION

Research in the past two decades has supported claims of natural stages in second language acquisition. These studies have provided evidence against the Contrastive Analysis Hypothesis, which states that all errors in a second language can be predicted by identifying the differences between the learner's native language and the target language (Ellis, 1990). Morpheme studies of the 1970s asserted that learners follow a certain order of acquisition regardless of language background, age (Ellis, 1990), method, text, teacher, or even error correction (VanPatten, 1987). More recent studies have illustrated more clearly the developmental stages that learners pass through in their acquisition of target grammatical structures. These studies have revealed patterns of acquisition of morphemes such as the copulas *ser/estar* and the prepositions *por/para* (VanPatten, 1987; Guntermann, 1992a and b; Ryan and Lafford, 1992; Lafford and Ryan, 1995). Other explanations, such as markedness (Rutherford, 1982; VanPatten, 1987; Guntermann, 1992a and b; Perdue and Klein, 1992) and cognitive theories (Gass, 1988, 1994; VanPatten, 1989; Tomlin and Villa, 1994; Bialystok, 1994), have sought to explain what is known as the Natural Order of Acquisition (VanPatten, 1987). One area that has been neglected, however, is the acquisition of Spanish definite articles by English-speaking learners.

The purpose of this study is to examine the acquisition of Spanish definite articles by two groups of English-speaking learners, with particular regard to the variability in their interlanguage systems. The main goal is to propose early developmental stages of acquisition of the definite article by the learners. We analyze these acquisition patterns based on Bickerton's (1975) "dynamic paradigm," a notion that "seeks to describe exactly what learners do in their developing interlanguage systems rather than how closely they approximate the target" (Master, 1988, p. 5). As Van Patten (1987, p. 102) pointed out 10 years ago, "there is a dearth of research detailing the acquisition or development of any one particular syntactic or morphological feature."

In light of these objectives, this analysis proposes to answer the following questions:

1. What stages of acquisition of Spanish definite articles can be perceived in the writing samples of learners in their first semester of Spanish study?

2. Does first language transfer influence these developmental patterns?

3. If first language transfer does not occur, what other factors may account for these stages of acquisition?

DATA COLLECTION [1]

The data for answering these questions were collected as described below.

Subjects

The subjects consisted of two groups of English-speaking learners: Group A included nine learners; Group B included ten. All learners were randomly chosen from a first-semester Spanish class.

Instrument

Writing samples of the two groups were gathered in order to provide the basis for a pseudo-longitudinal analysis of article usage. The writing samples were obtained as follows:

1. Group A (9): Writing was collected after 5 weeks of classes.

2. Group B (10): Writing was collected after 15 weeks of classes (Final Exam).

It was not possible to obtain data from the final exam of Group A. For this reason, we collected the writing samples from the final exam of a different group, Group B. It is recognized that this is a flaw in the design of the experiment. The results, therefore, are intended as an indication of what one can expect to find at this time period.

The writing assignment for Group A has five parts. The first four parts consist of oral comprehension exercises from which the learners gather information to use in the fifth part, the actual writing assignment. For this part, the learner writes a letter home of about 100 words. The writing assignment for Group B is a letter of at least 120 words in which the learners follow the description and instructions

provided. This composition is part of the final exam, and there are no preparatory exercises. See Appendix A for a more detailed description of the writing assignments for each group of learners.

Each learner from each group was assigned a number from 1 to 10. A list of noun phrases was created for each learner. Each entry included the following information:

1. Time of data collection (5 weeks or Final)
2. The noun phrase being analyzed plus minimal context
3. The article used
4. The article required

This study analyzed only the usage of the definite articles (singular forms: *el, la*; plural forms: *los, las*). Therefore, all the phrases produced by the learners where the use of a definite article is obligatory in Spanish (T) were counted, and those phrases in which the learners incorrectly used definite articles (I) were included also. The expressions that contained a correct definite article were marked as "C." It should be noted that for the purpose of this study, any form of the definite article (*el, la, los, las*) was counted as correct even if it was used with the wrong noun (for example, **el casa* 'the house'). This decision was reached because we are interested in the learner's selection of definite article over indefinite or zero article and not whether the learner has learned the gender notion in the Spanish noun phrase system. The total definite articles used, regardless of whether they were correctly or incorrectly provided (C + I), was labeled "U." Phrases requiring a definite article in which the learners used a zero article instead were labeled as "N."

Quantitative Analysis

The method of analysis used in this study was adapted from Master's (1988) study. The following notions were derived:

1. ***Correct in Obligatory Context (COC).*** To measure **accuracy** based on Master's proposed concept of "Supplied in Obligatory Context," this figure indicates the number of correct definite articles used (C), divided by the number of correct obligatory contexts in which the article should be used (T).
2. ***Used in Obligatory Context (UOC).*** To analyze **learner usage**, as the "dynamic paradigm" suggests, we divide the number of total definite articles used (U) by the number of obligatory contexts (T).
3. ***Incorrect in Nonobligatory Context (INOC).*** To obtain a complete picture of the usage of articles by learners, we added this notion, which is the number of incorrectly supplied definite articles (I) divided by the total number of definite articles used (U).
4. ***Needed in Obligatory Context (NOC).*** This value is the number of definite articles needed (N) di-

vided by the number of correct obligatory contexts (T).

As Master (1988, p. 9) points out, these notions are necessary to show how often definite articles are being used, regardless of accuracy, and with what degree of target-like use. Table 1 shows the variables and concepts used in this study.

Qualitative Analysis

This study analyzes the article production from the point of view of the relationship between form and function in discourse. We are trying to reveal any hidden systematization (Ellis, 1990, p. 50) in what seems an apparent variation of article production by learners. As mentioned before, we follow Master's (1988) approach and analyze the data in terms of learner usage (dynamic paradigm) and target language accuracy.

In our analysis, like that of Huebner (1983), we omitted proper nouns, idioms, and common expressions such as *en verano/en el verano* 'in summer/in the summer', and *en agosto* 'in August,' which, he observes, may be learned as formu-

Table 1
Definition of Terms
[Adopted from Master's (1988)]

C	Number of articles *correctly* used
I	Number of articles *incorrectly* used
N	Number of articles *needed* but not used
T	*Total* number of articles required (C + N)
U	Number of articles *used* (C + I)
COC	Percentage of correct articles used (C/T * 100)
UOC	Percentage of total articles used (U/T * 100)
INOC	Percentage of incorrect articles used (I/U * 100)
NOC	Percentage of articles needed (N/T * 100)

Measures:

	Needed	Not Needed
Used	C	I
Not Used	N	

lae. We included phrases such as *en/por la mañana* 'in the morning' and *regresar a casa/regresar a la casa* 'come back home,' however, even though these expressions are also learned as formulae, because we found some learners omitted the required articles in these noun phrases, and we wanted to seek the pattern behind this usage.

RESULTS

Analysis of Data

We begin our analysis by calculating the results per group. Tables 2A and 2B present a summary of the quantitative results. For Group A, we tallied a total of 104 obligatory contexts. The accuracy registered was 83%, with 86 articles correctly supplied and 18 places where the articles were required but not provided (17% NOC). The group, however, supplied 19 extra articles not required (18% INOC). Group B had a total of 90 correct obligatory contexts and provided 92 definite articles (102% UOC, which is the same percentage for Group A). The group's accuracy (COC) was 80%, a little lower than that of Group A. They supplied, however, more unnecessary definite

Table 2
Distribution of Definite Article Usage

A. Group A

LEV	STU	T	U	C	I	N	COC	UOC	INO-C	NOC
							%	%	%	%
5wk	1	5	2	2	0	3	40	40	0	60
	2	18	21	17	4	1	94	117	19	6
	3	3	1	1	0	2	33	33	0	67
	4	8	6	6	0	2	75	75	0	25
	5	22	22	19	3	3	86	100	14	14
	6	13	14	12	2	1	92	108	14	8
	7	13	12	10	2	3	77	92	17	23
	8	10	10	9	0	1	90	100	0	10
	9	12	18	10	8	2	83	150	44	17
		104	106	86	19	18	83	102	18	17

Table 2
Distribution of Definite Article Usage *(Continued)*

B. Group B

LEV	STU	T	U	C	I	N	COC	UOC	INO-C	NOC
							%	%	%	%
Fin	1	3	3	2	1	1	67	100	33	33
	2	13	12	11	1	2	85	92	8	15
	3	5	6	5	1	0	100	120	17	0
	4	7	10	7	3	0	100	143	30	0
	5	4	4	2	2	1	50	100	50	25
	6	9	16	9	7	0	100	177	44	0
	7	12	10	7	3	5	58	83	30	42
	8	15	11	11	0	4	73	73	0	27
	9	10	7	6	1	4	60	70	14	40
	10	12	13	12	1	0	100	108	8	0
		90	92	72	20	17	80	102	22	19

articles (22% INOC) than Group A (18%); also, the number of definite articles needed but not supplied produced a slightly higher percentage in Group B (19% NOC) than Group A (17% NOC).

Although the difference between the results of Group A and B does not seem to be significant, individual results are revealing and supportive of the tendencies observed in the general results. In Group A, as we already have pointed out, learners used fewer incorrect articles in non-obligatory contexts than Group B. It is interesting to see that out of the nine learners from Group A, four of them did not supply any extra incorrect articles (I), which means they have less overuse of articles than learners from Group B. Among the other five, the number of incorrect articles in non-obligatory contexts for all but one was less than 20%. Only one learner had a higher percentage (44%). The

results from Group B are also revealing. As noticed before, their number of incorrect articles in non-obligatory contexts as a group was slightly higher (22%). Their individual performances confirm these data: all but one learner supplied extra articles (I) not required, and five of the learners showed percentages of 30% or more.

The results shown by the number of articles required but not supplied support the tendencies noticed so far. Even though, as a group, learners from Group B seem to lack more needed articles than do learners from Group A (Group A had a 17% NOC versus 19% from Group B), a look at the individual tendencies shows a different picture: 4 out of 10 learners from Group B did not use articles at all. The rest of the learners of this group showed percentages of 40% or less. The data from Group A learners show that all were lacking articles that were obligatory. Two of these learners, in particular, showed underuse of articles in obligatory contexts over 50%.

In general, then, the errors detected in the two groups seem to be related to the difference of data collection times. The writing samples of Group A show that learners, after 5 weeks of classes, tended to omit required articles; meanwhile, the data collected from Group B show a different error after 15 weeks of classes, that of the overuse of articles, possibly due to the learners having entered a stage of experimentation after the longer exposure to Spanish.

Stages of Acquisition

The analysis of the data suggests that learners go through certain stages in the acquisition of articles in Spanish. We propose the following stages of acquisition shown in Appendix B, although the stages and suggested order are tentative. A true longitudinal study must be done before arriving at definite conclusions.

Stage 1. Omission of the Definite Article. This stage is demonstrated by the omission of articles even in formulaic expressions such as *en/por la mañana* 'in the morning,' *los sábados* 'on Saturdays,' etc., regardless of the fact that similar expressions are found in English (for example, **Limpio el apartamento todos __ dias mañana* 'I clean the apartment every day in the mornings').

This stage is supported by the following observations: All learners from Group A wrote sentences lacking a required article. Two of those learners failed to use required articles more than 50% of the time. In contrast, four learners from Group B did not have this problem at all (0% NOC), which means that they did not produce sentences like **todos _ días* 'everyday' or **no me gusta _ trabajo* 'I don't like my job.' Also, for no Group B learner did the percentage of sentences that lacked required articles exceed 42%.

Stage 2. Overuse of Articles (The los*-Stage).* In this stage, learners use definite articles in non-obligatory contexts. This stage seems to reflect an increased awareness of articles, because learners start to overuse them.

Observations that support this stage are the following: Group B had 22% of incorrect articles in non-

obligatory contexts. All learners from this group, with one exception, produced unnecessary articles (e.g., **El estudiante necesite traer* los *blue jeans,* los *pantalones,* las *camisas, y* las *camisetas. También necesite traer* los *zapatos,* las *botas, y suéteres* 'The student needs to bring blue jeans, pants, shirts, and T-shirts. He also needs to bring shoes, boots, and sweaters').

Group A had 18% of incorrect articles in non-obligatory contexts. Four learners from this group did not make this error (0% INOC). All this may be evidence that there is a direct correlation between the overuse of definite articles and the time of data collection. Thus, as mentioned before, as the learner is exposed to more Spanish, the incidence of overusing the articles increases, and learners supply articles even in a context where English may not require them *(e.g., The learner needs to bring _blue jeans, _pants, etc.).*

An additional observation is that learners from both groups tended to overuse articles when faced with a string of nouns. Six learners from Group A produced strings of nouns (list of classes or of clothing), and five of them overused the definite article. All the learners from Group B produced strings of nouns, with six of them having more than one string. In particular, learners 4, 5, and 6, who produced two sets of strings each, had one set in which the definite articles were overused and one set in which other articles or a combination of definite, indefinite, and zero articles was used. Learner 6, for example, wrote **El estudiante necesite traer* los *blue jeans,* los *pantalones,* las *camisas, y* las *camisetas. También necesite* los *zapatos,* las *botas y* los *suéteres* 'The student needs to bring blue jeans, pants, shirts and T-shirts. He also needs to bring shoes, boots, and sweaters' but he also wrote: *Los estudiante y yo jugamos Ø futbol Americano, Ø beisbol, Ø tenis, y Ø futbol* 'The students and I play football, baseball, tennis, and soccer.' It seems that learners from Group B began to use all of the articles more than learners from Group A in the same environment of listing elements, which brings us to the next stage [2].

Stage 3. Vacillation. After overusing articles, learners seem to go through a period of inconsistency in their selection of articles, called here *a stage of experimentation.* At first, they appear to choose articles (definite, indefinite, or zero article) at random; however, we noticed a subtle, systematic use of articles. Learners seem to make their own rules of usage based on number, and once they have made a rule, they seem to follow it consistently in any given string. As with the previous stage, it is important to notice that this rule seems to apply when the learners see *a string of nouns.*

The following are some of the patterns detected:

1. Three learners of Group B used definite articles with plural nouns and indefinite articles with singular nouns. For example,

   *. . . *también el estudiante necesita llevar* los *suéteres y* un *abrigo.*

'. . . also, the student needs to bring sweaters and a coat.'

2. One learner from Group A showed this tendency to use definite articles with plural nouns as well, although he used the zero article with the singular. For example,

   **Tomo* 0 *español, y* 0 *clase de niños developmental, y* las *matimáticas, y* las *ciencias.*
   'I am taking Spanish, child development, math, and sciences.'

3. One learner from Group B preferred the zero article with plurals and definite with singulars. For example,

**Necesite dos pars de blujeans,* 0 *zapatos de tenis,* la *chaqueta,* 0 *botas, y* 0 *suéteres,* 0 *faldas.*
'(She) needs two pairs of blue jeans, tennis shoes, a jacket, boots, sweaters, and skirts.'

4. One learner from Group B used the indefinite article for plurals and singulars. For example,

   Diga al estudiante traer unos *pantalones cortos y* unas *camisetas para llevar en verano y otoño, y* unos *blue jeans y* una *chaqueta para llevar en invierno.*
   'Tell the student to bring some shorts and some T-shirts to wear in summer and fall, and some blue jeans and a jacket to wear in winter.'

Stage 4. Grammatical Stage. Learners in this stage show correct usage of articles (definite, indefinite, and zero articles). The accuracy seems to increase as the learner is exposed to more Spanish. This level does not seem to be a consistent stage among beginning learners, however. It should be remembered that both groups of learners are beginners. The same learner could write a good sentence at one point of the assignment (e.g., *diga al estudiante traer* unos *pantalones,* 0 *suéteres, y* un *abrigo* 'tell the learner to bring some pants, sweaters, and a coat') and an incorrect one in the same assignment (e.g.,**Hola director de* los *estudiantes extranjeros* 'Hello, director of foreign students').

Some observations follow: Only one learner from Group A produced this kind of string with grammatical phrases, compared to four learners from Group B. Four learners from Group B had an accuracy rate of 100%. The highest accuracy rate in Group A was 94%. The lowest accuracy rate in Group B was 58%, compared to the two lowest of 33% and 40% in Group A. Again, all this suggests a correlation with the difference of data collection time between the two groups. The accuracy rate was lower in the writing samples collected after 5 weeks (Group A) than the accuracy rate observed in the data gathered after 15 weeks of classes (Group B).

DISCUSSION

These findings are not consistent with the strong form of the Contrastive Analysis Hypothesis, which assumes that interference occurs as a result of L1 transfer and states that "errors could be predicted by identifying the linguistic differences between the learners' L1 and

the target language" (Ellis, 1990, p. 29). In the compilation of occurrences due to L1 transfer, we found both groups have similar results. Out of the 47 sentences with errors produced by Group A, 16 (34%) were identified as due to L1 transfer (e.g., *En *lunes miramos un hombre toca la guitarra.* . . 'On Mondays, we watch a man playing the guitar. . .'). Learners from Group B included in their writings 37 sentences with errors, 13 of which (35%) were due to L1 transfer (e.g., **Lago Travis es muy bonito* 'Lake Travis is pretty'). This relatively significant incidence of L1 transfer agrees more with the weak form of the Contrastive Analysis Hypothesis, which states that L1 interference may explain some errors, but it is only one of many factors in the acquisition and cannot "be treated as a major source of error" (Ellis, 1990, p. 29). Moreover, as shown by our results, the Contrastive Analysis Hypothesis does not seem to explain the variability of errors observed in the usage of definite articles.

Results from our study suggest support of the Interlanguage Theory premise outlined by Ellis (1990) that suggests that learners construct a system of transitional linguistic rules with which they are continually experimenting in their approximation toward the target language. This hypothesis is supported by the fact that learners in the vacillation stage seem to develop a rule for article usage based on the plural or singular suffix of the noun. As Master (1988) suggests, the variation reflects certain strategies of interlanguage development, which can be seen in the learners' actual use of morphemes at different interlanguage levels [3]. In this way, as proposed by Corder (1967), their interlanguage system can be seen as a restructuring or recreating continuum.

Moreover, VanPatten's (1987) ideas on a natural order of acquisition may explain our results. He proposes the interaction of four factors to explain the order of acquisition. At the beginning, learners may fail to perceive the *communicative value* in articles and omit them (Stage 1, Omission) even in phrases that have similar constructions in English (e.g., *in the morning*, etc.) When learners discover that Spanish articles occur more frequently than in English (*frequency of input*), they *simplify* the rules of usage, thinking that Spanish articles are used in every context, and start using the articles indiscriminately (Stage 2, Overuse and Stage 3, Vacillation). *First language transfer*, as proposed by Andersen (1983), may also be responsible for the Vacillation Stage. When the learners enter a period of experimentation in their interlanguage development, they mistakenly perceive a similarity between the usage of Spanish and English articles. The role of L1, however, is limited "to those stages when there is convergence between L1 structure/function and developing interlanguage" (VanPatten, 1987, p. 111). VanPatten adds, "The interaction between L1 interference and interlanguage cannot violate the natural processes of acquisition that are in progress."

Another explanation for some of the stages proposed in this study, particularly the overuse of the definite articles, may be found in the sys-

tem used to learn vocabulary. The textbook used at this university, *Puntos de Partida* (Dorwick et al., 1993) presents a list of nouns with their corresponding definite articles in order to help the learners in the acquisition of the Spanish gender. As VanPatten (1989) cautions, however, learners at this early stage of acquisition may have difficulty attending to form in the input, because at the same time they are struggling to understand the meaning. Tomlin and Vila (1984), on the other hand, suggest that the learning of grammar can be enhanced when the learner somehow attends to form. This issue needs further study in order to isolate the causes for the overuse of definite articles.

CONCLUSIONS

Besides supporting previous studies that postulate a Natural Order of Acquisition, this study sheds light in an area of second language acquisition that has been neglected: the acquisition of Spanish articles by English-speakers. Several stages of acquisition were proposed in response to Research Question 1, based on what was observed in the writing samples of two groups of an introductory Spanish course. These stages are tentative since more research is needed to test these findings. Although we analyzed the writing assignments of two different groups at different times during the semester, both groups are still considered beginners.

Again, we stress the need for true longitudinal studies of the acquisition of definite articles as well as indefinite and zero articles. These studies are needed for comparison purposes in order to develop a more complete picture of the acquisition of these morphemes. Also required are studies that test the oral proficiency of learners with respect to definite articles. In general, as Parrish and Tarone (1988) have done, multitask research is needed in order to observe any task-related variation in the acquisition stages.

We found interesting results in response to the second and third research questions. Some L1 transfer errors were observed, and the percentages of L1 interference were very similar in both groups. L1 interference, however, seems to be one of four factors that interact as part of the interlanguage development that learners go through in their natural process of acquisition, as proposed by VanPatten (1987). These factors (L1 transfer, simplification, communicative value, and frequency of input) were useful in accounting for the stages of acquisition proposed in this study. The interlanguage issues outlined by Ellis (1990) were other factors in agreement with our results that could account for the variability observed in the learners' usage of definite articles.

NOTES

1 For this study I have a small corpus of data; therefore, any results obtained should still be considered tentative. This is a pilot study, and I would like to share my observations based on the data collected. These findings may serve as the foundation for questions that can be looked at in the future.

2 The analysis of the use of the other articles is beyond the scope

of this study; however, we noticed that when faced with a string of nouns, Group B learners tended to overuse all the articles (including the indefinite), while learners from Group A used only definite and zero articles.

3 In his study, Master (1988) established the interlanguage level of his subjects by their use of the negation morpheme (p. 10).

REFERENCES

Andersen, R. (1983). Transfer to somewhere. In S. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 177-201). Rowley, MA: Newbury House.

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition, 16* (2), 157-168.

Bickerton, D. (1975). *Creolization, linguistic universals, natural semantics and the brain.* Paper presented at the International Conference on Pidgins and Creoles (University of Hawaii, January 6-11, 1975).

Corder, P. (1967). The significance of learners' errors. *International Review of Applied Linguistics, 5,* 161-169.

Ellis, R. (1990). *Instructed second language acquisition: Learning in the classroom.* Oxford: Basil Blackwell.

Gass, S. (1988). Integrating research areas: A framework for second language studies. *Applied Linguistics, 8* (2), 198-217.

Guntermann, G. (1992). An analysis of interlanguage development over time: Part I, *por* and *para. Hispania, 75* (1), 177-187.

Guntermann, G. (1992. An analysis of interlanguage development over time: Part II, *ser* and *estar. Hispania, 75* (5), 1294-1303.

Huebner, T. (1983). *A longitudinal analysis of the acquisition of English.* Ann Arbor: Karoma.

Lafford, B., & Ryan. J. (1995). The acquisition of lexical meaning in a study abroad context: The Spanish prepositions *por* and *para. Hispania, 78* (3), 528-547.

Master, P. (1988). *Acquiring the English article system: A cross-linguistic interlanguage analysis.* Paper presented at the 22nd Annual Meeting of the Teachers of English to Speakers of Other Languages (Chicago, IL, March 8-13, 1988).

Parrish, B., & Tarone, E. (1988). Task-related variation in interlanguage: The case of articles. *Language Learning, 38* (1), 21-44.

Perdue, C., & Klein, W. (1992). Why does the production of some learners not grammaticalize? *Studies in Second Language Acquisition, (14),* 259-272.

Rutherford, W. (1982). Markedness in L2 acquisition. *Language Learning ,32* (1), 85-108.

Ryan, J., & Lafford, B. (1992). Acquisition of lexical meaning in a study abroad environment: *Ser* and *estar* and the Granada experience. *Hispania ,75* (3), 714-722.

Tomlin, R., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition, 16* (2), 183-203.

VanPatten, B. (1987). Classroom learners' acquisition of *ser* and

estar. In B. VanPatten, T. Dvorak, & J. Lee (Eds.) *Foreign language learning: A research perspective* (pp. 61-75). Cambridge: Newbury House.

VanPatten, B. (1989). Can learners attend to form and content while processing input? *Hispania, 72* (2), 409-417.

APPENDIX A
WRITING ASSIGNMENTS

WRITING ASSIGNMENT FOR GROUP A. There are five parts to this writing assignment. Parts 1 and 2 are done the first day. Part 3 and Part 4 are given as homework for the second day. The third day, the students do the actual writing assignment in the computer lab.

Parte 1: Lectura 'Reading' (This reading was read by the instructor in class. He wrote some of the main points in the blackboard:)

> *La Universidad de Texas, fundada en 1883, es un lugar muy atractivo para los estudiantes extranjeros. Más de 3500 estudiantes de más de 100 países asisten a UT. Esta universidad tiene unas bibliotecas extraordinarias, incluída la Nettie Benson Latin American Collection, una de las mejores colecciones de libros sobre Latinoamérica en el mundo.*
>
> *Además de su fama académica, UT está situada en una ciudad preciosa. A los estudiantes extranjeros les gusta la vida cultural (los conciertos de música clásica, los museos, las conferencias), el clima, los parques, y los lagos de Austin. También les gustan las atracciones turísticas como el capitolio, la biblioteca presidencial de LBJ, la famosa Calle Seis, y los murciélagos que viven debajo del puente de la Calle Congress.*
>
> *Para los estudiantes que necesitan perfeccionar su inglés antes de matricularse en UT, hay un programa excelente llamado Texas Intensive English Program en Dexter Hall en la Calle 24 cerca del campus. A veces estos estudiantes vienen a las clases de español para hablar con los estudiantes americanos sobre sus países de origen.*
>
> 'The University of Texas, founded in 1883, is a very attractive place for foreign students. More than 3500 students from more than 100 countries attend UT. This university has great libraries, including the Nettie Benson Latin American Collection, one of the best collections of books on Latin America in the world.

'Besides its academic fame, UT is situated in a beautiful city. Foreign learners like Austin's cultural life (classical music concerts, museums, conferences), its weather, its parks, and its lakes. They also like its tourist attractions like the Capitol, the LBJ Library, the famous Sixth Street, and the bats that live under the bridge on Congress Street.

'For students who need to improve their English before enrolling at UT, there is an excellent program called *Texas Intensive English Program* at Dexter Hall on 24th Street, near campus. Sometimes these students come to Spanish classes to talk with American students about their countries.'

Parte 2: Comprensión oral 'Listening comprehension' (The students do not see the script. The instructor reads this to them and then asks the four follow-up questions.)

Pepe Ramos es un estudiante extranjero que estudia química en la Universidad de Texas. Es de Limón, Costa Rica. Es la primera vez que vive lejos de su familia. Ahora vive en la residencia Jester con su compañero de cuarto que se llama Joe Weaver. Joe es de Lubbock. Pepe tiene sus momentos tristes, pero le gustan sus clases y su compañero de cuarto. No le gusta la comida de Jester por eso va a vivir en un apartamento el semestre que viene.

'Pepe Ramos is a foreign student who studies chemistry at the University of Texas. He is from Limón, Costa Rica. It is the first time he is away from his family. Now he lives at Jester with his roommate Joe Weaver. Joe is from Lubbock. Pepe has his sad moments, but he likes his classes and his roommate. He does not like the food at Jester, that is why he is going to live in an apartment next semester.'

Parte 3. *Preguntas* 'Questions':

¿De dónde es Pepe? 'Where is Pepe from?'
¿Qué estudia Pepe? 'What does Pepe study?'
¿Por qué tiene sus momentos tristes? 'Why does he have sad moments?'
¿Qué desea hacer el semestre que viene? 'What does he want to do next semester?'

Parte 4: Práctica oral 'Oral practice.' The instructor has students put the sentences on the board as they come in before class starts the following day. The instructor corrects them as a class.

Parte 5: Escritura 'Writing.' After the above four parts have been completed, students write the writing assignment, which is a letter written home by Pepe, the student from Costa Rica, or by Yoly, the Peruvian student. The letter should be about 100 words. They are given the *Querida Mamá* 'Dear mom' and different ways to sign *Abrazos* 'hugs' or *Besos* 'kisses.' They do this assignment in the Computer Lab.

WRITING ASSIGNMENT FOR GROUP B. The following are the instructions for a composition that students write as part of their final exam.

Composición 'Composition.' Write a composition of at least 120 words following the description provided. Remember to include at least three command forms in your composition.

Your family will be hosting a Costa Rican student. You are the only one who speaks Spanish in your family. Your parents have asked you to write a letter to the director of the program to give him the following information:

- Greet the director and introduce yourself and your family.
- Since the exchange student is coming from August to January, tell the director what kind of weather to expect and what kind of clothing a person needs to bring.
- Compare Austin to other cities in Texas.
- Then tell him that you have studied Spanish this year and that you are going to help the student in Austin.
- Describe some of the forms of entertainment in the Austin area.
- Tell what interesting things you did this year.
- Tell him to write back with any question he might have.

APPENDIX B
SAMPLE SENTENCES OF THE PROPOSED STAGES AND L1 INTERFERENCE

1. Omission of Definite Articles.

Example: **Limpio el apartmento todos 0 dias 0 mañana.*
'I clean the apartment every day in the morning.'

The *los*-stage.

Example: **Las clases llevo la geología, la matématica, y la español.*
'The classes I am taking are geology, math, and Spanish.'

Example: **El estudiante necesite traer los blue jeans, los pantalones, las camisas, y las camisetas. También necesite traer los zapatos, las botas, y los suéteres.*
'The student needs to bring blue jeans, pants, shirts, and T-shirts. He also needs to bring shoes, boots, and sweaters.'

3. Vacillation.

Definite article with plurals; indefinite article with singulars.

Example: **. . . debe llever los blue jeans, las camisetas y uno o dos chaquetas . . . también el estudiante necesita llevar los suéteres y un abrigo.*
'. . . he should bring blue jeans, T-shirts, and one or two jackets . . . also, the student needs to bring sweaters, and a coat.'

Definite for plurals; zero articles for singulars.

Example: **Tomo 0 español, y 0 clase de niños developmental, y las matimáticas, y las ciencias.*
'I am taking Spanish, child development, math, and sciences.'

Zero article for plurals; definite article for singulars.

Example: **Necesite dos pars de blue jeans, 0 zapatos de tenis, la chaqueta, 0 botas, 0 suéteres, y 0 faldas.*
'(She) needs two pairs of blue jeans, tennis shoes, a jacket, boots, sweaters, and skirts.'

Indefinite for plurals and singulars.

Example: *Diga al estudiante traer unos pantalones cortos y unas camisetas para llevar en verano y otoño, y unos blue jeans y una chaqueta para llevar en invierno.*
'Tell the student to bring some shorts, and some T-shirts to wear in summer and fall, and some blue jeans and a jacket to wear in winter.'

4. Grammatical stage.

Example: *Necesite llevar un suéter, una chaqueta, y 0 ropa para la invierno.*
'(He) needs to bring a sweater, a jacket, and clothing for winter.'

L1 interference examples:

*Ø *Lago Travis es muy bonito.*
'Lake Travis is pretty.'

**En Tejas, todo Ø año hace un buen tiempo.*
'In Texas, all year long the weather is nice.'

**En lunes miramos un hombre toca la guitarra....*
'On Mondays, we watch a man playing the guitar....'

The Spanish Psych Verb Construction: Beginning and Intermediate Learners' Patterns of Usage

CHRISTOPHER D. GASCÓN, The University of Texas at Austin

The Spanish psychological verb construction seems to be especially difficult for native English-speaking learners to acquire. Since some of the most common Spanish psych verbs, such as gustar *'to please' and* encantar *'to delight,' require a grammatical structure that is different from that of the English verbs frequently taken as their equivalents 'to like' and 'to love,' it is understandable that native English speakers may struggle with the construction. More precise information on how learners attempt to acquire this complex form would, however, prove valuable to Spanish instructors. This study presents data that describe the patterns of usage of students in first-, second-, third-, and fourth-semester university-level Spanish courses. An analysis of the data reveals that learners initially commit errors related to the transfer of structures from the native language. Gradually, however, transfer errors decline and developmental errors increase as learners experiment with elements of the target language. The study proposes that within this learning pattern, it is possible to discern distinct stages in the learners' progress toward acquisition.*

INTRODUCTION

Gustar and other psychological verbs present a challenge to native English speakers attempting to learn Spanish. Since the "psych verb" construction requires the learners to discern subject and beneficiary (or sufferer) in an inverted word order and to use the proper indirect object pronoun, the learners struggle to acquire this structure. It is thus of interest to instructors to know more about how learners develop their understanding of this construction so that they may be better able to formulate effective strategies for facilitating the acquisition process. This study addresses the following questions in an effort to achieve a clearer understanding of learners' acquisition of the Spanish psych verb construction:

1. What are the patterns of usage typical of learners attempting to acquire psych verbs such as *gustar* 'to please,' *importar* 'to matter,' *encantar* 'to delight,' *molestar* 'to bother,' and *interesar* 'to interest'? How do the learning patterns of students at beginning levels compare to those of intermediate students?

2. How can we account for the performance of the learners?

3. Does the acquisition of the psych verb construction reflect stages of interlanguage development?

Though it is beyond the scope of this study to recommend specific techniques for teaching the Spanish psych verb construction, it is hoped that the data and analysis presented here may assist instructors in finding effective ways to help learners acquire the structure.

It is possible for any Spanish instructor to make certain general predictions about how learners will handle the psych verb construction. First, it is expected that, overall, learners will experience difficulty in managing this construction. According to the Contrastive Analysis Hypothesis (Ellis, 1990, p. 25), the more closely a structure in the target language (L2) corresponds to its equivalent expression in the native language (L1), the easier it will be to learn, and vice versa. Since many of the most common English psych verb constructions (*like, love, am interested in*) utilize a "direct" structure while their equivalent expressions in Spanish (*gustar, encantar, interesar*) require a "reverse" construction [1], we may expect students to struggle with the form generally.

Second, it may be predicted that, due to frequency of occurrence in input (VanPatten, 1987, p. 87; Ellis, 1990, p. 96) and according to the output hypothesis (Ellis, pp. 117-119) and the discourse hypothesis (Ellis, pp. 119-121), learners will be more proficient in discussing their own likes and dislikes than in describing those of friends or parents. The first-person construction seems to be the one learners hear and are asked to produce most often in beginning and intermediate Spanish classes; thus, they should perform it best.

Third, it is generally expected that students at higher levels will outperform those at lower levels in all tasks. Results should suggest that students improve their psych verb proficiency as they advance from level to level.

Certain common errors may also be predicted. We can expect L1 transfer of the English word order S-V-O [2], leading learners to confuse the Spanish subject and object and commit errors like **me gusto la música* 'to me I am pleasing the music' [3]. We may also see the common developmental error of the use of *se* instead of the proper indirect object pronouns *le* or *les*.

BACKGROUND LITERATURE

Though psych verbs have been of interest to linguists studying comparative syntax and relational grammar, they have received little attention from second language acquisition researchers. Most of the literature on the subject of psych verbs focuses on the debate over the formal characterization of the construction; in particular, of its transitive properties. Though the increasingly more detailed definitions of transitivity that this debate has produced may help second language instructors systematize their own conceptions of how the various types of psych verbs work, it is doubtful that these definitions will prove to be of any practical use in teaching basic psych verb constructions to beginning learners. Nevertheless, these studies have established a basic lexicon that is helpful to any discussion on the subject. Thus, a brief review of some definitive studies, with an emphasis on the terminology they

use, is appropriate before proceeding to an explanation of the present study.

Belletti and Rizzi (1988) first distinguish between the *experiencer*, which is "the individual experiencing the mental state," and the *theme*, which is "the content or object of the mental state" (p. 292). They then postulate three distinct types of psych verb constructions based on Italian syntax. The first type of construction casts the experiencer into the role of the deep-structure subject and the theme into the accusative role. This psych verb configuration is the most canonical and is exemplified by the verb *temere*. In the second type of construction, exemplified by *preoccupare*, the theme is considered a "derived subject," while the experiencer is the accusative object. The third type of construction, exemplified by *piacere*, works in the same way, except that the experiencers are assigned dative status: they are the indirect beneficiary, rather than the direct recipient, of the effect.

The Belletti and Rizzi article has provoked a series of criticisms, revisions, and refinements by other relational and Chomskyan grammarians of its classification of the roles of experiencer and theme in the various psych constructions (Bouchard, 1992; Masullo, 1992; Saltarelli, 1992; Herschensohn, 1992; and Whitley, 1995) [4]. Of all of these, Whitley's study of transitivity is most pertinent to the present study. He first translates Belletti and Rizzi's tripartite conceptualization of Italian psych verbs into a Spanish version that features a four-part typology. The terminology he uses is different from theirs; while he accepts the term *experiencer* from their study, he prefers *cause* to their *theme*. In addition, he distinguishes between *direct* verbs, which cast the experiencer as subject (as in *I like it*), and *reverse* verbs, which treat the experiencer as object (as in *it pleases me*). Using these distinctions, he explains the four types of psych verbs:

> Type 1: Direct transitive; for example, *desear*. The experiencer acts as the subject, the cause as direct object.
>
> Type 2: Direct intransitive; for example, *gozar de/en*. The experiencer acts as subject, and the cause as an "oblique object" of a verb-specific preposition.
>
> Type 3: Reverse intransitive; for example, *gustar*. The experiencer is the indirect object, and the cause is cast as subject. The indirect object is optional: the experiencer may be generalized or impersonal, as in *La música rock gusta en todas partes*.
>
> Type 4: Reverse transitive; for example, *fascinar*. The cause functions as subject; the experiencer acts as the direct object. (573-574)

Whitley goes on to show, however, that such rigid syntactical categories break down when we analyze Spanish psych verbs more closely, and he proposes instead the idea of a transitivity "squish" based on se-

mantic distinctions. He borrows this term from John Ross (1973), who used it to describe a continuum of noun phrase types. He also utilizes a list of ten features proposed by Hopper and Thompson (1980) as a gauge of transitivity. He concludes with a call for other pedagogical explanations of *gustar* verbs that are more representative of the complexities of psych constructions, claiming that such work "would be especially valuable in second language acquisition" (p. 582).

There is little evidence, however, that comprehensive metalinguistic explanations of concepts such as transitivity facilitate beginning and intermediate learners' acquisition of basic psych verbs. It is more likely that input, output, and interactive activities have a greater effect on learners' mastery of psych verb constructions. Nevertheless, relational grammar does provide terminology and an explanation of the framework of psych configurations that are fundamental to any study of the concept.

DESCRIPTION OF STUDY

The first step in understanding better how learners attempt to acquire this complex construction is that of gathering and analyzing samples of their usage. The purpose of this pilot study has been to generate examples of the use of Spanish psych verbs by beginning and intermediate learners, to compare the performance of learners at different levels, and to look for patterns of usage that might suggest stages of interlanguage development.

Subjects

The subjects of the study were students in four University of Texas at Austin Spanish classes: 23 S1 (beginning level - first semester Spanish) students, 26 S2 (second semester Spanish) students, 36 S3 (third semester Spanish) students, and 20 S4 (fourth semester Spanish) students. At the time of data collection, the S1 students had been introduced to the verb *gustar* and had practiced its use to describe their own likes and dislikes. The use of *gustar* in third person constructions had not been emphasized, though it occasionally appeared in input. The students had not worked with other psych verbs, but they were informed at the time of the experiment that the other psych verbs, like *importar, encantar, molestar,* and *interesar,* worked generally like the verb *gustar* (that is, they feature a reverse structure).

The students in S2 had been introduced to some psych verbs other than *gustar* and had reviewed direct and indirect object pronouns. It is assumed that they also had some practice in the use of reverse constructions through studying the use of the subjunctive after expressions of emotion, such as *me sorprende que* or *me molesta que.*

The instructor of the S3 students had given little explicit attention to psych verbs in class, though the construction was recycled in input and output with some frequency.

The students in S4 had reviewed *gustar*-type verbs in a unit dedicated explicitly to the subject 10

class days prior to the experiment. In addition, instructors of fourth-semester Spanish emphasize this construction as essential to one of the seven communicative goals of the course, the expression of likes and dislikes.

Instrument

The learners completed three written tasks:

Task 1: Describe your own likes and dislikes.
Task 2: Describe a friend's likes and dislikes.
Task 3: Describe your parents' likes and dislikes.

They were instructed to vary their selection of verb and to choose from among the verbs *gustar, importar, encantar, interesar,* and *molestar.* The task was designed to generate samples of usage of psych verbs requiring reverse structure that were taught at most of the levels with the indirect object pronouns *me, le,* and *les.* The task called for a variety of verbs in order to ensure that performance would be particular to the psych verb construction and not simply to one or two specific verbs.

The percentage of correct responses (relative to the total number of attempts at the target structure made) was computed for each of the three tasks at each of the four levels. If the student made no error in conjugating the verb, selecting the correct object pronoun, or including the preposition *a* when appropriate, the sentence was judged as correct. Other errors, such as those in spelling or use of subjunctive/indicative, were disregarded.

Types of errors were also noted. They fell into three main categories, two of which have subcategories:

1. Omission of the preposition *a* 'to' when inclusion is required
   (1) Example: **Mi amigo no le interesa limpiar la casa* 'My friend does not interest cleaning the house.' The preposition *a* should precede *mi amigo.*

2. Object pronoun errors

   A. Omission of object pronouns *me* 'to me,' *le* 'to him/her,' or *les* 'to them' when inclusion is required
      (2) Example: **A mi amigo encanta comer* 'Eating delights to my friend.' The direct object pronoun *le* should precede *encanta.*

   B. Use of reflexive pronoun *se* instead of the correct object pronouns *le* 'to him/her' or *les* 'to them'
      (3) Example: **A José no se importa la clase* 'The class does not matter itself to José.' The pronoun *le* should be used instead of *se.*

   C. Confusion of object pronouns *le* 'to him/her' and *les* 'to them'
      (4) Example: **A mis padres le gusta la comida buena* 'Good food is pleasing to him to my parents.' The pronoun *les* should be used instead of *le.*

3. Verb morphology errors

   A. Subject-object confusion
      (5) Example: **Mis padres gustan ir al cine* 'My parents are pleasing to go to the movies.'

   B. Use of singular verb for plural subject (cause)
      (6) Example: **A ellos les gusta las fresas* 'The strawberries is pleasing to them.'

The first two categories are self-explanatory. With regard to verb morphology errors, subject-object confusion was judged as having occurred when (a) the form of the verb agreed with the form of the object pronoun instead of with the subject (cause), and (b) a check of other responses by the same student confirmed that this error was repeated consistently and systematically in other situations. For example, an S2 student performing Task 3 wrote these four sentences:

(7) **Mis padres gustan con yo estudio* 'My parents please with I study'

(8) *Les gustan mis amigos* 'They like my friends'

(9) **No les gustan con yo hago gradas mal* 'They are not pleasing to them with I make bad [grades]'

(10) **Les encantan mí* 'To them they delight me'

Since the student used the third-person plural conjugation in all cases regardless of what the cause (subject) was, it remains clear that he took the experiencers, his parents, to be the subject of the verb; thus, Examples (7), (9), and (10) would be recorded as subject-object confusion errors. Although the student might not have fully understood the construction of (8), the sentence is nevertheless structurally correct; thus, it was not marked as an error. The frequency of each of these types of errors was calculated for each task at each level.

These data enable a comparison of the performance of the learners from level to level. Based on this information, one may describe "profiles" of each of the four levels with respect to proficiency in the use of psych verbs. These profiles indicate general performance and the types of errors most common to each level.

RESULTS

Figure 1 summarizes the overall performances of each of the classes in each of the tasks. These data help to formulate an answer to the first research question posed: What are the patterns of usage typical of learners attempting to acquire psych verbs such as *gustar, importar, encantar, molestar,* and *interesar*? How do the learning patterns of students at beginning levels compare to those of intermediate students?

Task 1: Describing One's Own Likes and Dislikes

All levels performed this task with a relatively high degree of accuracy ranging from 75% in S1 to 93% in S3. Generally, each level tended to perform slightly better than the

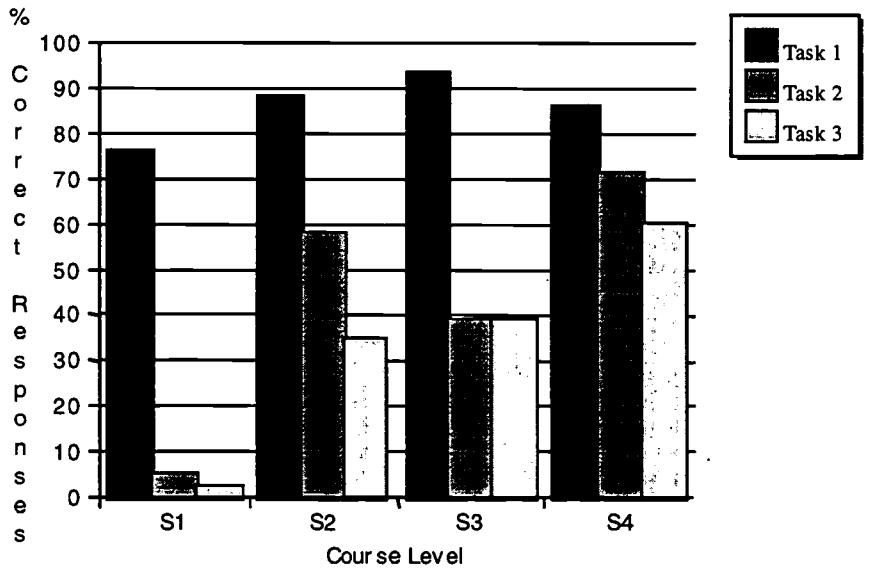


Figure 1. Summary of Overall Performance

previous level with the exception of the S4 class, which scored an 86%. The 7% difference between this class and the S3 class cannot, however, be considered significant since the sampling of responses generated by the S4 class numbered only half that of the S3 and S2 classes. Figure 2 shows the frequency of the three general types of errors that occurred in Task 1: omission of the preposition *a*, object pronoun errors, and verb morphology errors. No errors occurred with the preposition *a* because students avoided the emphatic *a mí* 'to me.' The pronoun *me* 'to me' was consistently used correctly. Conjugation errors were the only significant type of mistake committed in Task 1. It is of interest to note that the most common type of verb morphology error in S1 was that of subject-object confusion, which occurred with a 15% frequency. Such an error may take, for example, the following form:

(11) **Molesto la clase de inglés*
'I bother the English class'

This type of error gradually disappears, however; it occurred with only a 2% frequency in S4.

Task 2: Describing a Friend's Likes and Dislikes

Accuracy in the performance of this task trails behind that demonstrated

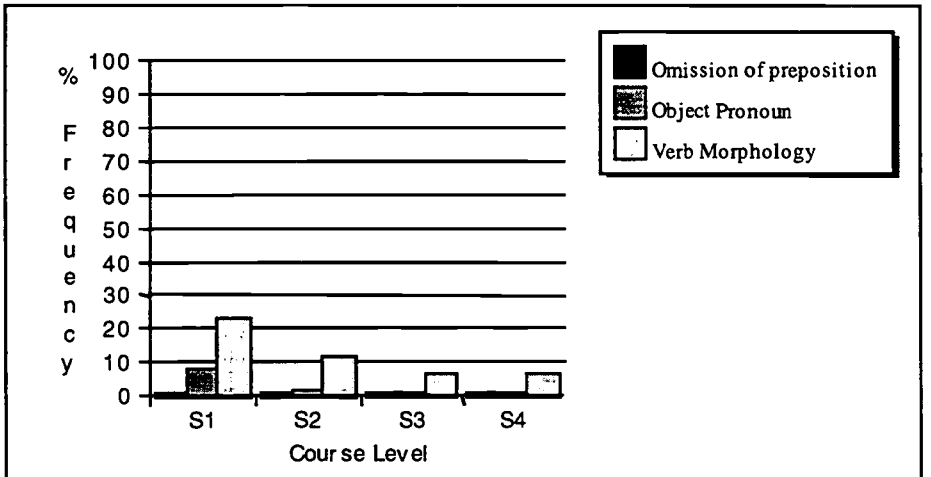


Figure 2. Task 1 Error Types

in Task 1, as seen in Figure 1. Nevertheless, the general trend again seems to be one of improved performance at higher levels, with the exception of the S3 class, which was actually outperformed by the S2 class. There is a marked difference between no proficiency with this construction in S1 to 58% accuracy in S2. This result is not surprising, given that the S1 class had had little exposure to this form or to object pronouns, while the S2 class had been instructed explicitly in third-person constructions with psych verbs other than *gustar* and had reviewed direct and indirect objects. The reason for the drop in performance at the S3 level may be that little explicit attention is given to psych verbs in the chapters of the textbook covered in this semester, nor did the instructor of the course report any review or practice of the construction other than an occasional recycling in communicative activities. The 71% accuracy score of the S4 class suggests that they benefited from the review chapter on *gustar*-like verbs they had recently completed.

Across all levels, the omission of the preposition *a* accounted for the greatest number of errors, as Figure 3 shows. Though the S2 class shows an exceptionally low frequency of this error, the general trend is one of slow improvement as levels increase. Object pronoun errors are less frequent overall and follow a definite trend of improved performance at higher levels. It is interesting to note that omission of the pronoun is the most common type of pronoun error in S1 (24% frequency), but that its occurrence dwindles to 15% in S2 and to approximately 7% in the S3 and S4

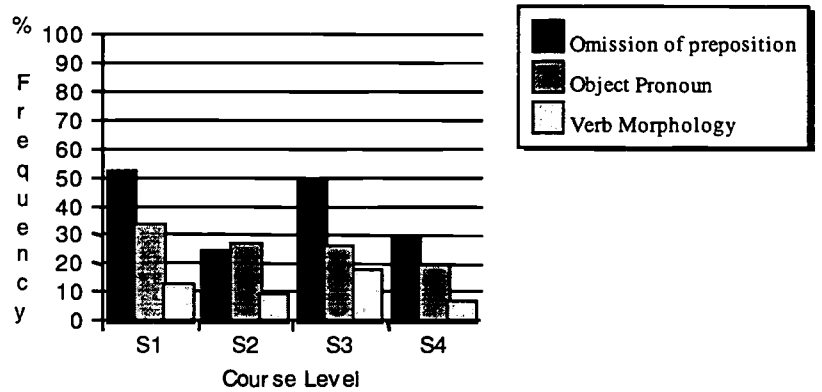


Figure 3. Task 2 Error Types

classes, while at the same time the frequency of *se* interference errors rises from 5% and 8% in S1 and S2 to nearly 20% in S3 before disappearing completely in S4.

Task 3: Describing Your Parents' Likes and Dislikes.

At all levels, accuracy in Task 3 was the lowest of the three. As with the other tasks, however, the trend is one of improved performance at higher levels, with the S4 class reaching a 60% success rate (see Figure 1). Figure 4 shows the frequency of the three general kinds of errors that occurred in Task 3. As in Task 2, omission of *a* is the most persistent error; it seems to occur with about the same frequency here as it did in the second task, and it never improves to less than a 30% rate of appearance. Control of verb morphology and choice of correct object pronoun improve dramatically from 100% error frequency in S1 to 10% and less in S4. Especially noteworthy again is the complete disappearance of *se* for *les* at the S4 level, while it occurred with 51% frequency in S1 and 11% frequency in both S2 and S3.

ANALYSIS

General Performance

We now address the second research question: How can we account for the performance of the learners?

The results generally affirm the expectations outlined in the first section of this paper. It was stated that the Contrastive Analysis Hypothesis would lead us to believe that the

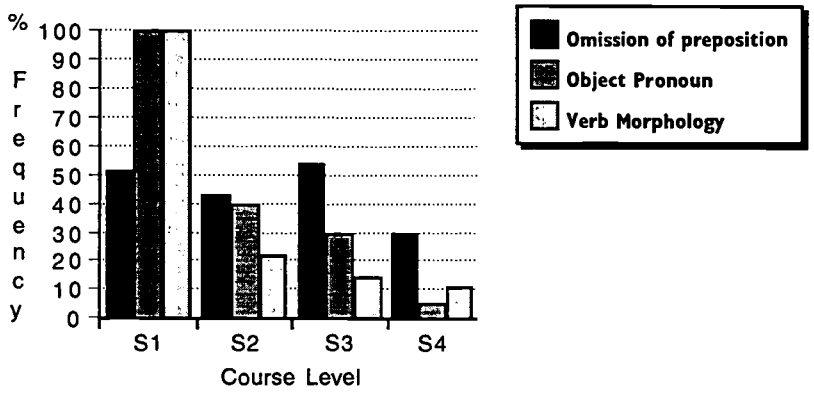


Figure 4. Task 3 Error Types

psych verb structure would be somewhat challenging for the English-speaking learner. The experiment confirms this idea. Though learners seem to come increasingly closer to acquisition of the form in Task 1, results suggest that many learners still will not have acquired the ability to use psych verbs correctly in third-person constructions even after two years of Spanish classes.

The expectation that higher level students would outperform lower level students was generally fulfilled by the data. Two exceptions to this trend did occur, however, at the S4 level in Task 1 and at the S3 level in Task 2. The first of these, as stated, does not constitute a numerically significant decline, while the second may be accounted for by the fact that the third-semester students received no review or practice exclusively dedicated to the concepts, as did the S2 students. It was mentioned that they were occasionally called upon to recycle the psych verb structure in communicative activities, but in such situations learners would more than likely be called upon to comment on their own likes, as in Task 1. This would explain how S3 students outperformed S2 students in Task 1, but not in Task 2.

The study also confirmed the prediction that learners would master Task 1 more easily than the other tasks. It was mentioned that any or all of the frequency, output, or discourse hypotheses could serve to explain this result. *Me gusta* is the form learners hear and produce most often. Furthermore, it is more likely to serve learners' discourse needs better than the other forms; that is, they are more likely to en-

gage in discourses that call upon them to discuss their own likes and dislikes, rather than someone else's. A related idea is that of communicative utility. VanPatten (1987) suggests that as long as a beginning learner believes that a linguistic feature has little communicative value, it will not be used and, therefore, will not be acquired until later. Perdue and Klein (1992) make a similar point that if certain elements of a learner's speech do not grammaticalize, it is because they are meeting their communicative needs without mastery of those elements and feel little need to change. Either of these versions of the communicative value idea would serve well to explain why the learners in this study show more mastery over Task 1 than Tasks 2 or 3.

Stages of Interlanguage Development

The third question posed at the outset of this study concerned the possibility of stages of interlanguage development. While a longitudinal study would be necessary to affirm any proposed order of acquisition, this pilot study may at least suggest a possible outline of certain stages learners pass through as they develop their analysis and control of psych verb constructions.

While performance analysis gives us a general idea of where learners are in their development toward acquisition, error analysis may help reveal some of the cognitive processes that are going on at the same time. Figure 5 shows the frequency in all three tasks of four different types of errors that were not shown in Figures 2 through 4 because they are subcategories of those three general error types. The first two types of errors listed in the legend, omission of the object pronoun and subject/object confusion, can be considered transfer-type errors. According to the Naive Lexical Hypothesis (Lafford and Ryan, 1995), English-speaking beginners attempt to build their Spanish sentences word for word from the English direct structure. Thus they produce sentences like *Ellos encantan su perro* in an attempt to express 'They love their dog.' Since the English utterance contains no object pronoun, the beginners omit it. They have also transferred English S-V-O word order, which produces in the Spanish a subject-object confusion error. Thus, these two types of errors can be understood to be symptomatic of L1 transfer.

The other two types of errors, the use of the reflexive pronoun *se* instead of the indirect object pronouns *me*, *le*, or *les*, and the use of *le* instead of *les* and vice-versa, are not related to transfer because these pronouns, for the most part, do not have English equivalents when used in psych verb constructions. They indicate, rather, that the learner is trying to develop hypotheses about how features particular to Spanish are used correctly. They may be called, then, developmental errors.

This distinction between transfer and developmental errors may help to define different cognitive stages of development in the learning of psych verbs. Figure 5 shows that errors prompted by L1 transfer occur with greater frequency at the S1 and S2 levels, while in S3 and S4, developmental errors are more fre-

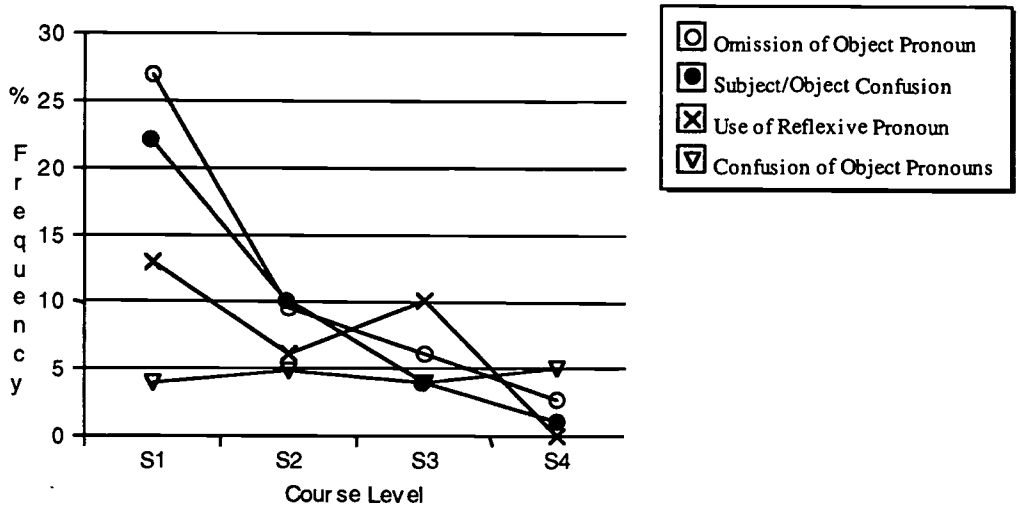


Figure 5. Frequency of Transfer and Developmental Errors

quent and transfer errors decline. In other words, when first confronted with the psych verb construction, the S1 students' interlanguage could be described in terms of Selinker's (1972) restructuring continuum, wherein the learners transfer L1 parameter settings to the L2. This stage could then be called the "transfer stage." In S2, it seems that the actual restructuring begins to take place: transfer error frequency decreases dramatically as the learners make adjustments to their interlanguage where they see that the L2 is different from the L1. This period might be called the "restructuring stage." At the S3 level we see that transfer errors become less frequent than developmental errors; here, then, the interlanguage resembles Corder's (1978) continuum of development, wherein learners rely more on L2 input than on L1 transfer to help set the parameters of the L2. This phase, then, is the first developmental stage. A significant finding at the S4 level was that *se* errors disappeared completely. This development would seem to mark an important step forward; students have learned to separate the *se* they use with reflexives, reciprocals, and double object pronouns from the *le* and *les* that are pertinent to the psych construction. Since learners at this level seem to have progressed developmentally, we may call this the second developmental stage.

Level Profiles

These stages may be combined with the patterns of usage described earlier to give us four profiles that

characterize early levels of progress toward acquisition of the psych verb construction. The descriptions will also serve to summarize the most significant findings of the study in general.

Transfer Stage: Learners at this level demonstrate proficiency with the *me gusta(n)* construction, but none with the *le/les gusta(n)* structures. A number of errors stem from attempts to transfer L1 parameters directly to the L2.

Restructuring Stage: Learners at this stage show increased proficiency with *me gusta* and a marked improvement over first-stage learners in the use of *le/les gusta*. They begin to realize that they cannot simply map structures from their L1 onto the L2. They start to adjust the mistakenly transferred parameters they had previously tried to follow. As they search for concepts to replace their previous hypotheses, they begin paying more attention to L2 input.

Developmental Stage I: Frequency of transfer errors is greatly reduced. Learners reach even greater mastery over the *me gusta* form, and they experiment actively with features particular to the L2 in an effort to form hypotheses concerning the *le/les gusta* forms. Recall may fail them here if there is no regular reinforcement of this structure. *Se* interference increases.

Developmental Stage II: When *se* is ruled out as an alternative to the object pronouns used in the psych verb structure, a new developmental stage begins. Control of verb morphology and choice of object pronoun is as high as it has ever been, though hypothesis formation based on L2 input continues. Transfer errors nearly disappear. Acquisition of the psych verb construction, at least as it is used in the first person singular and the third person, seems within reach.

LIMITATIONS

Of course, the findings of this pilot study are tentative because they were gathered synchronically; a longitudinal study following individual learners through their first four semesters of Spanish would provide stronger evidence of any order of acquisition observed. Oral interviews would also be a better indicator of degree of acquisition; the form-focused written tasks used for this study may have prompted greater control and monitoring than would be possible in a more meaning-focused interview. It would also allow for negotiation of meaning, which would ensure more uniform generation of samples: many students misunderstood the directions to the second task in this experiment and wrote about their friends' collective preferences rather than about one friend's likes and dislikes. Finally, it would be of interest to compare this study's results with those of an investigation of learners in natural environments to see if there is any relation between their 'stages' and those proposed here.

IMPLICATIONS AND CONCLUSION

Given that learners at all levels demonstrate a relatively high level

of competence with the first-person singular form of the Spanish psych verb construction and a relatively low level of competence with the third person forms, instructors may wish to consider how much reinforcement is being given to a given task type at each level. Though it is understandable that teachers may naturally call on students to perform tasks in the first person since one's expression of one's own likes or dislikes is a common communicative function, instructors may want to consider giving more reinforcement to third person forms as well. If we think of *me gusta* as the unmarked, canonical form of the psych verb structure, and of *le/les gusta* as the marked forms, then perhaps we should spend more time reinforcing the latter, if, as Guntermann (1992) and Rutherford (1982) maintain, instruction in more marked features can facilitate acquisition of less marked features, while instruction in less marked features may result in learners simplifying their interlanguages.

In the third person constructions, errors in the use of indirect objects and elements related to them, such as the preposition *a* 'to,' are more persistent than verb morphology errors. This seems to suggest that, whether as a result of instruction or of their own independent cognitive processes, learners seem to prioritize the elements in the Spanish psych verb construction. They attempt to control the most communicative and most easily translatable items first, such as the effect (verb) in question and its cause (the subject). Learners appear to give less attention, at least initially, to less easily translatable, more subtle forms, such as indirect object pronouns and the preposition *a*, which operate on L2, rather than L1, principles.

Such findings lead to certain questions regarding the instruction of the psych verb construction. First, is it possible to determine precisely what makes learners shift from a strategy that attempts to acquire L2 forms by restructuring familiar L1 forms to a strategy that recognizes the L2 forms as different from the L1 and seeks knowledge of its principles? Second, is it desirable for an instructor to attempt to facilitate such a shift in the learner's consciousness? If so, at what moment in the learner's progress should the instructor make such an attempt? If not, are there other ways in which instructors should adjust their sequence of instruction of the psych verb structure, given the sequence of learner acquisition hypothesized in this study? Answers to these questions would further enable instructors to better facilitate native English speakers' acquisition of the Spanish psych verb construction.

NOTES:

1 These terms, borrowed from Whitley (995), are defined in the next section.
2 VanPatten (1987) discusses a similar type of L1 transfer in the L2 acquisition of *ser* and *estar*.
3 Throughout this study, an asterisk indicates that the example provided is grammatically incorrect.
4 These theorists offer various interpretations of psych verb constructions designed to show how

the configurations either challenge or support the Universal Alignment Hypothesis, which, as stated by Perlmutter and Postal (1984), maintains that "there exist principles of universal grammar which predict the initial relation borne by each nominal in a given clause from the meaning of the clause" (p. 97).

REFERENCES

Belletti, A., & Rizzi, L. (1988). Psych-verbs and theta theory. *Natural Language and Linguistic Theory, 6* (3), 291-352.

Bouchard, D. (1992). Psych constructions and linking to conceptual structures. In P. Hirschbühler & K. Koerner (Eds.), *Romance languages and modern linguistic theory: Papers from the 20th Linguistic Symposium on Romance Languages* (pp. 25-44). Amsterdam: John Benjamins.

Corder, P. (1978). Language-learner language. In J. Richards (Ed.), *Understanding second and foreign language learning* (pp. 71-93). Rowley, MA: Newbury House.

Ellis, R. (1990). *Instructed second language acquisition.* Cambridge: Blackwell.

Guntermann, G. (1992). An analysis of interlanguage development over time: Part II, *ser* and *estar*. *Hispania, 75* (5), 1294-1303.

Herschensohn, J. (1992). A postfunctionalist perspective on French psych unaccusatives. In William Ashby (Ed.), *Linguistic perspectives on the romance languages: Selected papers from the 21st Linguistic Symposium on Romance Languages* (pp. 239-247). Amsterdam: John Benjamins.

Hopper, P., & Thompson, S. (1980). Transitivity in grammar and discourse. *Language, 56* (2), 251-299.

Lafford, B., & Ryan, J. (1995). The acquisition of lexical meaning in a study abroad context: The Spanish prepositions *por* and *para*. *Hispania, 78* (3), 528-547.

Masullo, P. (1992). Antipassive constructions in Spanish. In P. Hirschbühler & K. Koerner (Eds.), *Romance languages and modern linguistic theory: Papers from the 20th Linguistic Symposium on Romance Languages* (pp. 175-194). Amsterdam: John Benjamins.

Perdue, C., & Klein, W. (1992). Why does the production of some learners not grammaticalize? *Studies in Second Language Acquisition, 14* (3), 259-272.

Perlmutter, D., & Postal, P. (1984). The 1-advancement exclusiveness law. In D. Perlmutter & C. Rosen (Eds.), *Studies in relational grammar 2* (pp. 81-125). Chicago: University of Chicago Press.

Ross, J. (1973). A fake NP squish. In C. J. Bailey & R. Shuy (Eds.), *New ways of analyzing variation in English* (pp. 96-140). Washington: Georgetown University Press.

Rutherford, W. (1982). Markedness in L2 acquisition. *Language Learning, 32* (1), 85-108.

Saltarelli, M. (1992). The subject of psych-verbs and case theory. In P. Hirschbühler & K. Koerner (Eds.), *Romance languages and modern linguistic theory: Papers from the 20th Linguistic Symposium on*

Romance Languages (pp. 251-267). Amsterdam: John Benjamins.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics, 10* (3), 209-230.

VanPatten, B. (1987). Classroom learners' acquisition of *ser* and *estar*. In B. VanPatten, T. Dvorak, & J. Lee (Eds.), *Foreign language learning: A research perspective* (pp. 101-123). Cambridge: Newbury House.

Whitley, M. (1995). *Gustar* and other psych verbs: A problem in transitivity. *Hispania, 78* (3), 573-585.

The Relationship Between the Production and Perception of L2 Spanish Stops

MARY L. ZAMPINI, The University of Arizona

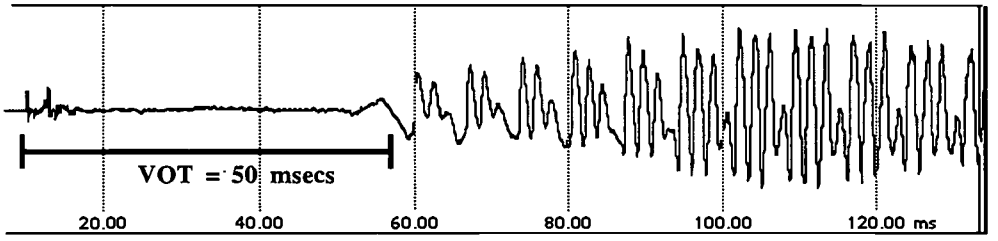
This paper explores the relationship between the second language (L2) production and perception of the Spanish stop consonants /p/ and /b/. An experiment was conducted that collected data on adult English-speaking learners' production of Spanish /p, b/ in a sentence context. The same learners also completed a series of perception experiments that examined their perceptual boundary between /p/ and /b/ as reflected by changes in voice onset time (VOT). The mean VOTs produced by the learners in the production experiment were compared to their perceptual VOT boundary. The results do not reveal a strong correlation between learners' perceptual capabilities and production of the L2 Spanish stops. In particular, while some learners' L2 productions approximated those of native Spanish speakers, their perceptual boundaries were similar to boundaries for monolingual English speakers, and vice versa.

INTRODUCTION

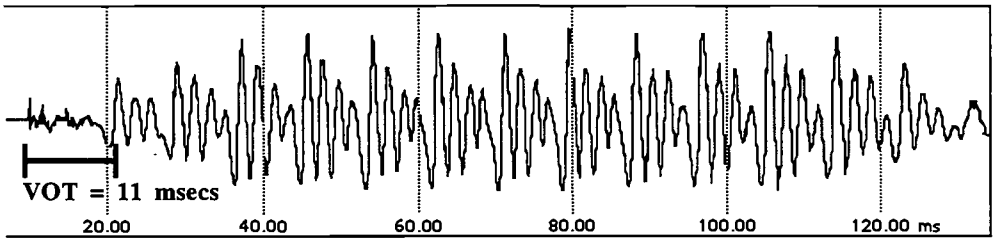
An important issue of second language (L2) pronunciation and phonological acquisition is whether the ability to perceive accurately a particular L2 contrast (e.g., the contrast between Spanish /p/ and /b/) is necessary for proper articulation of the contrasting phones. Flege's (1992) Speech Learning Model, for example, suggests that the inability to recognize or (re)categorize perceptual distinctions limits accurate L2 production. This paper will address that question with regard to the acquisition of the Spanish voiced and voiceless stop consonants by native English speakers.

The acquisition of the Spanish stops by English speakers is problematic for several reasons. While both languages contain a series of stop phonemes distinguished by voicing—voiceless /p t k/ contrasted with voiced /b d g/—the phonetic realization of the voiced and voiceless stops in the two languages differs in important respects. First, English /p t k/ are known as long-lag voiceless stops: they are realized with a relatively long voice onset time (VOT), which refers to the amount of time that elapses between the release burst of the stop and the onset of vocal fold vibration. This lag causes the aspiration that typically accompanies the production of /p t k/ in English, especially in word-initial position. The voiced stops, on the other hand, have short VOT values and are hence considered short-lag voiceless stops. This contrast is illustrated by the waveforms in Figures 1A and 1B, which show the first syllable of the words *poker* and *both*, produced by a native speaker of English, along with their respective VOT values.

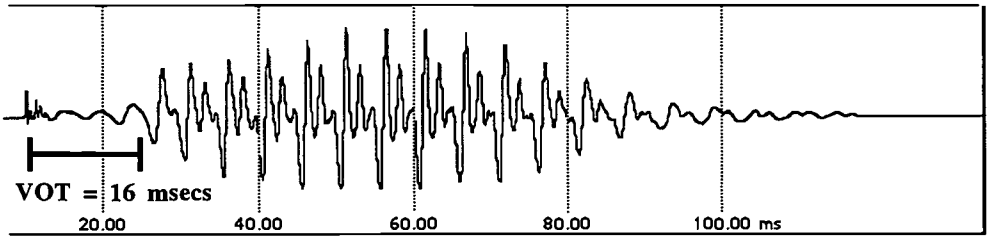
Unlike those in English, Spanish voiceless stops have short VOT values, while the voiced stops are realized with voicing lead (or prevoicing) in which



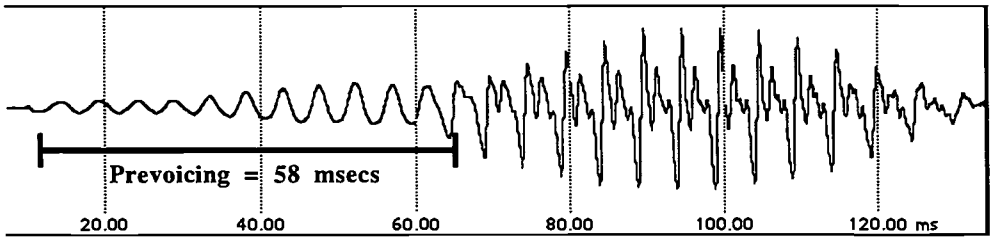
A. English *poker*



B. English *boath*



C. Spanish *poca*



D. Spanish *boca*

Figure 1. Sample Waveforms for English /p/ and /b/ and Spanish /p/ and /b/

the vocal folds begin vibrating before the release burst. Waveforms illustrating this distinction for the Spanish words *poca* and *boca* (produced by a native speakers of Spanish) appear in Figures 1C and 1D. Thus, from a phonological perspective, /p/ and /b/ differ with regard to the feature [voiced] in both languages. From a phonetic perspective, however, Spanish /p/ is more like English /b/ in that both belong to the short-lag stop category and are characterized by short VOT values, as shown in Figures 1B and 1C of Figure 1. To illustrate the differences between the two languages further, Table 1 provides the average VOT production values for native speakers of Spanish and English, as well as the range of VOT values found, reported in a classic study by Lisker and Abramson (1964).

Given the phonetic differences between the two languages, the challenge for English-speaking learners of Spanish becomes clear: they must reorganize the phonetic categories of the voiced and voiceless stop phonemes so as to reflect those of the target language. The learner must shorten the relative VOT of /p/ during Spanish production, so that this phone falls within the target range of a short-lag stop, and also eliminate voicing lag from the production of Spanish /b/, so that this phone becomes prevoiced. In addition, the phonetic category overlap between Spanish /p/ and English /b/ has

Table 1
Mean VOT Measurements (in msecs) of English and Spanish Stops (Lisker & Abramson, 1964)

[Positive values indicate voicing lag; negative values indicate prevoicing]

English			Spanish		
Stop	Mean VOT	VOT Range	Stop	Mean VOT	VOT Range
/p/	58	20 / 120	/p/	4	0 / 15
/t/	0	30 / 105	/t/	9	0 / 15
/k/	80	50 / 135	/k/	29	15 / 55
/b/	1	0 / 5	/b/	-138	-235 / -60
/d/	5	0 / 25	/d/	-110	-170 / -75
/g/	21	0 / 35	/g/	-108	-165 / -45

important implications for Spanish L2 speech perception. Given their phonetic similarities, learners may confuse Spanish /p/ for /b/ perceptually and must therefore adjust their perceptual categories (or boundaries) to reflect the Spanish system in order to avoid confusion and promote comprehension.

METHODS

The differences in the phonetic realization of the Spanish and English stop consonants with regard to VOT provide the motivation for the current study. In particular, the study addresses the following research questions:

1. Do learners acquire the appropriate phonetic categories with regard to the Spanish voiceless and voiced stops? If so, how do they manipulate the different acoustic cues of the speech signal in order to achieve the necessary distinction?

2. What effect does formal training in phonetics have on the acquisition of the Spanish stops?

3. Is there a relationship between the perception and production of the Spanish stops? That is, do learners with native-like pronunciations of the Spanish stops also show evidence of native-like boundaries between the voiceless and voiced phonemes in perception?

While all three questions have some bearing on the results to be discussed, the present work focuses primarily on the issues raised by the third question, that of the relationship between production and perception in L2 Spanish acquisition.

In order to address the stated research questions, an experimental study was designed to examine both the production and perception of word-initial Spanish stop consonants by L2 learners whose native language was English. The participants were enrolled in an advanced undergraduate Spanish phonetics course at the University of Arizona. Thirteen learners volunteered for the study and were asked to complete a series of production and perception experiments in a speech laboratory setting at several points throughout the semester.

For the production portion of the experiment, each learner was recorded during four repetitions of an English or Spanish sentence, each containing a target word that began with a stop consonant. The English productions were obtained once during the second week of the semester, while Spanish productions were obtained three times. The first Spanish recording occurred during the third week of the semester, one week following the English production session. At that point, the learners had not yet begun to study the articulation of individual Spanish phonemes; instead, they had studied basic concepts of phonetics, as well as Spanish syllable structure and syllabification. The second Spanish recording session took place three weeks later, immediately after the Spanish voiceless stops had been studied in class. The text used for the presentation and practice of these phones was Barrutia and Schwegler (1994). In this text, the difference be-

tween the Spanish and English voiceless stops is not described in terms of short-lag vs. long-lag categories and VOT; however, the text does tell the learner to try to avoid aspiration of /p t k/ when speaking Spanish by maintaining greater muscular tension of the articulators and vocal tract during production. The final Spanish production experiment took place near the end of the semester--during the fifteenth week of class (and nine weeks after the second session). The sentences that the learners read were the same for all recording sessions. A total of 32 sentences were used with target words containing a variety of word-initial consonants. The data and results discussed in the present study come from the target stops that appear in sentences like those in Table 2.

Table 2
Sample Sentences Used in the Production Exercises

English:
Please say the word *paces* to me.
Please say the word *bases* to me.

Spanish:
Diga la palabra peso *por favor.*
'Say the word *peso* please.'

Diga la palabra beso *por favor.*
'Say the word *beso* please.'

Once all the production data had been gathered, the sentences were digitized, and the voiceless closure intervals and VOT of the target stops were measured from the digitized waveforms using SoundEdit 16. Voiceless closure interval refers to the duration of closure before the release of the stop consonant that is characterized by a lack of vocal cord vibration. Prevoicing of the voiced stops was also measured where applicable; in such cases, the duration of prevoicing was measured as a negative VOT (see Table 1).

Finally, spirantized variants of the voiced Spanish stops were noted as well, and these tokens were eliminated from the analysis. Spirantization is a process whereby Spanish /b d g/ are realized as approximants in certain phonetic contexts. The stop allophones generally appear after a nasal consonant and in phrase-initial position (and /d/ appears as a stop after laterals, as well), whereas the spirantized allophones appear elsewhere. Since the spirants do not have the closure, release burst, or VOT associated with stops, they were not included in the present analysis.

For the perception portion of the study, the learners listened to computer-edited versions of the English and Spanish nonsense words, *pada* and *bada*. These nonword tokens were natural speech tokens produced by a fluent English-Spanish bilingual, and they were edited so as to vary from 40 msecs of prevoicing to 56 msecs of voicing lag at approximately 5-msec intervals. The VOT continuum consisted of a total of 20 tokens for each language. The learners listened to both English and Spanish versions of the words presented randomly and indicated whether each one began with /p/ or /b/ by pressing the appropriate but-

ton on a response box in the testing room. Each token of the continuum appeared 10 times throughout the experiment for a total of 200 responses per language. This experiment was also carried out three times during the same three weeks as the Spanish production exercises: Weeks 3, 6, and 15 of the semester. In addition, 15 monolingual speakers of English and 12 Spanish-English bilinguals completed the perception experiment, so as to provide a basis of comparison for the L2 learners' perception data. A mean VOT perceptual boundary for /b/ and /p/ was determined for each subject group based upon the percentage of /b/ (or /p/) responses. Separate boundaries were obtained for the English and Spanish versions of the tokens.

RESULTS AND DISCUSSION

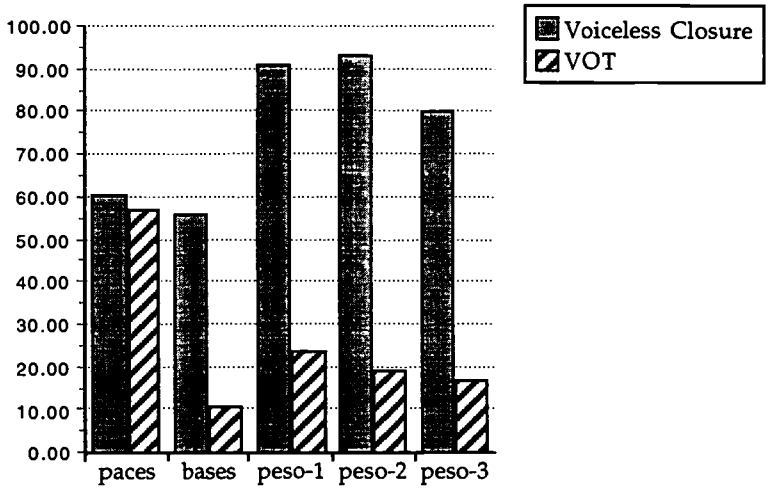
In the discussion of the results, the production results will be briefly described first, followed by a description of the perception results. Finally, the relationship between the learners' production and perception of Spanish /p/ and /b/ will be discussed.

Production Results

The production results are presented in Figures 2 and 3. These figures show the mean voiceless closure interval and VOT values for the English tokens *paces* and *bases* and the Spanish tokens *peso* ("weight") and *beso* ("kiss") for the L2 learner group as a whole.

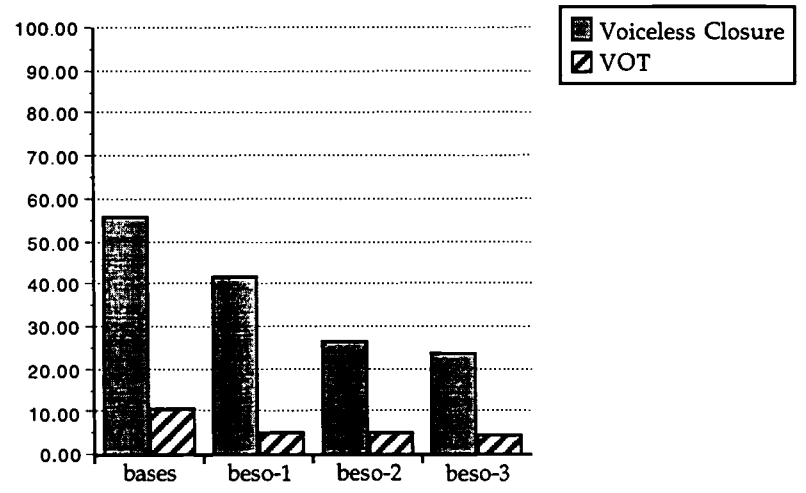
Consider first the information in Figure 2. As shown, the learners produced Spanish /p/ with VOT values that approach the average values reported for native Spanish speakers shown in Table 1. Although the learners' VOT values were somewhat longer than those of native Spanish speakers, they were still significantly shorter than their corresponding VOT values for English /p/. The learners also produced VOT values for Spanish /p/ that were similar to English /b/. Statistical testing of these data revealed that although the difference between the average VOT value of English /b/ in Figure 2 proved significant from that of *peso-1* and *peso-2*, there was no significant VOT difference between *bases* and *peso-3*. This indicates that by the end of the semester, the learners equated the L1 short-lag category, /b/, with the L2 short-lag category, /p/, at least with respect to VOT. Figure 2 also shows, however, that the learners distinguished Spanish /p/ from English /b/ through significantly longer closure intervals of the Spanish phone.

As for Spanish /b/, Figure 3 shows that the learners produced these tokens with somewhat shorter VOT values than in English; these differences, however, did not prove significant. Furthermore, the overall positive VOT means for L2 Spanish /b/ across all three sessions indicate that the learners failed to produce these tokens with prevoicing. As mentioned above, prevoicing is measured as a negative VOT; therefore, if the learners had prevoiced consistently (or learned to prevoice over the course of the semester), one would expect an overall negative VOT average. An examination of the individual data, however, revealed only two prevoiced /b/'s in the first session, three during the



peso-1: 1st session (Week 3) •*peso-2*: 2nd session (Week 6)*peso-3*: 3rd session (Week 15)

Figure 2. Mean Voiceless Closure and VOT Values (in msecs)— Learner Production of English /p, b/ and Spanish /p/



beso-1: 1st session (Week 3) • *beso-2*: 2nd session (Week 6) • *beso-3*: 3rd session (Week 15)

Figure 3. Mean Voiceless Closure and VOT Values (in msecs)— Learner Production of English and Spanish /b/

second session, and four during the third (out of a total of 56 tokens with word-initial /b/ in each session). Thus, the prevoicing associated with the Spanish voiced stops appears to take longer to acquire than the short-lag VOTs of the Spanish voiceless stops. (See Zampini (1998) for a detailed discussion of the production results presented in Figures 2 and 3, including a more detailed description of the statistical analyses.)

Perception Results

Turning now to the results of the perception experiments, consider first the information in Table 3, which shows the average VOT boundary that separates /p/ from /b/ perceptually for the L2 learners and the two control groups; Figure 4 presents the same information in graphical form.

First, it is interesting to note that all subject groups showed a consistent difference with regard to the VOT boundary for the Spanish and English versions of the nonsense words; namely, the Spanish boundary was consistently shorter than that for the English tokens. These two token types were identical in their acoustic characteristics prior to the release burst, so the voicing decision must

Table 3
Mean VOT Perceptual Boundaries (in msecs) Between /p/ and /b/

[Positive value indicates boundary in voicing lag range; negative value indicates boundary in prevoiced range.]

Group	English Tokens	Spanish Tokens	Mean
Monolingual English	14.5	2.77	**8.635**
Spanish/English Bilinguals	6.12	-8.96	**-1.42**
L2 Learners - 1st Session	8.22	-3.65	**2.285**
L2 Learners - 2nd Session	11.56	-7.59	**1.985**
L2 Learners - 3rd Session	11.85	-5.15	**3.35**

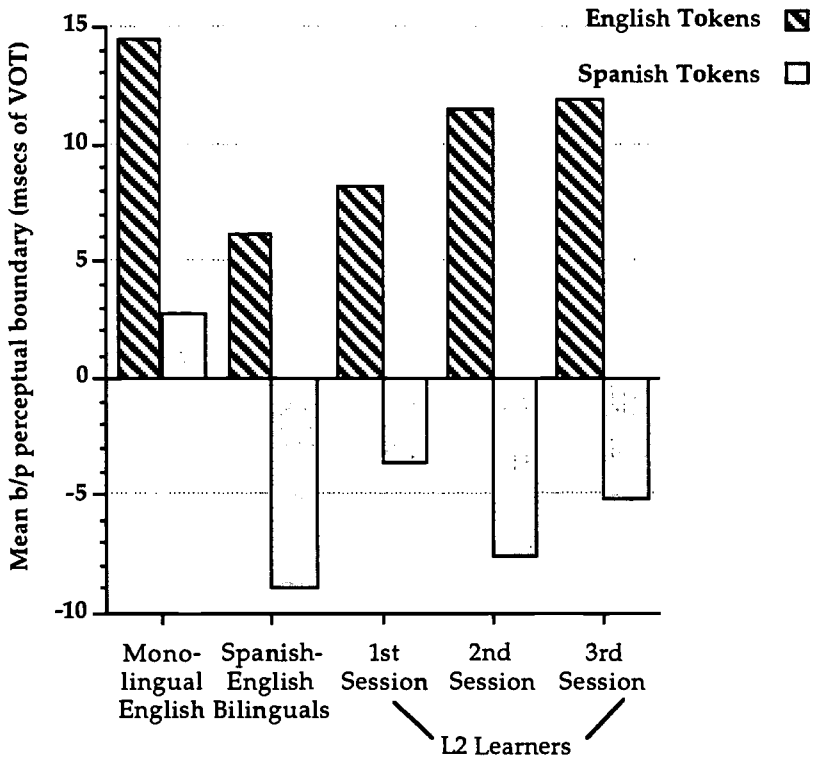


Figure 4. Mean VOT Perceptual Boundaries between /p/ and /b/

have been determined by aspects of the speech signal following the release burst; however, a discussion of the reasons for these differences is beyond the scope of the current paper.

Statistical analyses on the mean VOT boundaries revealed several significant effects. Consider first the boundaries for the English tokens in Figure 4. Analyses of these tokens revealed three significant differences. First, the English boundary difference between the monolingual English and Spanish-English bilingual speakers was significant. Second, there was also a significant difference in the English token boundaries for the monolingual English speakers and the L2 learners in the first L2 session, but not the second or third. Finally, the differences in English boundary between the L2 learners and the Spanish-English bilinguals were not significant for the first L2 session, but were for both the second and third sessions.

The Spanish token perceptual boundaries in Figure 4 also showed

several significant effects. First, the monolingual English speakers' Spanish boundary was significantly longer than that of the Spanish-English bilinguals; this pattern is similar to the one found for the English data of these two groups. Second, there was no significant difference in the Spanish perceptual boundaries of the monolingual English speakers compared to the L2 learners in the first session; however, the differences between these two groups in the second and third L2 sessions did prove significant. Third, none of the Spanish boundary differences between the L2 learners and Spanish-English bilinguals were significant. Finally, an examination of just the L2 learners' perceptual data revealed a significant difference in their English VOT boundaries between the first and second, and first and third, sessions, but not between the last two sessions. For the Spanish tokens, the only significant difference across sessions occurred between the first and second L2 sessions.

To summarize, the L2 learners started out with an English perceptual boundary that was significantly shorter than that of their monolingual English counterparts, but not significantly different from that of Spanish-English bilinguals; thus, their English boundary was more Spanish-like. These boundaries, however, shifted after training in the Spanish voiceless stops took place and became significantly longer and more English-like; this shift was sustained through the end of the semester. As for the Spanish perceptual boundaries, the L2 learners started out somewhere in between the two control groups, with a VOT boundary that was not significantly different from either group. They shifted toward more Spanish-like boundaries after training took place, however, as evidenced by the changes in the Spanish VOT boundary during the second session. Although this shift does not appear to be wholly sustained through the end of the semester, the L2 learners' Spanish boundary of the third session remained significantly different from the corresponding boundary of the monolingual English speakers, but not from the boundary of the Spanish-English bilinguals. Taken together, the changes in the L2 learners' English and Spanish perceptual boundaries indicate an attempt to maintain a clear distinction between the two languages by maximizing the perceptual distance between them. Thus, rather than showing evidence for one merged perceptual boundary between the voiced and voiceless stops regardless of language mode, the results suggest that the learners have two separate perceptual boundaries—one for each language—and that these boundaries become even more distinct with training.

The Relationship between Production and Perception

Finally, having discussed the individual production and perception results, we may now turn to the central issue of the paper: the relationship between the L2 learners' production and perception of the Spanish stops. If a positive relationship exists, one would expect that those learners who show short perceptual boundaries will also exhibit

short VOT production values, while those with longer perceptual boundaries will likewise exhibit longer VOT production values. Such an idealized relationship between production and perception is depicted in Figure 5.

In a similar fashion, if native-like perception in L2 is necessary for (or precedes) accurate production, one would expect that those learners with long perceptual boundaries will also have long VOT production values. Learners with short perceptual boundaries, however, may or may not have correspondingly short VOT production values. That is, if perception precedes production, learners that show accurate perceptual boundaries could still exhibit long production values if they are in a stage of acquisition in which production has not yet begun to change.

To examine this issue with regard to learner performance, correlations were obtained between the perceptual boundaries of a particular session and the corresponding productions of *peso* and *beso* from the same session. This information appears in Figures 6A - 6F.

In each of the graphs in Figure 6, the points represent the intersection of the Spanish perceptual boundary and mean Spanish VOT production value for each individual learner. None of the graphs approach the kind of idealized

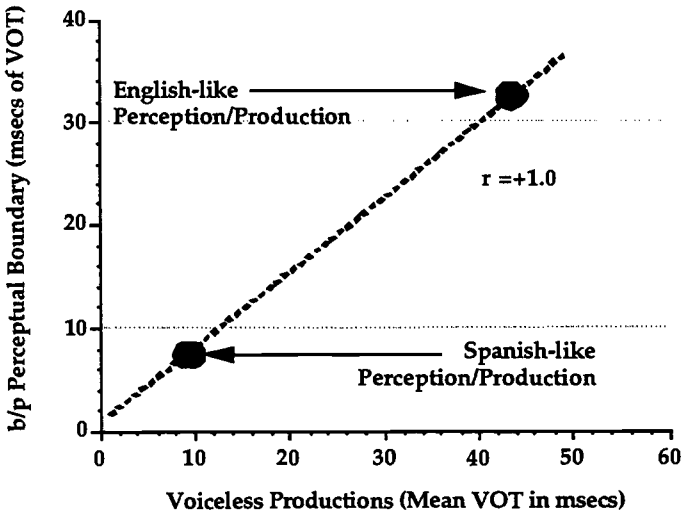


Figure 5. Idealized Relationship between Production and Perception

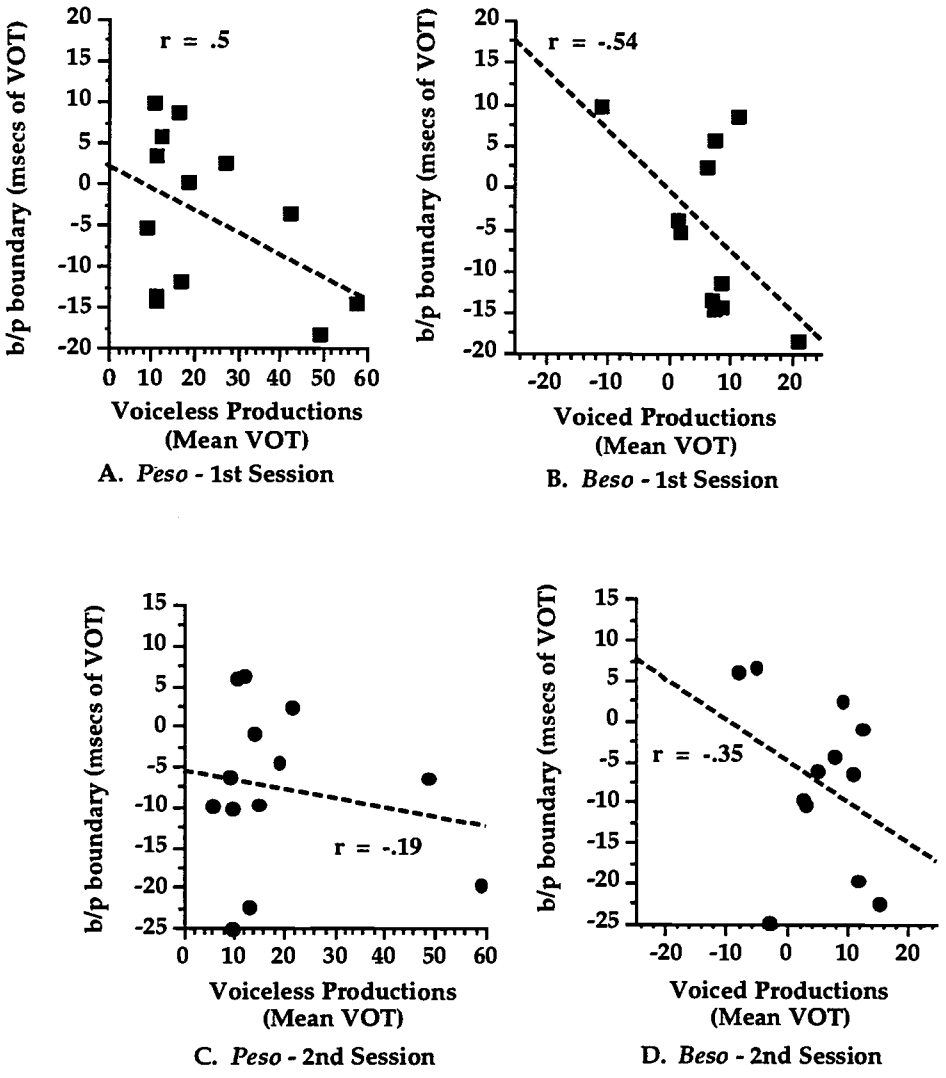


Figure 6. Correlations Between L2 Spanish Production and Perception (cont.)

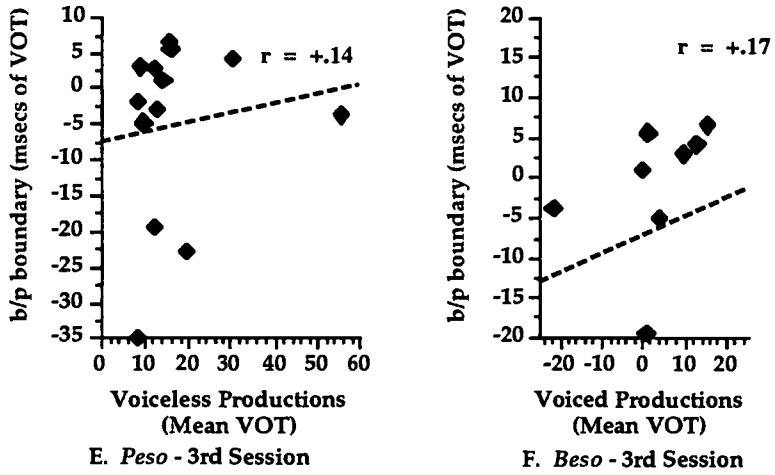


Figure 6. Correlations between L2 Spanish Production and Perception *(concluded)*

relationship shown in Figure 5. For example, consider the first session correlation data for the Spanish word *peso* in Figure 6A. Recall that if a positive relationship existed between production and perception, one would expect that learners with short VOT production values would also have short VOT perceptual boundaries. As seen in Figure 6A, however, those learners with the longest perceptual boundaries also have some of the shortest production values. This trend is also seen for the *peso* data in the second and third sessions, as shown in Figures 6C and 6E. These results, therefore, fail to support the hypothesis proposed by Flege's (1992) Speech Learning Model that inaccurate L2 perception will limit L2 production and suggest instead that perception does not necessarily precede production. In fact, the *peso* data in Figures 6A, 6C, and 6E appear to support an opposing hypothesis, namely, that L2 production may in some cases precede perception. That is, it may be the case that learners do not begin to adjust perceptual boundaries until they have attained accurate production categories. That this might be so is illustrated in Figure 7, which shows a predictive distribution of data for the opposing hypotheses.

If perception precedes production, one might expect a range of data along the horizontal axis, as depicted in Figure 7. Learners with short, Spanish-like perceptual boundaries may or may not exhibit correspondingly short production values, depending upon their stage of acquisition. If production precedes perception, on the other hand, one might expect a range of data along the vertical axis, since learners with short production values may or may not exhibit correspondingly short perceptual boundaries. This pattern

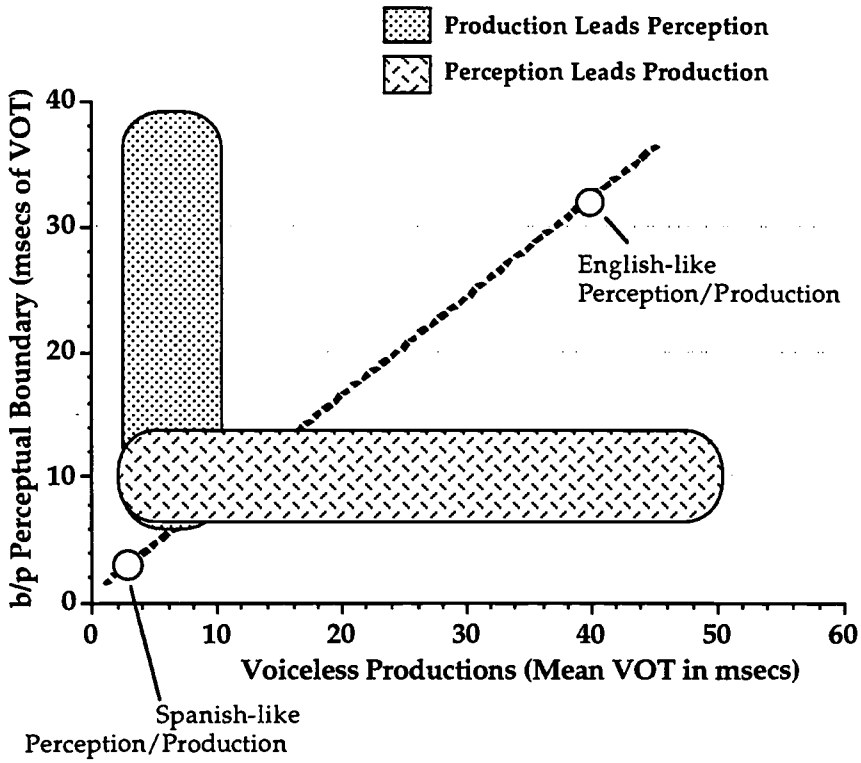


Figure 7. Relationship between Production and Perception - Two Hypotheses

would again depend upon each individual's stage of acquisition. Returning now to Figures 6A, C, and E, the overall spread of the individual points corresponds more closely to the predictions made by the hypothesis that production precedes perception. Thus, it appears that some learners learn to make the phonetic category substitution for Spanish /p/ before they make corresponding changes in perception. This result may not seem too surprising, since the phonetic category that they must learn for Spanish/p/, that of a short-lag stop, is one that already exists in the learners' first language. The phonetic category that learners must acquire for Spanish /b/, on the other hand, that of a prevoiced stop, does not exist in English. As a result, the substitution in production may take longer, whichcould, in turn, affect the interaction between production and perception. Consider, for example, the information in Figures 6B, 6D, and 6F, which correlate the Spanish perceptual boundary with the average VOT values for /b/. The correlations in these graphs are more scattered and do not appear to support either of

the predictions made by the hypotheses illustrated in Figure 7.

To summarize, the correlation data do not provide evidence for a positive relationship between the L2 production and perception of the Spanish stops. While some learners' L2 productions approximated those of native Spanish speakers, their perceptual boundaries were similar to boundaries for monolingual English speakers, and vice versa. In addition, the correlation data for Spanish /p/ suggest that production may precede perception, at least for this category. This indication does not necessarily mean, however, that inaccurate production will limit accurate perception; rather, it simply implies that the two may not be mutually dependent processes. The observed variation in the correlation data for Spanish /b/ further suggest that production and perception may be independent processes (at least for certain stages of acquisition or for certain types of phones), since no clear interaction was found.

These results have important implications for studies of both second language acquisition and speech processing. For example, although the results do not show a positive correlation between the perception and production with regard to VOT, there are other acoustic cues of the stop consonants that may play an influential role in the acquisition process. It was observed in Figure 2, for example, that learner production of the Spanish voiceless stops had significantly longer closure intervals than those for either the voiceless or voiced English stops; this difference provided a means of distinguishing the Spanish short-lag phones /p t k/ from the corresponding English ones, /b d g/. Given that the learners have learned to manipulate closure interval in order to achieve a given distinction, the interaction between VOT and closure may prove more important in L2 Spanish production and perception than either one of these two acoustic cues alone. Thus, research in both L2 production and perception is needed to gain additional insight into the relative importance of the different acoustic cues in second language acquisition, as well as into ways in which the weighting of these cues changes over time.

CONCLUSION

To conclude, this study has presented evidence for the acquisition of L2 Spanish stops by native English speakers as demonstrated by changes in production and perception over the course of the semester. Analyses that examined these processes individually showed significant changes toward Spanish-like production and perception categories over the course of the semester. Correlations of the production and perception results with regard to VOT, however, did not reveal a strong relationship between the two. The correlation data for Spanish /p/ provided some evidence for L2 acquisition in which accurate production precedes accurate perception, while the data for Spanish /b/ did not support either the notion that production precedes perception or that perception precedes production. It was suggested that the two processes may act independently during certain stages of acquisition. It may also be the case that, for some phones, per-

ception does precede production, while for others the reverse holds. Hence, this present study has served to illuminate a variety of potential interactions and to propose that the interaction between production and perception is more complex than is sometimes assumed.

REFERENCES

Barrutia, R., & Schwegler, A. (1994). *Fonética y fonología españolas*. New York: Wiley.

Flege, J. (1992). Speech learning in a second language. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, and application* (pp. 565-604). Timonium, MD: York Press.

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops. *Word, 20*, 384-422.

Zampini, M. L. (1998). VOT and voiceless closure interval in the production of second language Spanish stop consonants. Manuscript submitted for publication.

English Speakers' Acquisition of Voiceless Stops and Trills in L2 Spanish

JEFFREY T. REEDER, Sonoma State University

The purpose of this paper is threefold: (1) to define acoustically the learner progress in the acquisition of a set of phonological features, specifically interlingual differences in the voice onset time of voiceless stop consonants and the number of taps produced when attempting to produce the Spanish phoneme /r/; (2) to determine if and at what stage of acquisition the given L2 targets are realized; and (3) to use data from adult English-speaking learners of Spanish to test Flege's framework of second language speech acquisition known as the Speech Learning Model. The basis for addressing the above questions in this paper is a cross-sectional study of 40 native English-speaking learners of Spanish in a U.S. university. These learners, representing four different levels, provided data that were analyzed acoustically using computer-based speech analysis software. In addition to tracing the acquisition of a set of sounds through the four levels, this study provides evidence that the Spanish trill is acquired differently than the voiceless stops. Furthermore, Flege's SLM is shown to be a relevant model of second language speech acquisition with respect to the sounds and language combination examined in this study.

INTRODUCTION

This paper is divided into four sections. The first section provides the background for this study by summarizing current views of some of the factors constraining second language (L2) speech acquisition and by describing Flege's (1995) Speech Learning Model (SLM). Also in the first section is a description of the Spanish sounds that are studied in this work, accompanied by an overview of previous research done on the L2 acquisition of those Spanish sounds by English speakers. The second section of this paper describes the method used in this cross-sectional study, including a description of the participants and their experience with the study of Spanish, an explanation of the data collection procedure, and an outline of the scoring procedure. The next section presents the results of the perception test and the data from the participants' production attempts of the trill and the voiceless stops, and the final section summarizes the results and their implications for what is known about the acquisition of L2 Spanish speech.

BACKGROUND

In this section, current views of some of the factors governing second language (L2) speech acquisition are summarized. After this background information is given, a description of relevant portions of Flege's (1995) SLM is provided, given that it is a model of L2 speech acquisition that claims to be

applicable to learners of all ages. Next is a description of the Spanish sounds /r/, /p/, /t/, and /k/ that are studied here, as well as an encapsulation of the research that has already been done on the L2 acquisition of those sounds by native English speakers.

L2 Speech Perception Constraints

Early in their linguistic experience, humans learn to organize the enormous set of perceived sounds by classifying them as speech or non-speech and, for the former, as sounds with phonemic significance in the L1 phonological inventory, in much the same way as they organize the infinite variety of color into categories such as red, pink, or orange. With respect to L2 acquisition, the research suggests that learners in early stages routinely categorize L2 speech sounds that they perceive in terms of their L1 phonological inventory. To test the hypothesis that L1 phonemic categorization rules influence perception of other speech sounds, Scholes (1967) designed an experiment in which listeners from a variety of language backgrounds were presented aurally with a set of synthetic vowel stimuli. The resulting data showed that subjects tended to organize the stimuli in accordance with the vowel systems of their native languages, suggesting an L1 influence in vowel perception. In experiments in which subjects were asked to categorize initial occlusives, for which voice onset time (VOT) is an acoustic cue, Lisker & Abramson (1964) reported that listeners regularly separated a continuum of stop consonant stimuli varying in VOT value according to the categorizations in their L1.

Also seeking evidence of how categories are established, Williams (1979) engaged bilingual adults in perceptory discrimination tasks and found that the subjects established compromise VOT values falling between the categories present in the L1 and L2, suggesting that learners may alter their perceptive categories in response to stimuli. This finding was later confirmed by Flege (1987) and forms the basis for his "equivalence classification" hypothesis, which proposes that learners may group similar L1 and L2 phones into one category based on such compromise values. According to this hypothesis, a learner projects L1 phonetic categories onto the L2 whenever the sounds are judged by the learner to be equivalent; new phonetic categories are formed only when the learner perceives the sounds as different. One effect that Flege (1995) proposed for such equivalence classification is that cases of continued perceptual linkage of L1 and L2 sounds limit the accuracy with which L2 sounds may be produced (Flege, 1995).

L2 Speech Production Constraints

Just as perceptual categories are established for L1 processing, Borden, Harris, Fitch, and Yoshioka (1981) claim that speakers have mentally pre-established representations of muscular gestures that are necessary to produce the articulatory target. Accordingly, L2 production would presumably be limited, either by the degree of similarity between the L1 and L2 targets or by the degree

to which the learner is able to successfully establish new gestural representations. Borden (1980) suggests that self-perception plays an important role in establishing a link between the perception and production of novel phonetic targets in that the learner progressively modifies gestural representations until auditory feedback indicates to the learner that the L2 target has been met satisfactorily. Direct realist accounts of speech learning, such as that proposed by Best (1995), point out that learners have proprioceptive access to the gestures used to create speech sounds and are able to learn efficiently the important elements of the gestures used to create L1 speech sounds. Best proposed that this gestural proprioceptivity leads to the formation of relational "lower-order invariants," which may gradually give way to language-specific, "higher-order invariants," causing a reduced amount of lower-order phonetic detail to be detected and thus potentially interfering with the mechanisms used in the learning of new sounds.

A question that has generated controversy in the literature is whether L2 perception precedes production or whether accurate production can come before (or without) perception. Some studies, such as Lane (1963), Neufeld (1979), and Flege (1987), have suggested that accurate perception must come before production, but other research, such as Gass's (1984) study on English learners' VOT production, suggests that accurate production may precede perception.

The Speech Learning Model

In its current form, the Speech Learning Model (Flege, 1995) presents four postulates and seven hypotheses concerned with the ultimate attainment of L2 pronunciation. The SLM claims that learners of an L2 must create accurate perceptual "targets" to guide them in the production of L2 sounds; failure to do so will result in inaccurately produced targets. The first postulate of the SLM proposes that the same devices that are used by learners to learn their native language (L1) can be accessed at any age and applied to L2 learning. Since the present study treats the adult [1] acquisition of L2 Spanish, this postulate is of considerable importance since it provides the assumption that speech learning processes remain accessible to all L2 learners, regardless of age. Of the seven hypotheses of the model given by Flege (1995, p. 239), the second, third, and seventh are particularly relevant to the present study and are listed here:

Hypothesis 2: A new phonetic category can be established for an L2 sound that differs phonetically from the closest L1 sound if bilinguals discern at least some of the phonetic differences between the L1 and L2 sounds.

Hypothesis 3: The greater the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound, the more likely it is that phonetic differences between the sounds will be discerned.

Hypothesis 7: The production of a sound eventually corresponds to the properties represented in its phonetic category representation.

Hypotheses 2 and 3 predict that learners will be able to create a new phonetic category once they perceive that a sound differs from a corresponding sound in the L1; the likelihood of this occurring increases as the differences between the L1 and L2 sounds magnify. According to Hypothesis 7, once learners have established such a phonetic category representation for a novel sound, their production of that sound will eventually correspond to that of native speakers of the L2, provided their phonetic categories were accurately represented.

Description of Sounds

The Phonemes /p/, /t/, /k/

The phonemes /p/, /t/, and /k/ are present in both English and Spanish. Despite their apparent similarity, however, in English the allophones [p^h] (aspirated voiceless bilabial stop), [t^h] (aspirated voiceless alveolar stop), and [k^h] (aspirated voiceless velar stop) may occur in word-initial position or at the beginning of a stressed syllable, whereas in Spanish the three phonemes in question each have only one possible syllable-initial allophonic realization (the voiceless bilabial stop [p] for /p/, the voiceless dental stop [t] for /t/, and the voiceless velar stop [k] for /k/).

As originally proposed by Lisker and Abramson (1964), and as since expanded further by many other researchers, the standard acoustic correlate used to measure stop consonant production is VOT [2]. Within the VOT continuum, Lisker and Abramson found that most languages tend to cluster VOT around three or fewer categories of values: long voicing lead, zero onset or short lag, and long voicing lag. Keating (1984) proposed describing these as phonetic categories, such that Lisker and Abramson's categories would be phonetically realized as the following: *voiced* (also referred to as *prevoiced*), *voiceless unaspirated*, and *voiceless aspirated*, respectively. Figure 1 illustrates the phonological differences for stop consonants between English and Spanish along the VOT continuum. The figure shows that English uses all three phonetic categories, with the voiced and voiceless unaspirated categories used for voiced phonemes and the voiceless aspirated category used for voiceless phonemes. Spanish, however, uses only two of the phonetic categories. In Spanish, the voiced category represents voiced phonemes and the voiceless unaspirated category is used for voiceless phonemes. The voiceless aspirated phonetic category is not used in Spanish.

The Phoneme /r/

The Spanish trill /r/ has no counterpart in any dialect of American English, nor in most other dialects of English. Although this phoneme's allophonic distribution may include allophones that are voiced or voiceless fricatives, uvular trills, and voiced or voiceless alveolar trills, the most common allophone in most dialects and the one that is most frequently taught to learners of

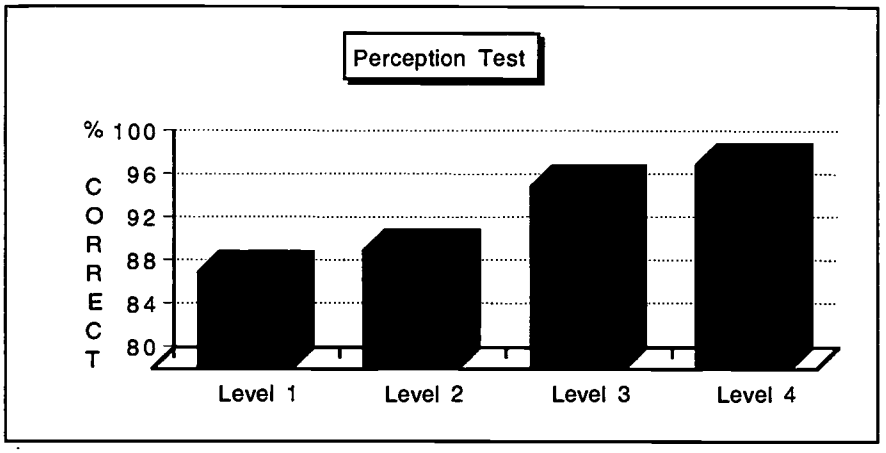


Figure 1. Perception Test Results, by Level

L2 Spanish is the voiced alveolar trill represented by the International Phonetic Alphabet symbol /r/. The most readily identifiable acoustic correlate of the /r/ is a regular interruption in the waveform and spectrograph caused by the brief, periodic cessations of phonation that correspond to each contact between the tongue and the alveolar area.

Existing Research On L2 Spanish Phonemes /p/, /t/, /k/, and /r/

Although interest in the acquisition of L2 Spanish stop consonants by English-speaking populations has developed relatively recently, there have been a number of important advances. González-Bueno (1997) tested two groups of intermediate learners of Spanish in a foreign language setting. The first, an experimental group, received explicit instruction and practice in the production of Spanish stop consonants, while a control group received no special instruction. Over the course of one semester, she found that learners in the experimental group were able to significantly shorten VOT in their production of /p/ and /g/, thus rendering their production more Spanish-like. Although the /t/, /k/, /b/, and /d/ also improved, the degree of improvement was not statistically significant, which González-Bueno attributes to the interactive operation of developmental and transfer processes.

Flege and Eefting (1988) conducted an experiment in which English monolinguals, Spanish monolinguals, and native Spanish bilinguals imitated a consonant-vowel continuum in which the VOT of the initial consonant varied. As expected, the Spanish monolinguals showed a tendency to produce stops

with Spanish-like VOT values (short lag, long lead) and the English monolinguals tended to produce English-like stops (short lag, long lag). With the bilingual group, however, the researchers reported stop production in all three VOT ranges, suggesting that those subjects had processed the stops in terms of the three phonetic categories present in both languages.

Contrasting with the amount of attention that has been given to the acquisition of L2 Spanish stop consonants as evidenced by the aforementioned studies, the trill has not been the subject of recent published studies. Consequently, this study hopes to fill a gap in the research by directing inquiry to the L2 acquisition of the /r/ among adult English speakers.

METHOD

This section of this paper describes the research method employed in this cross-sectional study. A description of basic characteristics of the participants as well as information about their background with L2 Spanish is given. Following the description of the subjects, an explanation of the data collection procedure and an outline of the scoring procedure are provided.

Subjects

The present study examines data gathered in 1997 from 45 volunteers affiliated with the Spanish program at a medium-sized, private university in the United States. Of these participants, 40 were native speakers of English and the remaining 5 were native speakers of Spanish. The native English-speaking subjects were recruited from among those who had no significant childhood background with Spanish, neither through significant formal study prior to age 12, through residence or extensive travel in Spanish-speaking communities, nor through family contact. The native English-speaking subjects are categorized as below and are described in Table 1.

Level 1 (Beginning Learners): The 10 participants in this group were students enrolled for credit in a first-semester university Spanish language course.

Level 2 (Intermediate Learners): The 10 participants in this group were students enrolled for credit in a third-semester university Spanish language course.

Level 3 (Advanced Learners): The 10 participants in this group were enrolled for credit in an upper-division or graduate-level university course in Spanish language, literature, or culture.

Level 4 (Very Advanced Learners): The 10 participants in this group were native speakers of English who were full-time instructional faculty teaching university courses in Spanish language, literature, or culture and who otherwise met the criteria for inclusion in the study.

The native Spanish-speaking participants were recruited from among teachers of Spanish at the same university. These native speakers, from Chile, Colombia,

Table 1
Participant Characteristics (Group Means)

	Level 1	Level 2	Level 3	Level 4
Age (years)	18.5	19.7	21.6	41.8
H.S. Span. (years)	2.2	2.5	2.3	2.2
Univ Span. (semesters)	1.1	2.6	7.9	21.6
Hrs Span used/day	1.2	0.9	1.8	4.1
Days in Span country	4.2	5.0	18.3	1615.5

Mexico, and Spain, were selected to provide speech data for purposes of comparison, and they represent many of the most common varieties of Spanish.

Data Elicitation

The data collection process consisted of three separate types of data-gathering methods. A questionnaire collected relevant biographical data and learner characteristics, a forced-choice minimal pair test prompted by a Spanish audio recording provided a measure of auditory discrimination, and three different types of audio recordings of each subject's Spanish captured L2 production. To ensure consistency and cross-level comparability of data, the data collection procedure was identical for each participant, regardless of level. All instructions were given in English.

After signing a form indicating their informed willingness to participate in the study, each participant completed the first part of the study, a questionnaire that identified the relevant individual learner characteristics and established a profile of each level of subjects, as seen in Table 1. The second part of data collection was a two-alternative, forced-choice discrimination test designed to measure perception of discrete target language items among 12 minimal pairs. The stimulus for this portion of the experiment was a previously recorded audio recording of a native Spanish speaker reading one of the minimal pair items; subjects circled the item they believed they heard on a score sheet. The recording was presented via a Sony TCM-919 cassette player.

The third part of this study recorded the subjects' production of Spanish. This part consisted of three different segments, each involving a different elicitation protocol. In the first of these segments, the re-

searcher instructed subjects to read from a list of ten Spanish words or phrases embedded in a carrier phrase common to all items (e.g. *"digo ~ esta vez"*). These items were selected to provide a wide variety of phonemic targets and to measure sensitivity to (and influence from) orthographic cues in an elicitation protocol where all or most of the participant's attention could be directed to pronunciation. In the second segment, also designed to measure discrete lexical items, participants were sequentially shown 11 picture cards, each with a drawing of a relatively common item. The task was for the subjects to say the names of the objects in Spanish. Finally, in the third speech elicitation protocol, participants provided a 30-second guided narration in Spanish in response to a written cue. This segment provided data from a context in which the participant's mental resources were presumably engaged in many elements of language production besides pronunciation.

Speech data for all participants were recorded as follows: Each subject was recorded individually in the researcher's office (which, while not an anechoic chamber, includes a number of sound muffling features). Having been prompted, subjects spoke into a Sony SV-9 microphone mounted on a Nissin tripod stand. The resulting input fed directly into the audio input of a Macintosh desktop computer, which then processed and recorded the signals digitally using *Signalyze* speech analysis software.

RESULTS

The results section presents the findings of the perception and speaking elicitation tests. These findings, presented in tabular and graphic formats, are also analyzed statistically to evaluate the significance of the findings.

Perception Test

The results of the two-alternative forced-choice perception test for each level are presented in Figure 1. The perception test in this study provides a general assessment of the participants' perception of Spanish minimal pairs, including 5 vocalic and 6 consonantal features. These results show that in this study learners at more advanced levels were better able to discriminate among Spanish minimal pairs than those at beginning levels, although learners at all levels showed a high percentage of correct responses, suggesting that most learners correctly perceive phonemic features in Spanish.

Spanish Production Data: Stop Consonants

The data collected from the speech elicitation tests appear in Tables 2 and 3. Table 2 indicates the shortest, longest, and mean VOT values recorded for each stop consonant (the data are also arranged graphically in Figures 2 through 7). The results of the ANOVA shown in Table 3 suggest that stop consonants are produced with a significantly more Spanish-like VOT when

Table 2
VOT Values for Each Phoneme, by Group

	/p/	/t/	/k/
Level 1			
Shortest VOT (ms)	36	26	41
Longest VOT (ms)	89	102	117
Mean VOT (ms)	54	53	73
Level 2			
Shortest VOT (ms)	27	30	53
Longest VOT (ms)	86	98	116
Mean VOT (ms)	51	55	74
Level 3			
Shortest VOT (ms)	13	15	34
Longest VOT (ms)	54	79	112
Mean VOT (ms)	36	36	60
Level 4			
Shortest VOT (ms)	17	16	31
Longest VOT (ms)	46	58	73
Mean VOT (ms)	29	29	49
Native Speaker [3]			
Shortest VOT (ms)	9	10	17
Longest VOT (ms)	27	33	41
Mean VOT (ms)	17	17	28

Table 3
ANOVA on the Difference of Mean VOT Values Between Levels

Comparison	Difference in Mean VOT (ms)
Level 1 > Level 2	0
Level 1 > Level 3	-48*
Level 1 > Level 4	-73**
Level 2 > Level 3	-48*
Level 2 > Level 4	-73**
Level 3 > Level 4	-25

*p <0.01
**p <.001

comparing Level 1 with Levels 3 and 4, as well as when comparing Level 2 with Levels 3 and 4. With some exception, reduction of VOT times was a general tendency throughout all of the levels, but only the increases evident between the first two andlast two levels reached confidence levels of 99% or greater.

Table 3 shows the results of an ANOVA on the difference of mean VOT values between levels for all three voiceless stops. The difference between the Level 1 values and the Level 2 values were, coincidentally, zero (the reduction in /p/ values from Level 1 to Level 2 was offset by the slightly higher values with /t/ and /k/). Statistically significant VOT reductions occur from Level 1 to Levels 3 and 4, and from Level 2 to Levels 3 and 4. Although a VOT reduction occurs from Level 3 to Level 4, it does not reach significance.

Figures 2 through 7 on the following pages illustrate the range of mean VOT values recorded from participants' speech samples. Figures 2 through 4 show the entire range of VOT mean values for /p/, /t/, and /k/, respectively, whereas the overall mean VOT for each level appears in Figures 5 through 7.

As shown in Figures 2 through 7 above, the three voiceless stop sounds follow the same pattern of acquisition. Learners in the earlier stages of acquisition, such as those in Levels 1 and 2, tend to produce the target sound inconsistently, as

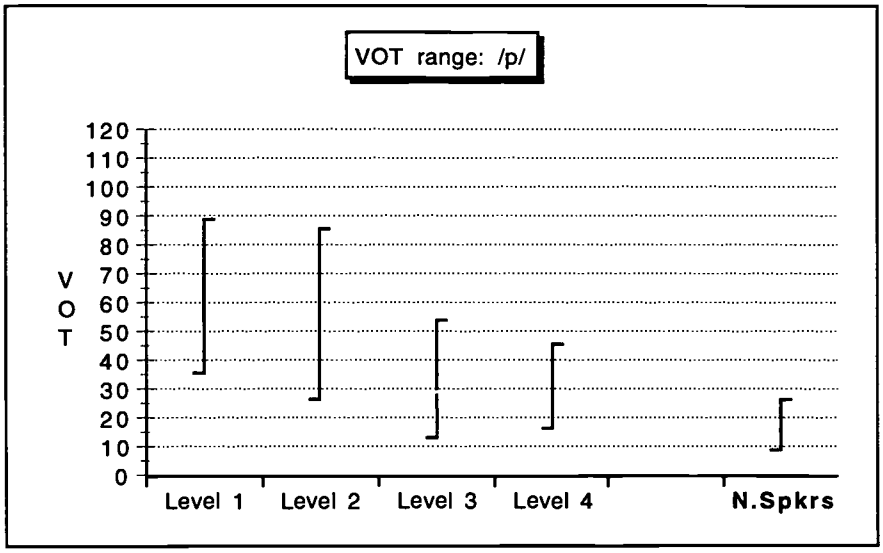


Figure 2. Range of Mean VOT Values for /p/, in Milliseconds.

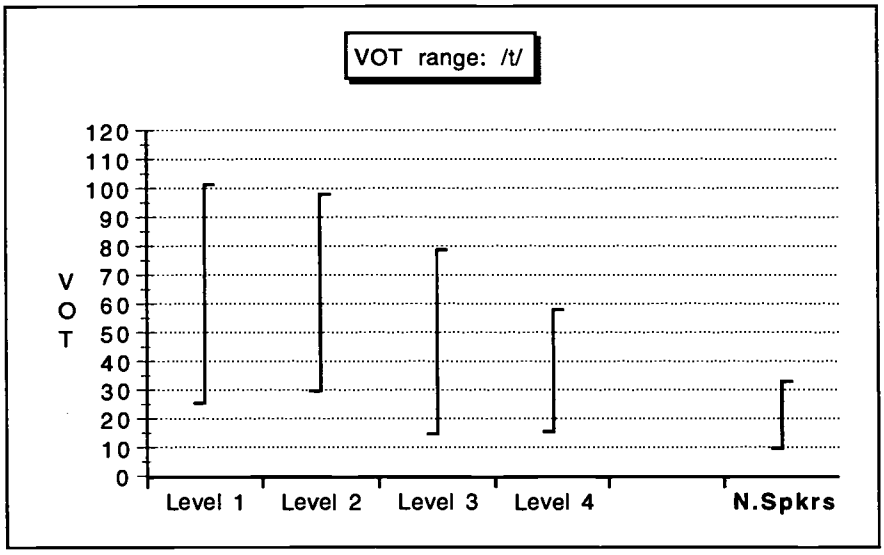


Figure 3. Range of Mean VOT Values for /t/, in Milliseconds

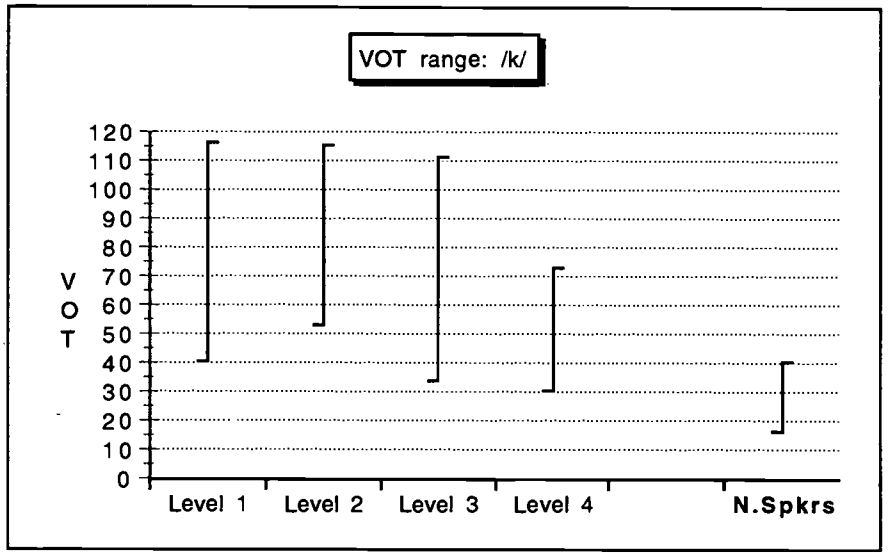


Figure 4. Range of Mean VOT Values for /k/, in Milliseconds.

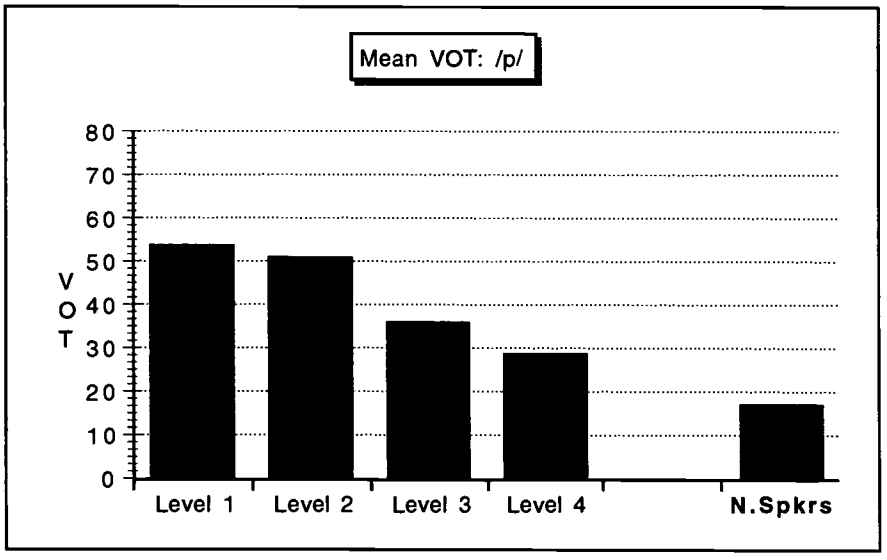


Figure 5. Group Mean VOT Values for /p/, in Milliseconds.

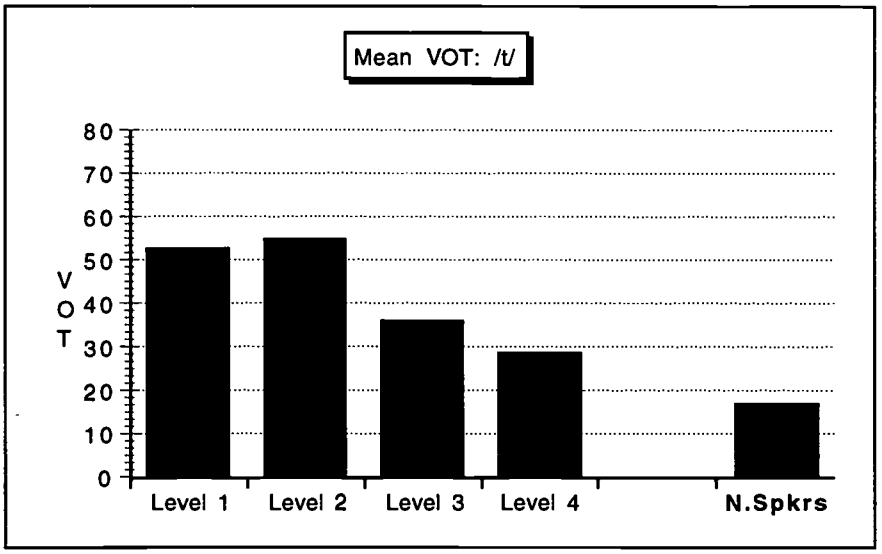


Figure 6. Group Mean VOT Values for /t/, in Milliseconds.

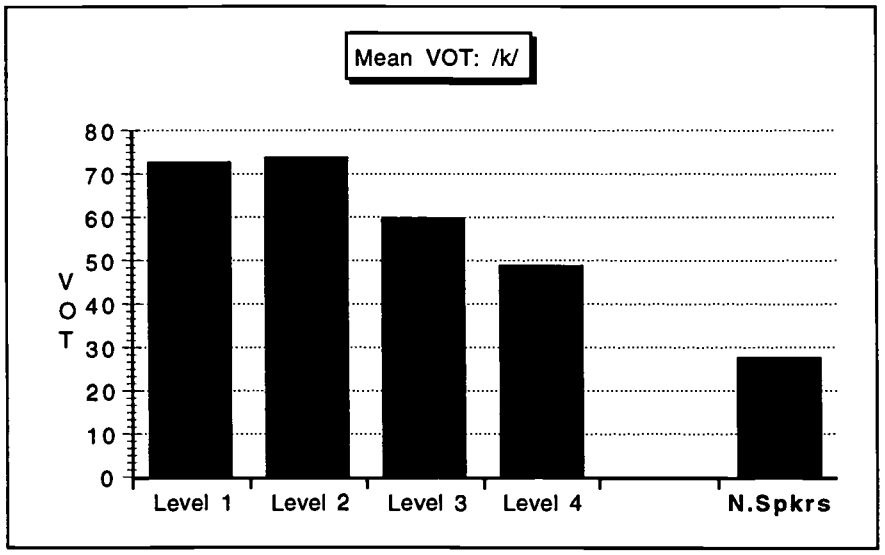


Figure 7. Group Mean VOT Values for /k/, in Milliseconds.

evidenced by a wide range of VOT values in their attempts at the L2 target.

For example, Figure 3 shows that there is more than 70 ms difference between the shortest and longest Level 1 mean VOT. Additionally,the less advanced learners show a tendency to equate the L2 sound with its L1 counterpart, resulting in the high, English-like mean VOT values that are shown in Figures 5 through 7. Although the learners in Levels 3 and 4 show considerably shorter, more Spanish-like mean VOT values than those in Levels 1 and 2, their VOT ranges are still not significantly Spanish-like and instead are intermediate between English and Spanish norms, suggesting the possibility of continued, although diminished, association between the L1 and L2 sounds.

Spanish Production Tests: Trill

With respect to the acquisition of the Spanish trill, the data show clearly definable progress. Evidence of this progress shown in Table 4 is that the mean number of trills produced by learners in each level steadily increases from 0.6 taps per /r/ attempt among the beginners to 2.6 taps per /r/ attempt among the very advanced learners. Similarly, the percentage of /r/ attempts with at least two taps increases from 7% to 83% from beginners to very advanced learners.

Given that this sound does not exist in the L1 of the participants in this study, it is not surprising that the beginning learners in Level 1 demonstrated an almost complete inability to produce the Spanish trill when necessary, evidenced by the low 7% of all Level 1 /r/ attempts

Table 4
Number of Closures (Trills) for /r/ Attempts

	Min	Max	Mean	% with 2+ trills
Level 1	0	4	0.6	7
Level 2	0	4	0.9	13
Level 3	0	5	1.8	37
Level 4	1	4	2.6	83
N. Spkrs.	2	5	3.2	100 [4]

that were produced within the native speaker range of 2-5 trills. Contrarily, the near complete mastery of this novel sound by the Level 4 participants suggests that learners are ultimately able to construct and employ an entirely new sound category.

DISCUSSION

The subjects in this study represent Spanish L2 learners of all levels, ranging from beginners whose exposure to Spanish came largely from input available in a formal foreign language study context to advanced learners with years of target language experience in both formal and naturalistic settings. Despite the existence of many differences between the populations represented in each of the four groups, all subjects shared several important characteristics. Before reaching age 12, none of the 40 participants in the four groups of native English-speakers had studied Spanish, been a part of a Spanish-speaking family, or lived in a Spanish-speaking community. Thus, this study can serve as a measure of the progress in L2 Spanish pronunciation of voiceless stops and trills by a group of English-speaking adult learners.

The results suggest that the phonological interlanguage in the groups tested, as measured by the test instrument used in this study, may be characterized as having an acoustically definable and statistically significant acquisitional sequence that initially shows many L1 characteristics and progresses over time to a more L2-like production, supporting Flege's postulate that the mechanisms and processes used in learning the L1 sound system remain intact throughout one's life and are available to be applied to L2 learning. Although the beginners were generally unable to produce the speech targets with phonetic accuracy, the advanced learners in this study were able to produce Spanish trills and voiceless stops that frequently matched or closely resembled those produced by native Spanish speakers.

The SLM claims that inaccurate productions in L2 sounds may result from the learners' failure to create appropriate perceptual "targets" to guide their production (Hypothesis 7), which may in turn be caused by the learners not discerning at least some of the phonetic differences between the L1 and L2 sounds (Hypothesis 2). This learner awareness and identification of the phonetic differences between the L1 and L2 are made more likely as the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound increases (Hypothesis 3). The SLM thus predicts that the native English speakers learning Spanish in this study will be more likely to create a new phonetic category for the /r/, which differs from any sound in the L1, than for either of the voiceless occlusives /p/, /t/, or /k/, which share many similarities between the two languages. Evidence of the formation of (or failure to form) new phonetic categories, according to Hypothesis 7, will come from learner production of that sound, which may eventually correspond to that of native speakers. The data presented in Table 4 on the acquisition of the /r/ indicate a near-complete mastery of the /r/ among the group of very advanced learners compared to a

near-zero ability among the beginners. Since the /r/ represents a novel phonetic category for the participants in this study, these findings support the SLM. The results of the data collected on the voiceless stops /p/, /t/, and /k/ also support Flege's model. As seen in Figures 4 through 6, the VOT values of the target voiceless stop phonemes were rarely produced within the native speaker ranges by the beginners; however, even the VOT ranges of the most advanced learners do not convincingly overlap the native speaker ranges. One explanation for this finding is that the phonetic dissimilarity between the Spanish and English stops is not as great as with the /r/ and thus learners are not as likely to perceive the need to actively and accurately create new phonetic categories.

In sum, this paper suggests that English-speaking adult learners' acquisition of the trill and the voiceless stops in L2 Spanish evolves in an acoustically definable manner. It also suggests that significant improvement is evident between many of the levels, even though none of the L2 Spanish sounds examined in this study were acquired completely enough to show significantly similar acoustic parameters to native speakers' production of the same sounds. Finally the evidence suggests that over the long term, the /r/, a sound for which no English counterpart exists, is acquired much more completely and consistently than the /p/, /t/, or /k/, all sounds that have similar English counterparts. This finding supports the SLM's (Flege, 1995) contention that L2 sounds for which no L1 equivalent exists are ultimately more likely to be acquired than sounds for which there are L1 equivalents.

ACKNOWLEDGMENTS

I thank Dale A. Koike and the participants at the University of Texas Spanish Second Language Symposium for their comments and suggestions on an earlier version of this paper, as well as the students, faculty, and administration of Baylor University who contributed to this study. Special thanks to Susye and Andrew for their continuous support.

NOTES

1 The term *adult* in this paper refers to any individual beyond puberty. Although I have chosen the term "adult language acquisition" for language acquisition after puberty and have set a criterion in participant selection which excludes those significantly exposed to Spanish prior to age 12, my doing so is arbitrary. Researchers such as Lenneberg (1967) and Scovel (1988) have proposed that language cannot be learned perfectly once a biologically predetermined critical period has been passed, due at least in part to neurological maturation. However, recent evidence (e.g., Flege, Munro, and MacKay, 1995) suggests that the acquisition of L2 speech, at least, follows a strong linear relationship with age and shows no evidence of a definable critical period.

2 Voice onset time (VOT) is the time between the release of the articulators in a stop consonant and the onset of vocal fold vibra-

tion in the following vowel. VOT is usually measured in milliseconds (ms), with the stop consonant release representing zero. When vocal fold vibrations begin before the articulators' release, as may occur with voiced stops, the convention is to assign a negative VOT value.

3 For comparison, Nathan (1987) presented the following mean VOT values (ranges in parenthesis) for seven Spanish speakers from Costa Rica, Colombia, and Venezuela: /p/ 18.8 (8.2/31), /t/ 22.6 (13.5/30.3), /k/ 40.7 (22.5/55.5). From the same source, mean values for English were given as /p/ 82.5, /t/ 105, and /k/ 117.

4 One of the native speakers produced the voiced alveolar trill allophone during the reading and picture identification portions of the data collection, but produced a voiceless alveolar slit fricative during the free-speaking portion. This allophone is fairly common in certain regions of Spanish America, and its appearance in the speaking task, but not the reading or identification tasks, is consistent with Dalbor (1980), who reports that many speakers use the trill in formal speech and the fricative in informal speech.

REFERENCES

Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-206). Timonium, MD: York Press.

Borden, G. (1980). Use of feedback in established and developing speech. In N. Lass (Ed.), *Speech and Language, Volume 3* (pp. 233-249). New York: Academic Press.

Borden, G., Harris, K., Fitch, H., & Yoshioka, H. (1981). Producing relatively unfamiliar speech gestures: A synthesis of perceptual targets and production rules. *Status Report on Speech Research, 66*, 85-117. New Haven, CT: Haskins Laboratories.

Dalbor, J. (1980). *Spanish pronunciation*. Fort Worth: Holt, Rinehart, and Winston.

Flege, J. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*, 47-65.

Flege, J. (1995). In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-272). Timonium, MD: York Press.

Flege, J., & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *Journal of the Acoustical Society of America, 83*, 729-740.

Flege, J., Munro, M., & MacKay, I. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication, 16*, 1-26.

Gass, S. (1984). Development of speech perception and speech production abilities in adult second language learners. *Applied Psycholinguistics, 5*, 51-74.

González-Bueno, M. (1997). The effects of formal instruction on the acquisition of Spanish stop consonants. In W. R. Glass & A. T. Pérez-Leroux (Eds.), *Contemporary perspectives on the acquisition of Spanish, Volume 2: Production, processing, and comprehension* (pp. 57-75). Somerville, MA: Cascadilla Press.

Keating, P. (1984). Phonetic and phonological representation of stop consonant voicing. *Language, 60,* 286-319.

Lane, H. (1963). Foreign accent and speech distortion. *Journal of the Acoustical Society of America, 35,* 451-453.

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20,* 384-422.

Nathan, G. S. 1987. On second language acquisition of voiced stops. *Journal of Phonetics, 15,* 313-322.

Neufeld, G. (1979). Towards a theory of language learning ability. *Language Learning, 29,* 227-241.

Scholes, R. (1967). Phoneme categorization of synthetic vocalic stimuli by speakers of Japanese, Spanish, Persian, and American English. *Language and Speech, 10,* 46-68.

Williams, L. (1979). The modification of speech perception and production in second-language learning. *Perception and Psychophysics, 26,* 95-104.

Prochievement Testing of Speaking: Matching Instructor Expectations, Learner Proficiency Level, and Task Type

BARBARA GONZALEZ PINO, The University of Texas at San Antonio

Earlier literature on classroom testing of speaking provides several models of both task types and rubrics for rating and suggestions regarding procedures for testing speaking with large numbers of learners. There is no clear, widely disseminated consensus in the profession, however, on the appropriate paradigm to guide the testing and rating of learner performance in a new language, neither from second language acquisition research nor from the best practices of successful teachers. While there is similarity of descriptors from one rubric to another in professional publications, these statements are at best somewhat subjective. Thus, the rating of learners' performance rests heavily on individual instructors' interpretations of those descriptors. The author conducted an initial investigation of instructor assumptions regarding student performance on speaking tests in her own program and identified several discrepant areas of instructor testing and rating practice. Further, faculty as a group will need to delineate more their specific expectations by level for a number of the rated features. The concerns identified coincided with those discussed recently in the literature, which suggests that other programs may also benefit from similar self-analysis.

INTRODUCTION

The language educator who is familiar with both the American Council on Teaching Foreign Languages (ACTFL) Proficiency Guidelines and the proficiency levels typically achieved by first- and second-year university learners of foreign languages may well have questions about some of the ways in which speaking is tested and rated in classes in some of those university programs. If learners are in the Novice Mid to Intermediate Mid range, are the testing tasks they are given always level-appropriate and should they be? Are the descriptors of the rating scales used by the instructors always appropriate and clear? Do all the instructors interpret general or ambiguous descriptors in the same way? Is the system of testing and rating that is in place implemented consistently by the instructors in a program and should it be? Do the tests match well with what is covered in classes? Do the instructors' expectations while rating reflect a solid understanding of what learners can do at their proficiency level? Do their expectations reflect an understanding of the restructuring of knowledge and performance that may occur in learners as they move to the Intermediate proficiency level? How do instructor-raters handle beyond-level tasks? We may not always know the answers to these questions as they pertain to our own programs, much less on a larger scale.

Concern about these questions and about the issue of fairness to students led the author to conduct a study in the lower-level program of her own institution to determine the procedures and criteria her instructors were actually using in testing and rating speaking and the extent to which these criteria coincided with program goals and current best practice. Clearly, as in most programs, instructors were testing tasks that were above the students' proficiency level, and their rating scales did not really distinguish level-appropriate tasks from above-level tasks. In addition, the degree of possible variations in implementation and interpretation of process, procedure, and criteria in testing and rating was unknown. Therefore, the author proposed to investigate these areas and use the results to initiate discussion among instructors in the program and in a broader professional setting.

THE LITERATURE

Throughout the recent literature on the testing of speaking, many concerns are raised about testing and rating procedures. Much of this literature, however, focuses on proficiency testing rather than on prochievement testing (Clark, 1989), a term that refers to the kind of proficiency-oriented achievement testing we do in foreign language classes on a regular basis, perhaps several times a semester. Nevertheless, in both the proficiency literature and the prochievement literature, aspects of testing and rating appropriately or inappropriately are discussed at length, and some of the features discussed in proficiency studies may have relevance for prochievement testing as well. In general, the profession defines achievement tests as those limited to a particular body of material just covered in class(es) and proficiency instruments as those testing the total range of skills and contexts a learner may be able to handle—regardless of where and when they may have been learned—and testing them through actual interaction in realistic situations. Prochievement tests are a combination of the preceding two types, testing students' ability to perform in only the contexts and situations that have been practiced in class.

Formats

A common concern is whether particular tasks or formats are best suited to certain proficiency levels or particular teaching and testing circumstances. According to Fulcher (1996), there is little evidence to suggest that any particular task format is more suited to one proficiency level than another. Indeed, most of the common formats can be used with learners at different points in their studies. With appropriate expectations, the picture, the topic, the interview, the multi-skilled or integrated task, and even the more demanding roleplay can all be adapted according to the level of the students. The crucial element in using the formats well is their content, which should comprise functions, topics or life situations, and grammatical features appropriate for the particular students and the material covered in their classes (Gonzalez Pino, 1989).

Instructor-Rater Expectations

The aspects of the topic of testing speaking that are most thoroughly covered in the literature are those of rating and the underlying instructor expectations that are such an important part of rating. Effective ways to rate have been studied for decades and in an extensive variety of formats and weighting schemes (Hart Gonzalez, 1994). Thompson (1996) found that rating may vary according to whether the test is taped or not and noted that a rater who is listening to a tape is not as likely to be distracted by the human qualities present in a live interview and is more likely to pay greater attention to form. Richards and Chambers (1996) studied a number of instructor-related variables in rating and reported that many of them have some effect on the rating process. They stated that training on how to rate improves consistency and that linguistic background counts, because native speakers rate more stringently. According to their study, the type of school in which teachers work matters; teachers in more elite or selective schools rate more stringently. One type of teacher experience significant in their findings is experience with learners at the level being rated. Length of overall teaching experience does not matter, however.

Richards and Chambers (1996) examined three types of rating scales in their studies: a norm-referenced categorical scale (one with weighted criteria and numeric scales for each but with no descriptors for the criteria), a criterion-referenced categorical scale (one with a set of criteria, each with a hierarchy of descriptors and numeric values), and a global criterion-referenced scale (one with descriptors and numeric values for each of several general levels of performance). They found that the two more global scales were more reliable, but they explained their finding by suggesting that the descriptors for the criterion-referenced categorical scale were vague and would require much greater specificity in order to function appropriately. Douglas (1994) found that most raters who used scoring rubrics were apparently affected by aspects of performance that were not mentioned in the rubrics. He noted that grammar and rhetorical complexity were particular problem areas for which teacher-raters might employ their own standards or substandards. Richards and Chambers (1996) discovered that pronunciation and grammar caused the greatest rating problems in their study, possibly because these two areas are concrete and yet have no specific detailed standards set out in common for the various levels for all raters to use.

In their 1995 study, Chambers and Richards also found that if the criteria to be used in rating were not described in some detail, teachers varied in their interpretations of the descriptors. Further, they found that teachers may expect strong performance on grammatical elements even if those elements are not appropriate to the task, are not appropriate to the students' level, and would not have been used by native speakers on the same task. Their study specifically compared learners' and native speakers' performances on the same set of tasks in order to compare the grammatical structures that were

used. They determined that teachers may expect forms that not even native speakers employ. They also found that learners who spoke more often received higher ratings, regardless of quality issues. Finally, they determined that these expectations frequently persisted despite differing expectations written into course syllabi, where certain features were cited for recognition and others for both recognition and production.

Thompson (1995) found that a group of proficiency raters tended to develop idiosyncratic testing and rating procedures as compared to other groups. Mullen (1978) recommended that more than one rater rate each test in order to eliminate the effect of rater inconsistency, a practice that has often been followed in proficiency testing since that time, although that procedure would not be practical in classroom testing multiple times per semester. Ross (1987) raised the issue of the appropriate mental construct to undergird norm-referenced scales, suggesting the proficient nonnative would be a better standard than the educated native speaker and highlighting the fact that we may vary in the standard to which we refer. Meredith (1990) suggested further that when rating, teachers must consider whether or not learners have had prior experience in the language; thus, he indicates yet another way in which our mental model and our expectations may vary. Whom do we expect our learners to be like? And do we expect a higher performance level of our false beginners than that indicated for all learners in a particular course?

Levels of Proficiency

Several other concerns in the literature center on the proficiency levels themselves. Stansfield and Kenyon's (1992) study reported that Intermediate and Advanced tasks are more difficult to rate than Novice and Superior and that sublevels of performance in the midranges are more problematic to distinguish from each other. Byrnes (1987) pointed out that Intermediates may make more errors than Novice Lows and Novice Mids, a seeming inconsistency. This indication, however, is related to Young's (1992) indication that there may be an uneven progression in language acquisition, even a "U-shaped" phenomenon in which Intermediate learners may seem to regress because, as they acquire new structures and vocabulary and reformulate their interlanguage, the restructuring destabilizes their performance for a time. The fact that they are creating in the language and relying less on memorized material has a similar effect. Thus, in addition to considering whether our expectations of learners are generally appropriate to their level, we must also consider the extent to which those expectations take into account these additional complexities in second language acquisition and in the rating process.

Textbooks

Finally, we can consider our textbooks as a type of professional literature to be examined and having clear implications regarding proficiency levels. First-year textbooks, so called whether they are used for the first year or a year and a half, invariably cover much of the structure

of the language in question and include functions and content that would be consistent with the Advanced and Superior proficiency levels, despite the fact that no learners (other than native speakers, perhaps) are expected to achieve those levels of proficiency during the first-year (or year-and-a-half) course. Second-year materials also typically include Intermediate through Superior material. The case can certainly be made that we are introducing materials at those levels as a pedagogical strategy to enable learners to begin to develop those particular skills. In each program, however, we still must decide on the appropriate way to evaluate performance on Advanced- and Superior-level material relative to performance on functions, structures, and topics for lower levels of proficiency, both when we design and when we rate our prochievement tests of speaking.

Summary of Literature

Clearly, then, in summary, the literature addresses some of our initial concerns by indicating that many variables affect teachers' rating of learners' speaking. Chief among these variables is the teacher's own set of expectations of students. Since these expectations could apply even in the face of specific statements on syllabi constraining such expectations and despite recommendations contrary to instructor expectations presented during training on how to rate, the concerns seem valid. Since only Richards and Chambers' (1996) and Gonzalez Pino's (1989) studies specifically concern class-related testing, while the others focus on proficiency testing of speaking, however, the investigator undertook to explore further the extent of variation in expectations among instructors who rate their own students' prochievement tests of speaking.

DESCRIPTION OF THE SURVEY

Several researchers have mentioned the need to ask raters to engage in self-assessment and in "think-aloud" protocols. They point out that a comparison of assigned ratings does not always permit an analysis of underlying differences in expectations since two raters can assign the same score for different reasons. Thus, this investigation begins with a self-assessment of testing and rating procedures that, it is hoped, will pinpoint further topics for investigation. The author will analyze consistency among raters and the relationships of responses to the constructs of the ACTFL Proficiency Guidelines and the tenets of second language acquisition

The Sample

Twenty instructors of lower-level language courses at the university level participated in the survey. These individuals teach in a communicatively oriented program in which the policy calls for daily emphasis on speaking skills. They administer and rate speaking tests for their own learners three times each semester. A set of 20 to 30 sample oral test items is provided to the instructors and the learners 2 weeks prior to each test. Each test comprises pictures, topics, interviews, and role-plays related to the chapters in question. The items, which were developed by a subcommittee of instructors for use by the entire group of 20

to 25 for coordinated examinations, cover the dozens of functions and topics included in the text. The text includes functions and topics appropriate to the Advanced and Superior levels reflective of the texts discussed above. The instructors then use the sample items as a repertoire or bank of items as they individually structure the way in which they will administer the test. The instructors determine whether they will test one-on-one or in learner pairs, in class or in their offices, whether they will use two or more formats on a given test (two are the departmental minimum), and whether they will tape record the test or not.

The instructor-raters vary in age, and their teaching experience ranges from 2 years to more than 30. They all have Master's degrees or higher, and all have a graduate specialization in the language in question. They are all professional language educators, even though a few are pursuing further graduate studies on a part-time basis. Some are foreign nationals, but all have experience in the U.S. educational setting. Almost all have had Oral Proficiency Interview familiarization training, and several have completed ACTFL OPI training. Many of them serve as Simulated Oral Proficiency Interview (SOPI) raters on a regular basis as well, and many have rated oral placement tests at the institution for a number of years. They have all attended training each year on how to administer and rate the tests. The amount of training per instructor thus varies with the number of years of experience in the progam. They all use the same set of rating scales, and all are of the norm-referenced categorical types, as adjusted for year 1 and year 2. Most of the instructor-raters have attended interrater reliability training sessions in which four or five actual student tapes have been rated by the group and in which expectations and interpretations of the criteria were discussed at length. Nevertheless, interaction in those sessions and other meetings of the group has highlighted on-going variation among the members in their expectations in the various categories being rated. The present survey should provide the opportunity to highlight specific areas of variation for further discussion and training.

The Instrument and Procedure

The author created a two-page checklist of 61 items relating to functions tested, formats used, and expectations held in rating (Appendix). There are 20 items on functions tested, thus sampling only a part of the curriculum in this area; 17 on formats used; and 24 on rater expectations of student performance. The instructors were asked to check all the statements that applied regarding their own procedures in testing and rating and their own expectations of learners in first- and second-year classes, which all of the instructors teach. In addition, they were provided space at the end of the questionnaire to write anything else they wished regarding their expectations of first- and second-year learners on speaking tests for their classes. The respondents were anonymous, since no place was provided on the survey for them to identify themselves. Anonymity was important, since one could assume that any in-

structors who felt their procedures or expectations did not match coordinated departmental expectations might not have wished to reveal them otherwise.

The instructors were accustomed to surveys and other efforts to research program functioning; therefore, they were simply asked via memo to fill out an attached survey in order to inform the coordinator's efforts to plan tester training for the semester in question. Forms were returned anonymously to the researcher's box over a period of days. Ninety percent of the instructor pool responded.

RESULTS AND DISCUSSION

As stated previously, respondents were asked to react to items covering functions tested, formats used, and expectations held. The results are shown in Table 1 and discussed in the following sections.

Functions

There was somewhat less variation among the instructor tester-raters in the area of functions tested than in the areas of formats and expectations. This finding might not seem surprising at first glance, given the coordinated nature of the program and the standard test samples distributed to faculty and learners; however, the same uniformity could have held true for formats but did not to the same extent.

The description function, which is a staple for Intermediate-level students and a logical starting point for Novices as well, was almost universally tested. All the instructors indicated that they included description of pictures and people, and 90% included description of places. Only 70% included description of objects, however.

Ninety percent of the instructor-raters had the learners ask questions on specific topics and get information about costs, times, and so on in real-life situations. Only 60% had students ask information questions about pictures. Interestingly, only 60% indicated that they had learners give information to others, although one would assume that participating in these question-asking formats would also include answering questions. Again, asking and answering information questions would seem appropriate functions for teaching and testing first- and second-year learners.

Seventy percent had learners roleplay greetings and introductions, and 70% had students express likes and dislikes. Both of these level-appropriate areas are part of the curriculum and of sample tests. Nevertheless, 30% of the instructors did not test them. In addition, 70% of the instructors had learners make requests as part of their roleplay; 30% did not, even though this possibility is also included in the sample tests.

All instructor-raters included narration in present and past tenses, and 90% included narration in the future tense. Again, the discrepancy is interesting, though small, as future-tense narration (be it formal or informal) is included in the sample tests. While present-tense narration could be considered appropriate for the learners' level of proficiency, past- and future-tense narration as Advanced-level tasks are in the realm of practice and goals more than of achievement and mastery.

Ninety percent of the instructors have the learners give directions for going somewhere and instructions for doing something, both of which are included in the sample tests but which vary in level appropriateness. Giving directions for going somewhere is considered Intermediate level, but giving instructions on how to do something can exceed the learners' level of proficiency, depending on the task or topic.

Only 60% of the instructor-raters include the comparison function on their tests, despite its inclusion in the curriculum and the sample tests. Only 60% included hypothesis, and 50% included persuasion. Forty percent included formal situations (work- or profession-related, for example), again despite the fact that there are such items available to them and such material is covered in both the first and second year. These functions are of the Advanced and Superior levels. Evidently some of these instructors have answered the question of how to rate learners on tasks that have been covered, but are beyond their level of proficiency, by eliminating the problem altogether and not including those functions on the speaking test at all.

Formats

There was somewhat greater variation among the instructor-raters on the questions regarding formats. Seventy percent used interviews, 50% used situations without complications, and 40% used situations with complications. Seventy percent used topics. Seventy percent used prepared material, referring to the sample tests distributed to students. Fifty percent also required the use of extemporaneous topics. Eighty percent expected performance at the phrase and sentence level; only 40% expected students to attempt performance at the paragraph level. Only 30% varied formats so that students would have to adapt their language to different registers. Seventy percent used formats that called for giving personal answers, not just general information, and 70% used formats that elicited variable answers. Given that as many as 30% of the instructors do not use some of the formats at all, one could assume that some learners are receiving more well-rounded assessment than others, if not also more well-rounded preparation.

The roleplay formats could be considered more difficult to perform (and to rate) than interviews, which consist of asking and answering questions, but roleplay can be an Intermediate-level format. Therefore, the fact that only half the instructors use roleplay is a concern because that format is the best simulation of real-life use of language. Having the learners speak extemporaneously is also more difficult than adhering to a specific repertoire of material; yet, it too is an essential skill that half of these learners are not attempting on tests. If only 40% of the instructor-raters expect students to attempt paragraph-level speech in the first 2 years of language study, this is yet another decision that has been made regarding what is too difficult for learners. The issue of how to rate the learners on beyond-level tasks does not arise for half of them because the

Table 1
Table of Responses
[Percentages of Positive Responses to Items by Instructor-Raters]

QUESTION THEMES	%	QUESTION THEMES	%	QUESTION THEMES	%
FUNCTIONS TESTED				Require students to speak without hesitation	30
Descriptions:		Situations without complications	50		
Pictures	100			Require students to use vocabulary covered	90
People	100	Situations with complications	40		
Places	90				
Objects	70	Topics	70	Require students to use accurately all grammar covered	60
Ask questions					
Situations	90	Prepared material	70		
Pictures	60				
		Extemporaneous material	50	Require accurate use of past, present, and future	80
Answer Questions	60				
		Sentence -level formats	80		
Roleplay					
Greetings	70			Require students to handle any topics covered	80
Introductions	70	Paragraph-level formats	40		
Express likes/ dislikes	70				
		Formats vary registers	30	Expect coherence	80
Make requests	70				
				Expect cohesion	50
Narration		Require personal information	70		
Present	100			Expect sociolinguistic appropriateness	30
Past	100				
Future	90	Require variable answers	70		
Giving directions	90			Expect students to perform only Novice-Intermediate tasks well	50
		EXPECTATIONS			
Giving instructions	90	More than two errors allowed for an A grade	80		
Comparison	90				
				Expect students to perform Advanced tasks well if covered	30
Hypothesis	60	Require students to pronounce accurately	40		
Persuasion	50				
Formal situations, work-related	50	Require students to pronounce understandably	50	Expect students to perform Superior tasks well if covered	50
FORMATS					
Interviews	70				

tasks simply are not required of them. Adapting language for different registers is also a beyond-level task, but one that is covered in the program. It appears in few of the speaking tests, apparently because it is also thought to be too difficult for the learners.

Instructor-Rater Expectations

The responses to the questions regarding instructor-raters' expectations of learners when rating revealed the greatest variation of all the variables. Eighty percent of the instructor-raters agreed that an A student could have more than one or two errors in a test speech sample, so they began on a similar footing. They were divided on pronunciation, however, with 40% indicating that learners must pronounce accurately and 50% indicating that learners should pronounce understandably but not necessarily entirely accurately. Thirty percent expected learners to speak without hesitation, which could be difficult for Novices even with the sort of semi-prepared repertoire testing used. Ninety percent expected learners to know and use the appropriate vocabulary that had been covered, and 60% expected the accurate use of all grammatical structures covered in the current semester and previously. This latter expectation is especially interesting, given that, as noted previously, many of the structures covered would not be mastered until the learners rose one or two more levels in their proficiency. In previous sections we saw that instructors omitted some functions and formats deemed too difficult; this type of exception occurs at about the same rate for grammar, which is nearly half the time. The one difference, however, at least for the grammar topics included, was the tenses, since accurate performance with past, present, and future was expected by 80% of the instructors.

Eighty percent of the instructors felt that learners should be able to handle all the topics covered. Eighty percent said they expected coherence, and 50%, cohesion, which are Advanced-level expectations. Thirty percent expected sociolinguistic appropriateness, which, while a low figure, nevertheless reflects a group of instructors who have another Advanced-level expectation for Novice and Intermediate learners. Seventy percent say they hold these expectations for the semi-prepared repertoire material, but only 10% hold these expectations for extemporaneous material, which may render the expectations somewhat more reasonable.

Half the respondents expected learners to perform well only on Novice and Intermediate material, and 30% expected them to perform well on Advanced material that had been covered. Only 10% expected learners to perform well on Superior material that had been covered. These responses are not entirely consistent with the percentage of instructors who expected Advanced-level grammar and functions, which was 60%. Thus, many instructors may expect higher-level functioning of students, even though only 30% of them at most marked these Advanced and Superior-level items.

Half the respondents agreed that Novices would perform fairly accurately because they were using

primarily memorized material and that Intermediates would perform relatively less accurately because they were now creating in the language. Apparently the other half of the respondents were not aware that the literature does appear to support those positions. Sixty percent agreed that students would make more errors on new material than on old material, a truism for which we would have expected a greater level of support. Sixty percent agreed that students would make more errors on extemporaneous material than on prepared material, where again we would have expected a higher level of support.

In the comments sections, the only respondents who provided additional information did not add categories of expectations. They merely reinforced the answers they had marked previously by elaborating on the reasons they expected students to speak without hesitation or the reasons they expected grammatical accuracy.

CONCLUSIONS

Clearly, this study reinforces findings in the literature that even with seemingly well-defined expectations for students in syllabi and clarification of expectations through discussion and training for faculty, testing and rating procedures can and do vary. The entire area of beyond-level material evidently needs to be discussed more carefully in this program and most likely in any university or high-school program in which the issue has not yet been raised. Those areas that are being covered in the courses need to be tested, and clarification is needed about how to include them and rate them appropriately. Language-specific expectations for pronunciation and grammar need to be discussed in some detail, just as Richards and Chambers (1996) found. In particular, expectations regarding the use of past and future tenses require further definition, and expectations regarding other structures need to be explored. Instructions, comparisons, hypothesis, and persuasion are other areas for which discussion is indicated. The concerns about fairness to learners that instructors may be expressing by omitting some areas from tests should be addressed so that adaptation and not elimination is the solution. Clarifying and modifying rating rubrics is essential to ensuring greater agreement on what instructors are expecting.

In addition, there clearly needs to be a broader use of varied formats in the testing so that learners form a broader communicative base and so they are not affected by always having to perform in their weakest format, should that be the case. Instructors should include formats that call for adaptation to the situation and interlocutor. As noted, roleplay is often neglected, and it is the format that best affords opportunity to negotiate meaning and attend to sociolinguistic details (Omaggio, 1980). Teachers should give learners opportunities to perform at the paragraph level so that they can be led in that direction. They should ask students to perform extemporaneously as well as with their prepared repertoire in order to facilitate learners' ability to use the language in the real world.

Perhaps we need to develop brief tester-rater manuals for use within our programs with content based on faculty consensus from discussion, training, and further study of the literature. In addition, as Kenyon and Stansfield (1993) suggest, the creation of a set of reference tapes for use within a program could be very beneficial. If instructors had some sample student responses and ratings for the different formats used on each of the tests in each of the courses in the program, their integration into the process when newly hired and their on-going comfort and consistency would be better ensured.

Further discussion and training, including inter-rater reliability training, is needed on all these topics. Even though there will undoubtedly always be some degree of variation from one rater to another, we as professionals have focused heavily on proficiency testing and rather little on prochievement or classroom testing. With an increased focus on the quality of our ratings in the tests we give most frequently, we can only enhance the effectiveness of our programs and our students' achievement and proficiency.

REFERENCES

Byrnes, H. (1987). Proficiency as a framework for second language acquisition. *The Modern Language Journal, 71* (1), 44-49.

Chambers, F., & Richards, B. (1995). The free conversation and the assessment of oral proficiency. *Language Learning Journal, 11,* 6-10.

Clark, J. (1989). Multipurpose language tests: Is a conceptual and operational synthesis possible? *Language teaching, testing, and technology*. Washington, D. C.: Georgetown University Press.

Douglas, D. (1994). Quantity and quality in speaking test performance. *Language International,* 125-143.

Fulcher, G. (1996). Testing tasks: Issues in task design and the group oral. *Language Testing, 13* (1), 23-52.

Gonzalez Pino, B. (1989). Prochievement testing of speaking. *Foreign Language Annals, 22* (3), 478-487.

Hart Gonzalez, L. (1994). Raters and scales in oral proficiency testing: The FSI experience. Paper presented at the Sixteenth Annual Language Testing Research Colloquium, Washington, D.C.

Kenyon, D., & Stansfield, C. (1993) Evaluating the efficacy of rater self-training. *Language Testing Research*. Cambridge: The Fifteenth Annual Language Testing Research Colloquium.

Meredith, R. (1990). The oral proficiency interview in real life: Sharpening the scale. *Modern Language Journal, 74* (3), 288-296.

Mullen, K. (1978). Direct evaluation of second language proficiency: The effect of rater and scale in oral interviews. *Language Learning, 28* (2), 302-308.

Omaggio, Alice. (1980). Priorities for the 1980s. In D. Lange (Ed.), *Proceedings of the National Conference on Professional Priorities* (pp. 47-53). Hastings-on-Hudson, NY: ACTFL.

Richards, B., & Chambers, F. (1996). Reliability and validity in the GCSC oral examination. *Language Learning Journal, 14,* 28-34.

Ross, S. (1987). An experiment with a narrative discourse test. *Language Testing Research*. Monterey, CA: The Ninth Annual Language Testing Research Colloquium. Monterey, California.

Stansfield, C., & Kenyon, D. (1992). The development and validation of a Simulated Oral Proficiency Interview. *Modern Language Journal, 76* (2), 129-141.

Thompson, I. (1995). A study of interrater reliability of the ACTFL Oral Proficiency Interview in five European languages: ESL, French, German, Russian, and Spanish. *Foreign Language Annals, 28* (3), 407-422.

Thompson, I. (1996). Assessing foreign language skills: Data from Russian. *Modern Language Journal, 80* (1), 47-65.

Young, R. (1992). *Expert-novice differences in oral foreign language proficiency*. Paper presented at the Colloquium on Non-Native Speaker International Discourse at the Fourteenth Annual Meeting of the American Association for Applied Linguistics, Seattle, Washington.

APPENDIX
ADMINISTERING AND RATING ORAL TESTS

Check all that apply in your point of view and in the way that you administer and rate oral tests.

1. In oral tests (tests of speaking) for my first- and/or second-year foreign language students, at some point during the year the students must

   _ describe pictures
   _ describe objects
   _ describe people
   _ describe places
   _ ask questions based on a picture
   _ ask questions on topics, such as family, studies, etc.
   _ ask questions to get information about cost, times, etc.
   _ express likes and dislikes
   _ greet others, perform introductions, say farewell
   _ make requests
   _ give information on a varietv of topics
   _ narrate in the present; e.g., say what they do on weekends, during a typical day, etc.
   _ narrate in the past; e.g., say what they did on a weekend, holiday, typical day, etc.
   _ narrate in the future; e.g., say what they will do on a holiday, in their

future work, etc.
_ explain a process, such as how to make a particular dish
_ give directions or instructions., such as how to go from one place to another
_ compare two (or more) pictures, people, places, objects
_ say what they would do in a hypothetical situation
_ try to persuade someone of something
_ use formal language (introduce a speaker, start a formal talk, explain an abstract topic, such as socialism, etc.)

2. When taking speaking tests, my first- and/or second-year foreign language students are expected to

_ interview a partner (another student or the teacher)
_ roleplay situations without complications
_ roleplay situations with complications
_ perform extemporaneously sometimes
_ perform with prepared situations, topics, presentations, etc.
_ vary the way they speak to suit the audience (listener) and the situation (be more or less formal)
_ speak at the phrase level
_ speak at the sentence level
_ give personal answers based on their own information, experiences, preferences, opinions, etc.
_ give variable answers (answering open-ended questions rather than those which have only one right answer, such as what day it is)
_ answer closed questions (those with one right answer)
_ speak in context (on a particular topic or situation)
_ handle a random selection of topics, questions, situations, etc., from a pool of them which have been practiced and/or prepared during the testing period

3. When I rate the speaking tests of my foreign language students in first or second-year courses, I expect students to perform as follows for a grade of A:

_ have only one or two errors in the speech sample
_ pronounce accurately
_ pronounce understandably, but not always accurately
_ speak virtually without hesitation
_ know and use the appropriate vocabulary (with no English)
_ use accurately the grammatical structures that have been covered in this level class and in previous levels
_ complete adequately all the types of tasks or functions that have been covered
_ talk adequately about any and all of the content areas that have been

covered (family, clothing, current events, jobs, etc.)
_ speak in a culturally or sociolinguistically appropriate manner
_ organize their thoughts logically
_ transition appropriately from one idea to another
_ do all of the above when speaking with prepared material
_ do all of the above when speaking extemporaneously
_ perform well only on Novice tasks (ACTFL scale)
_ perform well only on Intermediate tasks
_ perform well on Advanced tasks if they have been covered
_ perform well on Superior tasks if they have been covered
_ accurate performance from Novices because they are memorizing their material
_ less accurate performance from Intermediates because they are now creating in the language
_ more errors on new material than on old
_ more errors on extemporaneous formats
_ fewer errors on prepared formats
_ phrase- or sentence-level or length performance or responses
_ paragraph-level or length performance or responses

4. Other: Your additional comments about what you expect from first and/or second-year foreign language students on their speaking tests:

Current Issues in the Spanish Language Proficiency of Bilingual Education Teachers

MICHAEL D. GUERRERO, The University of Texas at Austin

The purpose of this paper is to examine some critical issues regarding the Spanish language proficiency of bilingual education teachers, primarily those teachers from the Spanish language-origin community. Recent longitudinal studies have demonstrated a positive relationship between sustained native-language instruction and student achievement. Because this finding is encouraging, it is time to take a closer look at the context in which bilingual education teachers develop their Spanish language proficiency. This examination reveals that given the present subtractive sociolinguistic context in the U.S., the likelihood of bilingual teachers developing native-like Spanish language proficiency is an uphill battle. Further, the Spanish language preparation bilingual education teachers receive at institutions of higher education is not commensurate with the task of developing a high level of Spanish language proficiency. Finally, the Spanish language proficiency measures used to gauge the Spanish language proficiency of bilingual education teachers are not without their problems. Under the present circumstances, bilingual education teachers with the ability to provide sustained native-language instruction will continue to be the exception.

INTRODUCTION

There is little doubt that the longitudinal findings of both the Ramirez, Yuen, and Ramey (1991) and Thomas and Collier (1997) studies provided sorely needed empirical evidence that the sustained use of the English learner's native language, in this instance Spanish, is a key variable positively associated with student achievement. Clearly, there are other variables that probably contribute to the success of the learner, such as program design, parental involvement, a shared vision among staff, appropriate assessment practices and strong leadership provided by the principals. The use of the native language, however, arguably assumes a more central role in the success of the learner. Language issues must be considered while the school staff designs the program. Similarly, language issues are probably considered as the staff engages in developing and implementing parental involvement activities, assessment policies, and vision building. Principals probably work diligently to recruit and hire highly qualified staff, especially staff with solid Spanish language skills.

The most central of all the language decisions made in a bilingual program relate to the classroom teacher's use of the Spanish language for instructional purposes and the ability of the teacher to use the language for academic purposes. This centrality hinges on the fact that it is the classroom teacher who spends the majority of the school day with the learners in attempting to implement the bilingual program.

The conscious and deliberate consideration of the role of the non-English language in successful bilingual education programs is symptomatic of the implementation of such programs in a society that has a strong subtractive (Lambert, 1977) and linguicist orientation (Phillipson, 1988). This subtractive orientation is best exemplified by the fact that few schools aim to maintain and continue developing the learner's first language. These programs are generally referred to as transitional bilingual education programs as they aim to transition the learner from Spanish language instruction to English language instruction as quickly as possible and with no long term commitment to native language development.

Regarding the notion of linguicism, Phillipson (1988, p. 341) states

> The *forms* that linguicism takes are many. For instance, structural linguicism may be *overt*, e.g. use of a given language is prohibited in institutional settings such as schools. Or linguicism may be *covert*, e.g. certain languages are *de facto* not used in teacher training, or as languages of instruction, or in aid activities, even if use of the languages is not explicitly forbidden. The prevailing ideology may be *consciously* linguicist, e.g. teachers instruct pupils not to use their mother tongue, because they are under the delusion that a ban of this kind will help the learning of another language.

The point is that bilingual education teachers are not immune to the subtractive and linguicist orientation of U.S. society. The majority of these teachers and their families were schooled in the U.S. and hence subjected to the array of language practices that perpetuate language shift and loss among speakers of Spanish language origin. In short, the implementation of a bilingual education program that includes sustained native-language instruction goes against the linguistic grain of this country and will require much conscious and deliberate planning on the part of the school staff.

Providing sustained native language instruction logically entails the availability of bilingual education teachers with the facility to do so. The development of this facility, however, is contingent upon meaningful language development opportunities, which may be hard to come by through mandatory (K-12) public schooling in the U.S. Consequently, most bilingual education teachers will rely on the required language related course work at a teacher training institution to develop this skill. Unfortunately, universities and colleges are part of the same subtractive sociolinguistic milieu. Spanish language development requires

time, more time than most post-secondary institutions are able or willing to offer prospective bilingual education teachers.

The irony of the situation resides in the fact that well intentioned state departments of education require bilingual education teachers to pass a formal Spanish language examination. The nature of the tests, however, varies greatly (Grant, 1995), and this author believes that their construct validity is open to question. More importantly, the language standards inherent to these measures may not be adequate for making a judgment about an individual's ability to deliver sustained native-language instruction. The social consequences generated by using these tests must be carefully examined.

Sustained native-language instruction is desirable, but the social conditions for providing bilingual education teachers the opportunities they need to develop native-like proficiency in Spanish are not widely available. It is the intent of this paper to begin placing these issues into proper perspective.

SUSTAINED NATIVE LANGUAGE INSTRUCTION

Collier (1995, p. 3) makes the statement that

> To assure cognitive and academic success in a second language, a student's first language system, oral and written, must be developed to a high cognitive level at least through the elementary school years.

One can assume that, in order for learners to develop their first language to a high cognitive level, the learners must also have access to this level of language or to target-language speakers, especially teachers (Wong Fillmore, 1989). It is the teacher who carries the responsibility of modeling spoken and written Spanish that will provide even native Spanish-speaking children the opportunity to develop further their oral and literacy skills.

Collier also maintains that in order for a learner to develop academically, in an efficient manner, the student must receive instruction in the native language. Academic development includes growth in each of the content areas such as math, science, and social studies. The implication is that the classroom teacher will also serve as a key target-language speaker from whom the learner should be able to acquire academic language proficiency in the non-English language.

Stated differently, the bilingual education teacher should be able to deliver instruction in the non-English language across the curriculum as well as the mainstream classroom teacher does in English (Gaarder, 1977). This ability will likely transcend a simple knowledge of technical vocabulary in the content areas. It will require the ability to comprehend (listen and read) and produce (speak and write) the non-English language with appropriate syntax, cohesive markers, rhetorical organization, functions, gestures, figures of speech, and cultural references, all of which may vary depending on the subject matter taught. Trueba (1989, p. 113) adds:

> In bilingual education, lack of mastery of the language of instruction causes serious problems for the teachers; it affects their classroom management, their clarity in explaining subject matter, and the quality of relationships with native speakers of that language. If a teacher does not know the target language well, children's linguistic and cognitive development also suffers, because they are deprived of guidance and feedback in situations where correct and precise use of the language is required to understand a concept or the logical foundations of reasoning.

Again, there are other factors, such as program design, methodology, assessment, and educational policies, that can influence the academic success of English-language learners. Nonetheless, no one can deny that each classroom teacher occupies a critical role in the academic success of English learners. Moreover, it is not the language ability of each teacher in isolation that matters, but rather the collective, consistent, and sustained use of the native language over several years that will determine student outcomes as Thomas and Collier's research suggests (1997).

BEFORE THEY BECOME BILINGUAL TEACHERS

The main position of this paper is that prospective bilingual education teachers from the Spanish-language-origin community often do not reach expected levels of proficiency in the Spanish language due to a variety of factors. Basically, prospective bilingual education teachers are members of the wider society and are subjected to the same subtractive and linguicist practices and policies as everybody else. These practices, unfortunately, begin to impact the prospective bilingual education teacher negatively at a very early age and continue to do so throughout their public education experience.

Even before schooling begins, Spanish-speaking parents struggle with the decision as to whether or not they should teach their children Spanish. Grosjean (1982, p. 124) states, "in the United States, there are innumerable examples of immigrant parents encouraging, if not forcing, their children to learn English, with the potential consequence that some may become rootless and alienated from their native language group."

From a linguistic perspective, and to the degree to which Spanish-speaking parents withhold linguistic input from their young children, it is at this point that the lion's share of the damage may occur. According to the model of communicative language ability set forth by Bachman (1990), linguistic input that begins to shape the organizational and pragmatic competence of the young child may be withheld. The linguistic and social consequences, as Hernández-Chávez (1993, p. 58) states, are that

> Large numbers of Chicano children and young people from Spanish speaking families either no longer learn the language or acquire but a limited facility in it. As a result, patterns of communication are disrupted, cultural and social structures break down and youth become alienated from their communities.

A frequently cited reason that Spanish-speaking immigrants do not transmit the Spanish language to their children is rooted in the parents' belief that if their children learn English, they will secure good jobs and prosper. Peñalosa (1980) and Zentella (1990) argue that this belief is more a myth than reality. Chicanos and Puerto Ricans continue to be economically marginalized even after acquiring English.

The economic argument aside, Spanish-speaking immigrant parents also receive numerous messages from different components of society indicating that their children should be taught only English. The present movement to make English the official language of the U.S. is a case in point. The recent judicial case in which a judge equated a mother's speaking Spanish to her young daughter with child abuse is yet another (Morales, 1995). Further, the parents themselves may have been victimized for using Spanish at school.

As young children from the Spanish language community enter schooling, the message to abandon the Spanish language is further reinforced. Wong Fillmore (1991, p. 20), in a compelling study of preschool programs designed to serve language minority children, concludes that many of these children lose their primary language as they learn English. The researcher explains

> Consider what happens when young children find themselves in the attractive new world of the American school. What do they do when they discover that the only language that is spoken there is one that they do not know? How do they respond when they realize that the only language they know has no function or value in that new social world, and that in fact, it constitutes a barrier to their participation in the social life of the school? They do just as the promoters of early education for language minority students hope they will. They learn English, and too often, they drop their primary languages as they do. In time, many of these children lose their first languages.

Unfortunately, there are also few opportunities offered through the K-12 educational system in this country to promote the maintenance and development of non-English languages among school age children. In a study conducted by the U.S. Department of Education (1993), a number of findings relevant to this discussion were reported. The study found that

1. Only 17% of schools provide a significant degree of primary language instruction.

2. ESL is the predominant instructional approach.

3. Of the 363,000 teachers providing services to Limited English Proficient students, only 10% are certified bilingual teachers.

4. The majority of teachers serving Spanish-speaking pupils have no proficiency in Spanish.

With regard to the finding that the majority of teachers that serve Spanish-speaking pupils lack profi-

ciency in the language, even those few students that do find their way into a bilingual program cannot count on having the kind of access they need to continue developing their Spanish language academic proficiency. Escamilla (1992) studied various features of 25 elementary bilingual maintenance programs over a 2-year period. With regard to the uses to which Spanish and English were put, the researcher reports that in some classrooms Spanish was used primarily for direction giving and discipline. English was used for academic instruction and conversation.

It should also be noted that bilingual education in the U.S. is most readily associated with elementary school programs as opposed to secondary education programs (Faltis & Arias, 1993). Not only are there proportionately fewer bilingual education programs at the secondary level, but there are also fewer programs that are aimed at continued development of the learner's Spanish language skills.

Consequently, with each successive year of schooling, the likelihood of opportunities for prospective bilingual education teachers to develop academic Spanish language proficiency is further reduced. The result of such an educational experience for the majority of members from a Spanish-language-origin community is language shift that generally results in language loss. A number of studies using U.S. Census data support the trend of language shift and loss among Spanish-language-origin people in the U.S. (Bills, 1989; Veltman, 1988; Hernández-Chávez, 1996).

These findings are especially important since it is youth of Spanish language origin who will probably become bilingual teachers. Fewer and fewer members of this group will raise their children to speak Spanish, and those that do may pass on a model of Spanish-language proficiency unlike that of native speakers. This trend is facilitated by the lack of high-quality bilingual education programs in the U.S. In short, the pool from which to draw proficient speakers and writers of Spanish is continuously dwindling.

Merino and Faltis (1993) indicate that sustained native-language instruction appears to be contingent upon two factors, teacher language proficiency and the implementation of a well-articulated, late-exit (K-6) bilingual education program. In short, with so few exemplary, developmental bilingual education programs, it is unlikely that the pressing demand for prospective bilingual education teachers that are proficient in Spanish will be met through education. Moreover, with continuously decreasing numbers of individuals proficient in the Spanish language, it is unlikely that any meaningful number of programs designed to use sustained native language-instruction could ever be implemented at any given point in time.

This discussion would be incomplete if some thought was not given as to why there are so few exemplary bilingual programs in the U.S. and why there is a dire need for well-trained bilingual education teachers with native-like academic proficiency in Spanish. The reason is arguably intimately related to the societal value placed on bilingualism

in this country. Kjolseth (1983, p. 48) maintains

> We are not bumpkins but quite ordinary and normal humans who develop language skills when they are effectively called for, and do not when they are not. And although easily made the scapegoat, our schools are not to blame, because schools reflect the cultural policies —i.e., values of our dominant groups—and are merely the places where our main cultural myths are translated into curriculum.

The language values of the dominant groups in this society are, as Kjolseth (1983) has described, schizophrenic. On the one hand, in this society it is admirable when native speakers of English learn non-English languages through foreign language study, even as imperfect as their mastery of the languages will be. On the other hand, before members of a non-English-language-speaking group can perfect their native language through schooling, they must postpone, perhaps abandon, this endeavor and first (or only) acquire English. The point is that, if the dominant groups in our society truly value bilingualism, it must be supported in a manner that is logical and that generates the best results. Lyons (1990, p. 79) explains,

> [C]onsider that an undergraduate student preparing to be a teacher would receive in four years only 600 hours, at five hours per week, of foreign language instruction. The average graduate of such a teacher-training program lacks the skills to use properly, much less teach, a foreign language to children. Only rarely would he or she possess foreign-language skills suitable for the "imitative capacities of young children."
>
> Time could be turned to our advantage, however, if we were to conserve, develop, and capitalize on the language skills of the language minority students in our schools. These skills, developed through tens of thousands of hours of mother tongue instruction, offer both a quick fix and a long term solution to the problem of American monolingualism.
>
> The average language minority child entering kindergarten has a higher level of language mastery than the average graduate of the intensive and expensive 47 week Defense Language Institute program.

It is beyond the scope of this discussion to examine closely why such language policies are firmly rooted in U.S. society. Is it out of nationalism, compassion, ignorance, linguicism, or simply the need to sustain a steady supply of individuals to fill undesirable and low-status jobs (Spener, 1988)? Perhaps a case could be made for each rationale. Regardless of the motivation, it is clear that when it comes to cultivating language resources in the U.S., the orientation is subtractive and linguicist.

In sum, prospective bilingual education teachers must survive the sociolinguistic forces that gradually and predictably deteriorate the Spanish language abilities of the general Spanish-language-origin community even before schooling begins. Schooling, bilingual education programs, and foreign language training in particular do little to enhance the Spanish language abilities of prospective bilingual education teach-

ers, because they are predicated upon illogical premises, ideologies that over the decades have only proven how ineffective they are. Consequently, it is unreasonable to expect prospective bilingual education teachers to have gained an age-appropriate level of academic Spanish language proficiency prior to teacher training.

THE ROLE OF INSTITUTIONS OF HIGHER EDUCATION

Virtually no research has been dedicated to the design of bilingual education teacher training programs, including the language training component of these programs; that is, little research has been conducted that might provide empirical evidence for the effectiveness of such programs. Rodriguez (1980, p. 372) states,

> Legislative regulations and State Board of Education guidelines press teacher trainers with myriad lists for bilingual teacher competencies. While all such competency lists are said to be synonymous with effective bilingual teachers, they are vulnerable to criticism for several reasons. To begin, there is as yet little or no empirical evidence that existing competencies are valid. Most competencies for bilingual education teachers are generated by experts.

Little appears to have changed since Rodriguez made this statement. Grant (1992, p. 431) observes that

> While the lack of a substantial body of solid research is a serious problem in teacher education in general, it is a doubly serious problem when it comes to research on the preparation of teachers to work in culturally diverse schools, especially when that preparation includes working with limited English proficient (LEP) students.

Dalton and Moir (1992, p. 416) speak more specifically to the paucity of research on the effectiveness of bilingual education teacher-training programs. These authors state,

> It appears that in practice little program evaluation is specifically designed for internal use in program improvement or to increase understanding about developmental processes. This means that the suitability of teacher education curricula for the communities served, the effect of the program on professional and LEP student consumers, and experiences of program participants remain largely unexplored.

What can be inferred from these observations is that little is also known about the effectiveness of different approaches on developing the prospective bilingual education teacher's Spanish language proficiency. As previously stated, many teachers instructing Spanish-speaking students have no proficiency in Spanish (U.S. Department of Education, 1993). Assuming the teachers to which this finding applies have already taken the required course work to instruct in a bilingual setting, it is also safe to assume that the Spanish language training they received did not fully meet their needs. This trend is not new. A decade earlier, Waggoner and O'Malley

(1984, p. 25) reached a similar conclusion:

> [A]pproximately four out of five teachers using a non-English language in instruction during 1980-81 did not have the language skills or basic professional preparation to do so.

The fact of the matter is that there are many fundamental empirical questions that must be explored if this situation is to begin changing. Consider, for example, the number of courses required of prospective bilingual education teachers that are taught in Spanish. Is there an empirical rationale for establishing a set number (e.g., two or three) of language-related courses? What evidence is there that the amount of course work offered in Spanish is commensurate with the language goals of the bilingual education teacher training program?

The content of the courses offered in Spanish is an equally important and related question. Prospective bilingual education teachers need to have opportunities to develop their Spanish language academic proficiency. How to assist prospective bilingual education teachers with the development of their academic proficiency in Spanish is yet another fundamental empirical question.

Regarding the quality of the courses generally offered in Spanish, at least one related issue should be raised. Faculty who offer courses taught in Spanish must assume the role of a language model in much the same way K-12 bilingual education teachers must for their students. Whether or not the faculty possess the requisite language skills also remains an important question. Assuming that the majority of bilingual education faculty were schooled in the same subtractive sociolinguistic milieu, it may well be that faculty are also in need of further opportunities to develop their own academic Spanish language proficiency.

The bilingual education teacher training practice of requiring prospective bilingual education teachers to take Spanish language courses through a Foreign Language Department must also be carefully examined. Empirical evidence is needed that can shed light on the effectiveness of this long-standing practice. In what ways does this kind of course work aid prospective bilingual education teachers to meet the linguistic demands of a bilingual education setting? Overall, are the Spanish language learning opportunities provided to the prospective bilingual education student teacher sufficient to meet this demand?

The culminating experience for a prospective bilingual education teacher is student teaching within a bilingual setting. Assuming that the student teacher is placed in a bilingual education program that uses sustained Spanish language instruction, perhaps the most critical empirical question could be addressed. What kinds of language skills are needed in order to provide sustained native language instruction?

In sum, available data (U.S. Department of Education, 1993; Waggoner & O'Malley, 1984) indicate that the Spanish language development opportunities offered through bilingual education teacher training programs are less than adequate given

the needs of the participants. Part of the reason for this dilemma stems from the use of pre-service language development opportunities that have not been empirically substantiated. The professional judgment of college and university faculty is a reasonable starting point for setting up the language component of a bilingual teacher training program, but this judgment must also be tested for its validity.

SPANISH LANGUAGE PROFICIENCY TESTS

There are a number of states that have implemented a Spanish language testing policy for prospective bilingual education teachers (Grant, 1995). The assumption is that the prospective bilingual education teacher will have been prepared to meet the language demands of the test during their teacher-training experience. One can also assume that the intent of such a policy is to ensure that a bilingual education teacher is proficient enough in the Spanish language to fulfill the linguistic demands associated with a bilingual education classroom or program. Nonetheless, it is the validity of the test used on which the value of these kinds of policies depends.

It is safe to say that the social consequences (Messick, 1989) associated with the use of these kinds of tests, for the learners and for society in general, are considerable. If the tests are valid and measure what they purport to, then the social consequences associated with their use will be positive in most cases. Only those teachers who are able to teach across the curriculum will find their way into the classroom. By extension, learners will be more likely to achieve academically.

Unfortunately, these tests do not appear to be fulfilling their function. Recall that over the last two decades empirical findings suggest that bilingual education teachers generally have a less than adequate command of the Spanish language (U.S. Department of Education, 1993; Waggoner & O'Malley, 1984). How could this be if tests are in place, at least in some states, to ensure that the bilingual education teacher is proficient in the Spanish language? The obvious explanation is that the tests are of questionable validity.

It is beyond the scope of this short paper to report on the psychometric properties of each of the tests currently used in the U.S. The point to be made here is a general one, but a critical one. These high-stakes tests must possess construct validity, and this test quality is intimately linked to instances of its use in the appropriate setting. The vast majority of bilingual education programs in the U.S., however, are transitional bilingual programs at grades K-3. The goal of this type of program is to transition the learner into all English instruction as soon as possible. Further, and as stated many times in this paper, many practicing bilingual education teachers lack proficiency in the language. Consequently, as these tests are developed, and bilingual teachers are observed by the test developers to examine which language abilities are used, how they are used, and the level of proficiency modeled by the teacher, the construct validity of the test is shaped. The end result is a test with construct validity based on

weak language models operating within an educational context with English, not bilingualism, as the ultimate goal.

Guerrero (1994) examined the unified validity (Messick, 1989) of the Spanish language proficiency test designed for bilingual education teachers in New Mexico. In terms of this measure's subtractive orientation, an examinee can pass the written part of the test, a letter to parents consisting of at least 150 words, with as many as 20 errors. As another example, excerpts used to measure the reading ability of teachers were taken from no higher than fourth grade text books (Valdés, 1989). The test is presently used to endorse teachers at all grade levels (K-12). One of the more disturbing findings is that approximately 80% of the examinees (n= 217) taking the test for the first time did not pass the test. Stated differently, the majority of this sample were not amply prepared to meet the relatively low level demands of this test.

Norfleet (1994) examined the reliability and validity of the Spanish language proficiency test used for bilingual endorsement purposes in Arizona public schools. The general conclusion reached by the researcher is that the test, developed in 1981, continues to serve its intended purpose. Norfleet (1994, p. 238) explains,

> Although some of the results indicate that the test appears to be accomplishing its main objective, the measurement of the ability to use Spanish in the bilingual classroom, major revisions for the ACTSPE [Arizona Classroom Teacher Spanish Proficiency Exam] are essential in other areas.

The same subtractive orientation can be detected in this test as well. For example, the test was designed for the elementary grades (Barkin-Riegelhaupt, 1985), but the same test is also used to measure the Spanish language proficiency of prospective bilingual education teachers at all levels (K-12). Further, the oral parts of the test are weighted more heavily in scoring the test than the parts involving literacy. The message conveyed is that the ability to speak the Spanish language is more important than the ability to write it. In the present context of transitional bilingual education in the U.S., this message is accurate.

Grant (1995) indicates that 28 states across the country offer either certification or endorsement in bilingual education. Unfortunately, twelve of these states do not test for teacher language proficiency, and three states measure only oral language proficiency. Seven states allow the bilingual education teacher preparation institutions to establish their own language testing procedures and criteria. These practices, however, do not mean that each institution within a given state adheres to the same procedures and criteria. Only six states have adopted tests that entail more than one language measure (e.g., for speaking, reading, writing, or culture). Grant (1995, p. 5) seems to suggest that only two states, Arizona and California, have developed tests that target "the proficiency needed by bilingual teachers for teaching." New Mexico is currently using a test that was also

intended to be linked to the classroom uses of Spanish (Valdés, 1989).

Throughout this country, the Spanish language proficiency of bilingual teachers seems to equate with only oral proficiency. Further, based on the variety of language measures used, there is little consensus, regionally or nationally, regarding what this ability entails. In the few cases where Spanish literacy skills are required, there appears to be a propensity for the standards to be lower than what might be expected of an English-speaking teacher and to assign less value to the literacy skills than to oral skills. Last, in the cases of states with no language testing policy at all, the whole issue of Spanish language proficiency is simply disregarded.

In sum, there are a number of states that have implemented language testing policies for prospective bilingual education teachers. While the spirit of these policies is well-intentioned, the policies reflect the subtractive orientation of bilingual education programs in the U.S. In effect, these tests help perpetuate less than adequate Spanish language abilities not only among prospective and practicing bilingual education teachers, but also among the students they will teach. Ada (1986, p. 390) speaks to this dilemma when she states,

> Bilingual teachers may feel inadequate in their language ability because of several factors. Those teachers whose mother tongue is English may not have had the opportunity to acquire full mastery of a second language—a sad reflection on our limited and deficient foreign language teaching. Members of language minorities who chose to become bilingual teachers may also have been victims of language oppression as children, when they were scolded or punished in school for using their home language. Therefore, it should not be surprising that many bilingual teachers lack confidence in their literacy skills. Yet if these individuals can acknowledge that the language inadequacy stems from deeply rooted institutionalized oppression. . ., they will be better able to understand what their students may be going through.

CONCLUSION

This paper has attempted to examine some of the critical issues related to the Spanish language proficiency of bilingual teachers. For the majority of bilingual education teachers, and within the present subtractive sociolinguistic context, the development of advanced level, teacher-like proficiency in Spanish is an ambitious goal. U.S. society, and schooling in particular, make it especially difficult for bilingual education teachers to develop the ability to provide sustained native-language instruction.

The burden of developing this ability is presently placed on bilingual education teacher training programs. Unfortunately, these programs attempt to meet the language needs of prospective bilingual education teachers based on language practices with little or no demonstrated empirical support. This language practice is much in line with those used for children of limited English proficiency in the majority of school programs; that is, in both cases, edu-

cators assume that the learners will acquire academic proficiency in the target language quickly or can forego native language instruction altogether.

With regard to the policy of mandating Spanish language testing for bilingual education teachers, the policy can only be as valid as the instrument used. Presently, the professional language norm upheld through the use of these tests is subtractively oriented, much in keeping with the majority of existing bilingual education programs. Nonetheless, if an individual manages to meet the prescribed expectations, the perception is that this individual is able to fulfill the language demands of a bilingual setting. In effect, this individual probably can fulfill these demands since the majority of bilingual education programs are early elementary programs and seek to transition the learner to all English instruction as quickly as possible.

The promise of sustained native-language instruction is great, too great to ignore or neglect. Before more children can benefit from bilingual education programs with an additive orientation, however, many fundamental linguistic changes must take place in the experiences, practices, and policies that affect prospective bilingual education teachers in this country.

REFERENCES

Ada, A. (1986). Creative education for bilingual teachers. *Harvard Education Review, 56* (4), 386-394.

Bachman, L. (1990). *Fundamental considerations in language testing*. Reading, MA: Oxford University Press.

Barkin-Riegelhaupt, F. (1985). Testing bilingual language proficiency: An applied approach. In L. Elias-Olivares (Ed.), *Spanish language use and public life in the U.S.A.* (pp. 165-180). Berlin: Mouton de Gruyter.

Bills, G. (1989). The U.S. Census of 1980 and Spanish in the Southwest. *International Journal on the Sociology of Language, 79*, 11-28.

Collier, V. (1995). Acquiring a second language for school. *Directions in Language & Education, 1* (4). Washington, DC: National Clearinghouse for Bilingual Education.

Dalton, S., & Moir, E. (1992). Evaluating limited English proficient (LEP) teacher training and in-service programs. *Second National Research Symposium on limited English proficient student issues: Focus on evaluation and measurement*. (Vol. 1, pp. 415-453). Washington, D.C.: Office of Bilingual Education and Minority Languages Affairs.

Escamilla, K. (1992). Theory to practice: A look at maintenance bilingual classrooms. *The Journal of Educational Issues of Language Minority Students, 11*, 1-25. Boise, ID: Boise State University.

Faltis, C., & Arias, M. (1993). Speakers of languages other than English in the secondary school: Accomplishments and struggles. *Peabody Journal of Education, 69* (1), 6-29.

Gaarder, B. (1977). *Bilingual schooling and the survival of Spanish in the United States*. Rowley: MA: Newbury House.

Grant, C. (1992). Educational research and teacher training for successfully teaching limited English proficient students. *Proceedings of the Second National Research Symposium on limited English proficient student issues: Focus on evaluation and measurement* (Vol. 2, pp. 431-473). Washington, D.C.: Office of Bilingual Education and Minority Languages Affairs.

Grant, L. (1995). Testing bilingual teacher's language proficiency: The case of Arizona. (Educational Resource Information Center) ED 384 222.

Grosjean, F. (1982). *Life with two languages. An introduction to bilingualism*. Cambridge: Harvard University Press.

Guerrero, M. (1994). *A critical analysis of the validity of the Four Skills Exam*. Unpublished doctoral dissertation. University of New Mexico, Albuquerque, N. Mex.

Hernández-Chávez, E. (1993). Native language loss and its implications for revitalization of Spanish in Chicano communities. In B. Merino, H. Trueba, & F. Samaniego (Eds.), *Language and culture in learning: Teaching Spanish to native speakers of Spanish* (pp. 58-74). Washington, D.C.: Falmer Press.

Hernández-Chávez, E. (1996). La pérdida del español entre los chicanos. Paper presented at the Annual New Mexico Association for Bilingual Education. Albuquerque, NM.

Kjolseth, R. (1983). Cultural politics of bilingualism. *Transaction Social Science and Modern Society, 20* (4), 40-48.

Lambert, W. (1977). The effects of bilingualism on the individual: Cognitive and sociocultural consequences. In P.A. Hornby (Ed.), *Bilingualism: Psychological, social and educational implications* (pp. 15-27). New York: Academic Press.

Lyons, J. (1990). The past and future directions of federal bilingual education policy. *ANNALS, AAPSS, 508*, 66-80.

Merino, B., & Faltis, C. (1993). Language and culture in the preparation of bilingual teachers. In M.B. Arias & U. Casanova (Eds.), *Bilingual education: Politics, practice, research* (pp. 171-198). Chicago: University of Chicago Press.

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

Morales, A. (1995, August 30). Lawmaker blasts judge's comments on language. *Austin American Statesman*, p. B1.

Norfleet, L. (1994). *The reliability and validity of the Arizona Classroom Teacher Spanish Proficiency Exam*. Unpublished doctoral dissertation. Northern Arizona University, Flagstaff, AZ.

Peñalosa, F. (1980). Chicano bilingualism and the world system. In R. Padilla (Ed.), *Theory in bilingual education. Ethnoperspectives in bilingual education research* (Vol. II, pp. 3-17). Ypsilanti, MI: Eastern Michigan University Press.

Phillipson, R. (1988). Linguicism: Structures and ideologies in linguistic imperialism. In T. Skut-

nabb-Kangas (Ed.), *Minority education: From shame to struggle* (pp. 339-358). Clevedon, England: Multilingual Matters.

Ramírez, D., Yuen, S., & Ramey, D. (1991). *Final report: Longitudinal study of structured English immersion strategy, early-exit and late-exit transitional bilingual education programs for language minority children*. San Mateo, CA: Aguirre International.

Rodríguez, A. (1980). Empirically defining competencies for effective bilingual teachers: A preliminary study. In R. Padilla (Ed.), *Theory in bilingual education. Ethnoperspectives in bilingual education research.* (Vol. II, pp. 372-387). Ypsilanti, MI: Eastern Michigan University Press.

Spener, D. (1988). Transitional bilingual education and the socialization of immigrants. *Harvard Education Review, 58* (2), 133-153.

Thomas, W., & Collier, V. (1997). *School effectiveness for language minority students*. Washington, D.C.: National Clearinghouse for Bilingual Education.

Trueba, H. (1989). *Raising silent voices: Educating the linguistic minorities for the twenty-first century*. Boston: Heinle & Heinle.

U.S. Department of Education. (1993). *Descriptive study of services to limited English proficient students*. Washington, D.C.

Valdéz, G. (1989). Testing bilingual proficiency for specialized occupations: Issues and implications. In B.R. Gifford (Ed.), *Test policy and test performance: Education, language and culture*, (pp. 207-229). Norwell, MA: Kluwer Academic.

Veltman, C. (1988). *The future of the Spanish language in the United States*. New York City & Washington, D.C.: Hispanic Policy Development Project.

Waggoner, D., & O'Malley, J. (1984). Teachers of limited English proficient children in the United States. *Journal of the National Association for Bilingual Education, 9* (3), 25-42.

Wong Fillmore,L. (1989). Language learning in social context. The view from research in second language learning. In R. Dietrich & C. Graumann (Eds.), *Language processing in social context*, (277-302). North Holland: Elsevier Science Publishers.

Wong Fillmore, L. (1991). When learning a second language means losing the first. *Early Childhood Quarterly, 6*, 323-346.

Zentella, A. (1990). El impacto de la realidad socio-económica en las comunidades hispanoparlantes de los Estados Unidos: Reto a la teoría y metodología lingüística. In J. Bergen (Ed.), *Spanish in the United States: Sociolinguistic Issues* (152-166). Washington, D.C.: Georgetown University Press.

The Multiple Challenges of Multimedia: Development, Implementation, and Evaluation

MARGARET ANN KASSEN, The Catholic University of America

This article presents an overview of an ongoing initiative to develop, integrate, and evaluate multimedia lessons to enhance listening comprehension in Spanish language classes at The Catholic University of America. In an effort to focus on the process involved, the two phases that have been completed are examined, including the rationale for the project, the selection of the Libra authoring system, the building of the authoring team, the design and revision of the lessons, and their integration into the curriculum. Findings from the evaluation components are explored with particular emphasis on the evidence pointing to the need to understand better how learners actually interact with multimedia. Based on the lessons learned from this project, guidelines are suggested for those who may be interested in pursuing a similar undertaking.

INTRODUCTION

In the second language teaching field today, there is great interest in technology and its potential to enhance second language learning. Multimedia has a particular appeal, due in part to what Pusack and Otto (1997) refer to as its "capacity to access and control via computer a full range of familiar media: text, motion video, photo images, sound and graphics" (p. 2). Many of the materials teachers have previously used effectively to promote language learning can be placed at the learners' fingertips in the digital environment with the promise of increased learning outcomes. The allure of multimedia must be tempered, however, with the realities that confront any instructional innovation: the challenges of development, implementation, and evaluation. How does one go about developing multimedia lessons for use in multiple-section courses? How can these lessons be integrated in the existing curriculum? How can multimedia be evaluated? These fundamental questions will be examined in the context of a multimedia project for beginning level Spanish undertaken at The Catholic University of America (CUA).

BACKGROUND OF THE PROJECT

The Department of Modern Languages at CUA was interested in exploring language learning technology, particularly as a way to enhance students' out-of-class language learning. This option had become feasible due to our recently installed Multimedia Language Center, a small, independent study room equipped with computers, VCRs, laserdisc, etc. As a first step, the media director and language coordinator attended a workshop on Libra, a multimedia authoring system designed to facilitate listening comprehension. Libra was developed by Robert Fischer and Mike Farris at Southwest Texas State University at San Marcos [1] with the support of the Fund for the Improvement of Post-Secondary Education (FIPSE). Because the listening comprehen-

sion goals of CUA's Spanish program *Destinos* were compatible with those of Libra, the team applied for and received a grant to develop, implement, and evaluate Libra lessons for *Destinos* as a part of Fischer's FIPSE project.

LISTENING COMPREHENSION AND TECHNOLOGY

A focus on listening comprehension at the novice and intermediate levels is well supported by research and theoretical models that point to comprehensible input and intake as fundamental to second language acquisition. As Rubin (1994) noted in her review of the relevant literature, the actual process of listening has received considerable research attention. Comprehension has come to be seen as a dynamic, interactive, cognitive process in which listeners construct mental representations of their understanding of texts, using and coordinating both concept-driven (top-down) and sentence and word-level (bottom-up) processes. Furthermore, in real-world, face-to-face communication, listeners have immediate access to information about the context of their interactions (participants and setting) and paralinguistic cues (such as eye movement and gestures), and they generally have some control of the interaction by means of negotiation and use of strategies (Joiner, 1997).

Given this understanding of the listening process, how can technology best be brought to bear on listening skill development? While traditional cassettes provide the audio component and video the visual information, these media do not facilitate the interaction of learners with the text other than through their cumbersome replay option. Multimedia, however, because of its digital nature, allows direct access to and manipulation of particular segments of text and, by means of various help features, offers the user added options for managing the interaction. As Joiner (1997) notes,

> Computer-assisted multimedia comes closer than the other audio and visual media to meeting the standard for listening embodied in face-to-face communication. Presence, interactivity, control, multisensory input, and multiple sources of assistance can be incorporated into this sophisticated technology. (p. 90)

LIBRA AUTHORING SYSTEM

Libra is a hypercard-type program for Macintosh [2] that allows the teacher to create a stack of question cards written to target selected segments of a videotext. These question cards appear on the computer screen and are linked to video clips that have either been digitized and are accessible on the same screen or that are shown on a laserdisc player, as was the case in this study. The Libra documentation recommends beginning each lesson with a textmap to acquaint learners graphically with the structure of the text. These advance organizers may reappear before each segment to lead users from one segment to another logically (see Figure 1). The five question templates available in the program are flexible and may be used to focus learners' attention on various aspects of listening comprehension,

from main ideas to details and specific words. For multiple choice, checklist, binary checklist, and icon sorting questions (see Figures 2, 3, 4, and 5), learners select the correct responses by clicking or dragging and get immediate feedback. The fifth question type is open-ended and allows users to create their own answers by typing into a notebook. While the teacher sets parameters by deciding which portion of the video is available on a given card, learners are given considerable control of their learning: they may rewind, replay, and fast-forward as needed, and they may access a number of help features such as a dictionary, a list of characters, a story summary, script, and special notes.

THE PROJECT

The multimedia project at CUA has gone through two phases, each of which involved developing, implementing, and evaluating lessons.

Phase 1: Development and Implementation

In the initial pilot study in the spring of 1995, the language coordinator/media director team and a

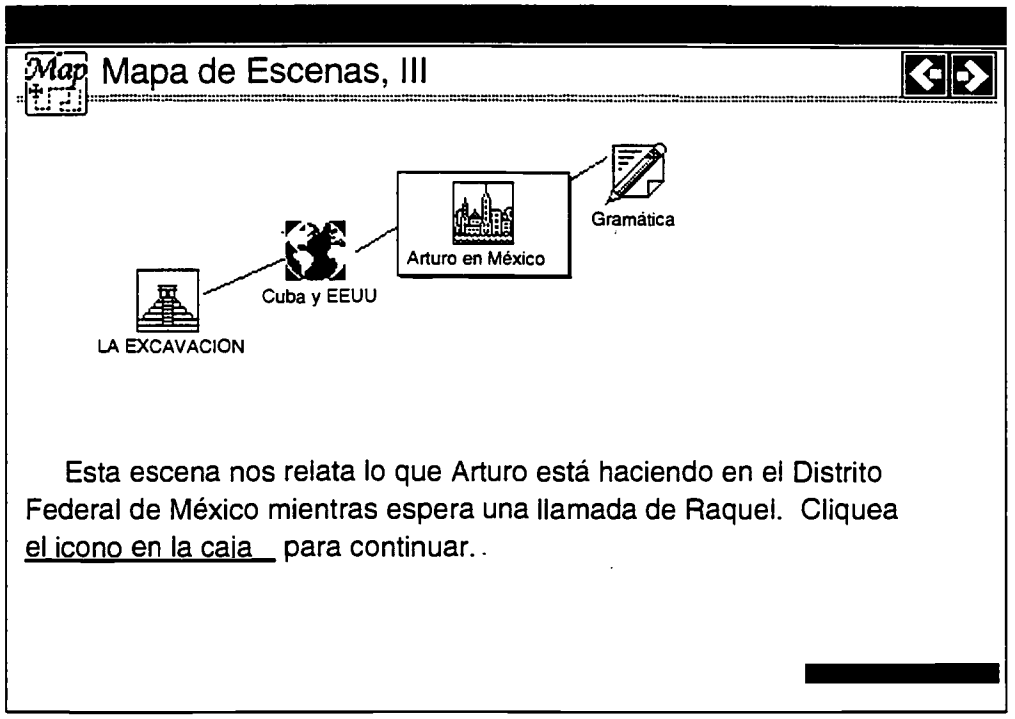


Figure 1. Advance Organizer

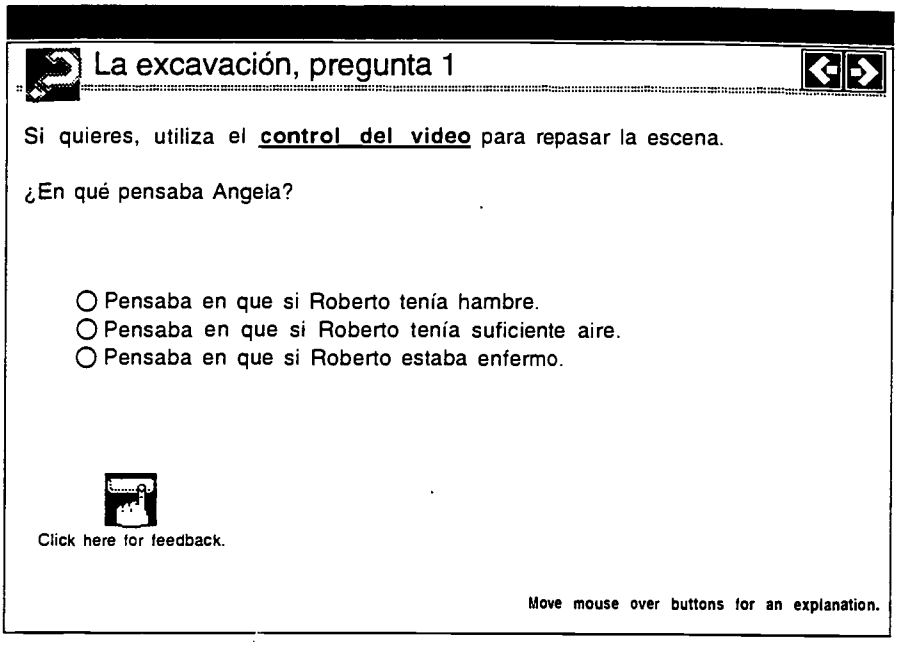


Figure 2. Multiple Choice Questions

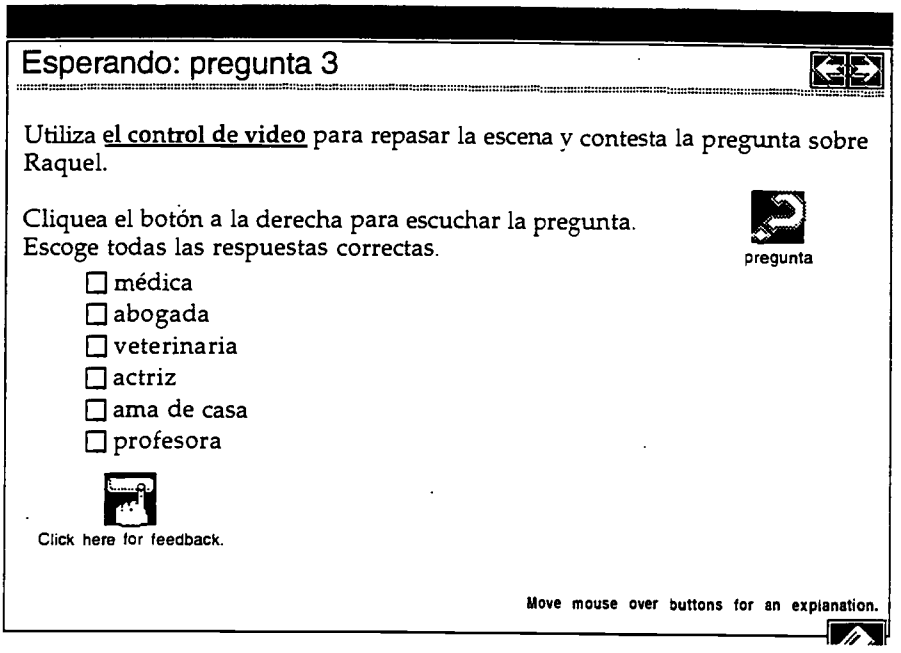


Figure 3. Checklist Questions

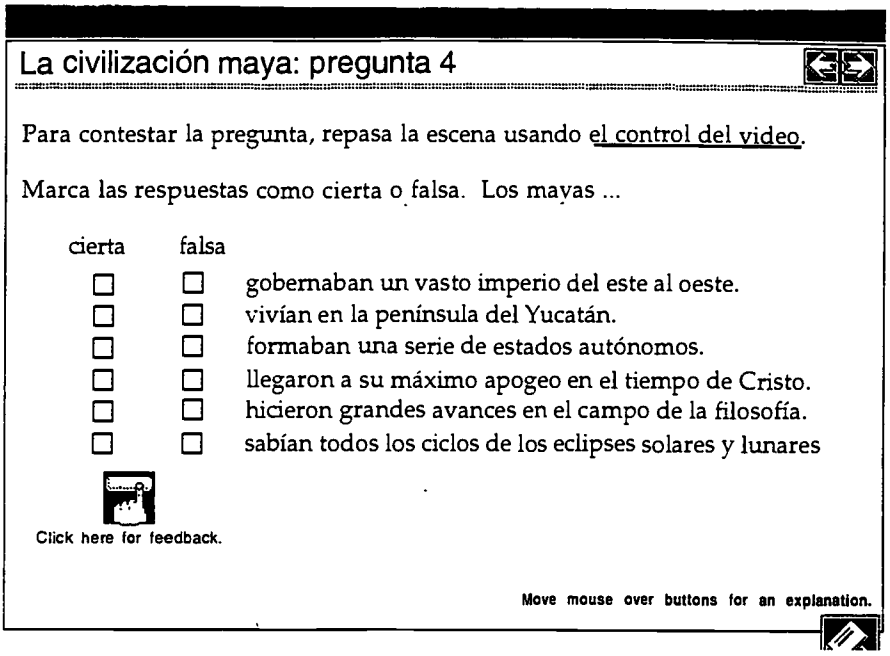


Figure 4. Binary Checklist Questions

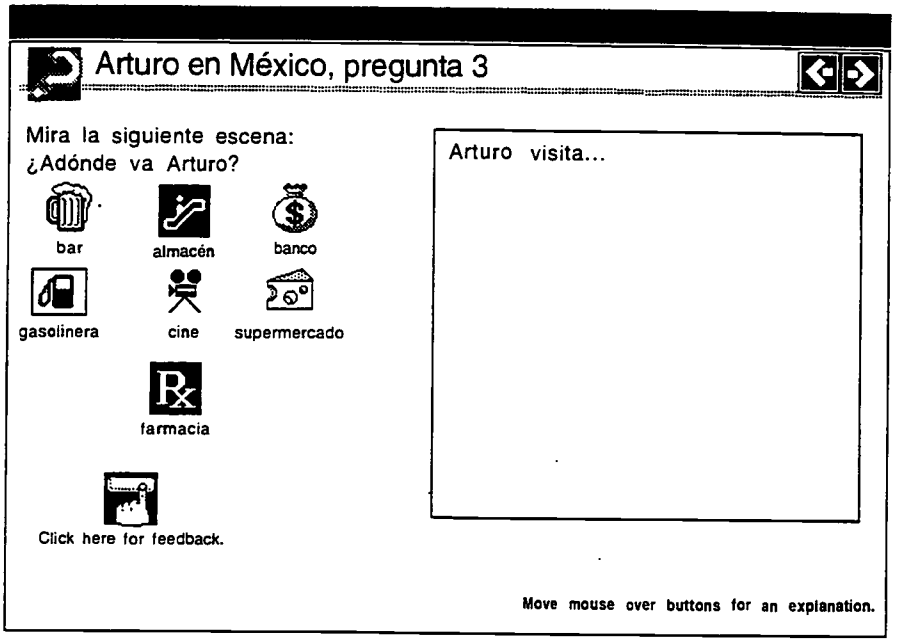


Figure 5. Icon Sorting Questions

graduate teaching assistant (TA) designed Libra lessons of approximately 20 cards each to accompany eight *Destinos* episodes used in second- and third-semester Spanish. The 25-minute episodes were segmented into scenes, and text maps led the learners to view each scene in its entirety without stopping. Subsequent cards asked questions on both main ideas and details in the scene and provided learners the opportunity to review relevant portions of the scene as needed. Due to time limitations, no glossary was provided, and scripts and character lists were not available for all lessons.

At the second-semester level, two instructors of four class sections with a total of 59 students participated in Phase 1. The third-semester group involved three instructors of six sections with 106 students. For each lesson, students had one week to go to the Multimedia Language Center individually or with a partner to complete the assignment. They either completed the Libra lesson linked with the *Destinos* videodisc in the experimental group, or they viewed the video and filled in a written worksheet in the control group. The questions in both conditions were substantially the same, though some of the written questions had to be adapted from electronic to print medium. The instructors were requested not to discuss the episode until the day it was to be completed.

Because the episodes assigned for out-of-class work were an integral part of the *Destinos* program, the project team felt that they would integrate smoothly into the classwork just as the other videos viewed in the lab did. To reinforce the notion that these assignments were indeed required, the syllabus listed their due dates, and the course description stated that they counted as 10% of the final course grade.

Phase 1: Evaluation

The evaluation of the pilot study was based on a time-series experimental design in which the students alternated between the experimental Libra condition and the control group. A modified recall protocol pretest served to establish preexisting differences among the students. Similar recall protocols in English, collected in class the day students were to have completed viewing the assigned episode, served as the posttest. Due primarily to an uncooperative instructor, the second-semester data were quite incomplete and yielded no statistically significant results. The third-semester data were reduced by student absences, but complete data sets were available on 17 students in the experimental group and 20 in the control group. An analysis of covariance revealed that on two of the four posttests, the experimental group significantly outperformed the control group. A third posttest approached significance (see FIPSE, 1994, for complete data and explanation). Thus, despite the limited data, the pilot study provided evidence of the positive impact that the Libra-based viewing of the *Destinos* videos had on comprehension.

Student questionnaires administered at the end of the semester

complemented the quantitative analysis. These questionnaires revealed a generally positive attitude toward the multimedia lessons (see Table 1). The areas of concern identified by the student users included the length of time required to complete the lessons and the lack of understanding of the program on the part of some of the instructors, students, and lab attendants. In its own evaluation of the project, the development team posited two additional factors that may have contributed to the lack of regular student participation: insufficient integration into classwork and the redundancy built into the *Destinos* program, which, appropriately for a soap opera, allows viewers to follow the story line even if they miss an episode.

Phase 2: Development and Implementation

The second phase of the project focused on addressing the concerns raised in the pilot and on improving participation. The Phase 1 authoring team and two additional TAs reconfigured four of the lessons for second-semester Spanish (see sample cards in Figures 2 to 5). With the continued support of the FIPSE project, one of the TAs had received a grant to attend a Libra workshop, and he then helped teach the authoring system to the other TA team member as they rewrote the lessons.

Two major changes were incorporated into the lesson revisions. First, given the redundant nature of the videos, the team reduced the

Table 1
Attitude Questionnaire Item from Phase 1 (n=80)

My attitude toward using computers to learn a foreign language:

	Before This Semester	After This Semester
Hostile	1	2
Hesitant	13	14
Indifferent	30	22
Curious	31	21
Enthusiastic	5	21

amount of video to be viewed in each lesson to three to four segments: one or two that presented the key story line events, one that focused on culture, and one that targeted functional language use. Second, several adjustments were made to the lessons to allow the learner greater flexibility in interacting with the text. A video control panel was added to every card so students could fast-forward, rewind, and replay as needed at any time. More help features, in the form of scripts, character lists, and a glossary, were provided in three of the four lessons. A third change involved the addition of video-related grammar exercises, but because this component was not included in the pilot study, it will not be examined in this article.

Significant changes were made in the implementation of the revised lessons. The two second-semester Spanish instructors were closely involved in the development of the lessons, and, as a result of their investment in the process, they understood Libra clearly and were able to explain it adequately to their students. The instructor team made a conscious effort to integrate the viewings more tightly with coursework. They agreed to hold detailed discussions of the targeted scenes in class, to give brief dictations and quizzes on them, and to incorporate them directly into exam questions. Care was also taken to train the lab assistants more thoroughly on the program.

During the spring semester of 1997, the 55 students enrolled in the four sections of second-semester Spanish taught by the two TAs who helped revise the Libra lessons participated in Phase 2 of the project. As requested by the students in the pilot study and recommended by Masters-Wicks et al. (1996), an orientation to the program was held in class before the first lesson was assigned. Using transparencies of sample computer screens, the instructors explained the question types and the general navigation features. As before, the lessons represented 10% of the final grade, and due dates were listed on the syllabus.

The final evaluation, which was qualitative in nature, was done via a user questionnaire (see the appendix). The questionnaire included Likert-scale items to assess student reactions and open-ended questions to gain more insight into the students' own perceptions of the lessons. Instead of waiting until the end of the semester to assess students' views, the students completed the Phase 2 questionnaire in the Multimedia Language Center following completion of the last lesson. It was hoped that, by answering the questions immediately after the lesson, learners would provide more accurate information because their experiences would be fresh in their minds.

Phase 2: Evaluation Results and Discussion

A total of 30 students out of the 55 enrolled completed the questionnaire. The instructors were surprised at this low response rate because more than 30 students reported doing the Libra lessons. Some students may have done all the lessons except the final one when the questionnaire was distributed. Perhaps the

lab attendants were not diligent enough in handing out and collecting the questionnaires as students completed the lesson. While the causes cannot be determined with certainty, the 55% response rate is evidence of the difficulty of investigating a phenomenon that is essentially an out-of-class, independent activity.

The four Likert-scale questions are presented in Table 2. The responses to Questions 1 and 2 suggest that this second phase more successfully provided the support learners needed to feel comfortable using the computer materials. The third and fourth questions again reflect a generally positive (though not enthusiastic) attitude toward computer-mediated learning: fewer students were negative at the end of the project than at the beginning.

The following five open-ended questions composed the second part of the survey:

1. What was the main benefit you derived from doing the Libra computer lessons?

2. What strategies did you use to maximize this benefit?

3. What did you like the most about doing the Libra computer lessons?

4. If you could improve one aspect of the Libra computer lesson, what would it be?

Table 2
Likert-Scale Questionnaire Items from Phase 2 (n=30)

	SA	A	N	D	SD
1. I had sufficient orientation to the *Libra* program before doing the lessons on my own.	10	9	3	6	1
2. The lab attendants are helpful to me when I use the multimedia program.	8	13	5	2	0
3. At the beginning of the semester, my attitude toward using computers to learn a foreign language was positive.	8	8	10	4	0
4. After the completing the *Libra* computer lessons, my attitude toward using computers to learn a foreign language is positive.	8	13	8	1	0

SA = Strongly Agree, A=Agree, N=No Opinion, D=Disagree,
SD=Strongly Disagree

5. In your opinion, what would be the ideal number of Libra computer lessons to do [per semester]? How much time would you be willing to spend on each?

While there was considerable variation in the wording of individual student responses, the answers to these questions were revealing, in particular when considering the patterns that emerged across student comments. Tables 3, 4, and 5 summarize the responses to the first three questions, including the category or pattern of response, the number of students who made comments with that focus, and some sample comments.

When questioned about their perceptions of the benefits of the Libra lessons (see Table 3 for summary), half of the students (14) expressed the view that the lessons helped them "hear" or "understand" the videos. This response is in line with the listening comprehension goals of the authoring system itself and our own programmatic objectives and provides encouraging evidence of Libra's face validity. Closely related to the comprehension focus was the observation of three students that the lessons emphasized the storyline. This awareness of main events appears to support the use of the textmap as a way to offer learners support in building the

Table 3
Questionnaire Responses, Phase 2: Student-Perceived Benefit of Multimedia Lessons (n =30)

Question 1. What was the main benefit you derived from doing the Libra computer lessons?

Category of Response	Number of Responses	Sample Comment
Listening comprehension	14	• To see and hear lesson at my speed • Understanding more of what was said by the characters
Focus on content	3	• Reinforced the storyline and made it more clear
Practice	3	• To watch a lesson and do exercises about it
Vocabulary	2	• I gained a larger vocabulary

macropropositional structure of the text. Additionally, three students identified practice as Libra's greatest benefit, an apparent recognition of being more actively involved than in traditional video viewing. Finally, two students specified that Libra was useful in improving their vocabulary. While no vocabulary assessment was done in this study, Chun and Plass (1996) found evidence of considerable vocabulary learning in their work with multimedia to enhance reading comprehension. Their work and student comments in this study raise questions about the possibility of incidental vocabulary learning in the Libra multimedia environment, questions that need to be addressed in future research.

The second question asked students to reflect on and identify the strategies they had used while doing the Libra lessons (see Table 4). Most frequently cited (by ten students) was the use of the rewind/replay button, with comments such as "I listened to segments more than once using review buttons." Students' willingness to interact with the computer in order to adjust for their own learning needs is an example of the positive side of learner control in multimedia.

A second strategy mentioned by three students was attending more closely to the material; as one student said, "[I] listened to it more than I pay attention to *Destinos*." Multimedia lessons written with systems such as Libra have the potential to facilitate attention in a number of ways: video is segmented and efficiently accessible, lessons are task-oriented, earphones reduce

Table 4
Questionnaire Responses, Phase 2: Strategy Use (n=30)

Question 2. What strategies did you use to maximize this benefit?

Category of Response	Number of Responses	Sample Comment
No response	12	
Replay	10	• Using the rewind button to review when confused or unsure
Attention	3	• Listened to it more than I pay attention to *Destinos*
Visuals	2	• Watching the videos themselves

sound distortion, and students physically interact with the computer via mouse clicks to progress through the lesson. Because paying attention is one of the metacognitive strategies considered essential for language learning (Oxford, 1990), this capacity may be one of multimedia's important strengths.

In a response that may be related to learning style preferences, two students claimed to attend to the videos themselves. Since the visual stimulus helps to contextualize the input, focusing on the images may be a particularly helpful strategy to visual learners. Though learning styles were not examined in this investigation, there is growing research interest in this area (Beauvois & Eledge, 1996; Chun & Plass, 1996; and Meunier, 1997). Other strategies mentioned (each by one student) included doing the lessons with a partner, reading the questions in advance, and consulting the feedback.

Perhaps most notable is what students did not say in response to this question. Twelve of the 28 students did not offer any strategies at all. Though the lack of comment does not necessarily mean that the students did not make use of any strategies, it does raise a fundamental question: Are students sufficiently aware of their learning needs in order to address them? In the traditional, teacher-mediated classroom, metacognitive strategies that enable learners to address their learning needs are essential, but they may be even more critical in the autonomous learning environment of multimedia.

As a case in point, the questionnaire results suggest that the students may not have taken full advantage of some of the help features. While 80 to 90% of the students claimed to use the replay function "sometimes" or "often," 25% of the students reported "never" using the script or glossary, 40% reported using them "sometimes," and the remainder did not respond. The uneven use of help functions may be in part due to issues such as the level of difficulty of the lesson questions, the screen design of these features, the lack of adequate training on their use, and forgetfulness on the part of the students. Though the version of Libra used in this project did not allow for tracking student actions [3], studies that examined computer logs for computer-assisted writing (Scott, 1990) and reading (Davis & Lyman-Hager, 1997) came to a similar conclusion: students did not make adequate use of the resources provided. Based on their findings with third-semester French students, Davis and Lyman-Hager suggest that "students need training to use computers in the most beneficial way" (p. 68).

What are the strategies that "good multimedia learners" use? Can or should these strategies be taught? If so, how? How is strategy use affected by proficiency level and motivation? Hoven (1997) proposes that not all second language learners, especially in the initial stages of their learning, want or are able to take control (105). She advocates a hierarchy of task types to raise students' awareness of strategies. As recommended by Pusack and Otto

(1997), "issues of control and interactivity in multimedia instruction urgently require more attention from researchers to inform development" (p. 10).

When asked what they liked best about the multimedia lessons, students responded quite pragmatically (see Table 5). Nine students indicated that the lessons helped them achieve their short-range goals ("It helped my understanding of *Destinos*") and their longer-range goals ("Improved my listening skills," "Very helpful on tests"). Five students highlighted the learner-centeredness of the lessons, for example, being able to "choose [their] own time [to do the lessons]," "partner up", and "start[ing] and stop[ping] when I need to."

The ability to check their work via the feedback button was cited as a plus by four students. In the literature, access to immediate, non-threatening feedback is widely seen as an advantage of computer-assisted learning (Bush, 1997; Pennington, 1996; Pusack & Otto, 1997). In the foreign language learning context, the research on error correction has been focused primarily on oral feedback, and results have been mixed (for a summary, see Chaudron, 1988). As yet, there is little research on the impact of computer-provided

Table 5
Questionnaire Responses, Phase 2: What Students Liked the Most About the Lessons (n =30)

Question 3. What did you like the most about doing the Libra computer lessons?

Category of Response	Number of Responses	Sample Comment
Goal-oriented	9	• Made the lesson more clear • Very helpful on tests • Improved my listening skills
Control	5	• Having the option of starting and stopping when I need to • Choose own time • Can partner up if you want
Feedback	4	• Checking your answers

feedback on language learners. How do learners react to computer feedback? Is it anxiety-producing? What types of feedback are most helpful?

In a study of feedback on computer-assisted grammar practice, Nagata and Swisher (1995) found evidence that 'smart' feedback, which gave further information about errors, was more useful to students than the generic variety. What feedback is most appropriate for comprehension-based tasks? The feedback provided in the Libra lessons was generic, offering such responses as the following: "All of the answers so far are correct. Keep going." "None of your answers are correct." "Some of your answers are correct and some are not." Despite the general nature of this feedback, 68% of the students claimed to use the feedback "sometimes" or "often," and 68% described it as "helpful" or "very helpful." There were, however, two requests for improved feedback. One student noted, "If I got it wrong, I never knew why." The lack of understanding of this potentially valuable component of the multimedia environment points clearly to a need for further research.

The final questions provided input for further cycles of development and implementation. For example, one student recommended showing the video on the same screen as the questions, a possibility with newer versions of the Libra program, but feasible only with the addition of a video server to store the lesson segments. The generally positive attitude of the students to the Libra lessons was reflected in their responses to the last question. Three students wanted to do one multimedia lesson for each of the 17 chapters studied during the semester. Others offered suggestions ranging from 3 to 6 lessons, with 5 as the median. Almost half of the students expressed a willingness to spend from 20 to 40 minutes per lesson. At the present time, the Libra lessons that were not updated in Phase 2 are currently being restructured to be 20 to 30 minutes in length.

CONCLUSION AND IMPLICATIONS

The effective use of multimedia presents challenges at many levels. For those who are interested in exploring this instructional tool, several recommendations follow from the CUA experience.

1. **Select authorware according to its theoretical foundation and your project goals.**

As the array of possibilities for authorware is growing rapidly, the selection process needs to be made with care. Authorware should be selected according to its compatibility with current language learning models and the project goals. Libra was created with comprehension theory in mind and our goals were listening comprehension. Other programs such as GALT and Annotext are designed to facilitate reading comprehension (for a summary of currently available authorware by language skill, see Martínez & Herren, 1998).

2. **Create broad support for your project.**

Due to the time, skill, and energy demands of working with multimedia, broad support is necessary.

The development team might include the language coordinator, another faculty member(s), the media director, graduate teaching assistant(s), and undergraduate assistants. Outside support from the software creator and grant project director is necessary for technical reasons and can also be valuable in building support for the project at higher levels in the university. Backing from the department chair and dean is important as release time and funds are necessary for project members to attend workshops and develop and rework lessons. This high-level support is also key when faculty submit multimedia projects as part of their activity reports and tenure and promotion reviews.

3. **Assess facilities and technical issues in advance.**

Because university facilities, academic computer support, and software requirements vary widely, it is not useful to outline specific technical guidelines. The basics of availability and compatibility of equipment and software need to be addressed, however, as do logistical considerations, such as the following: How many stations are available for how many students? Should students reserve time slots? Should they work in pairs? Are the lessons properly loaded and easily accessible on each station? Are lab assistants familiar enough with the lessons to help students as needed? Dealing with these issues will help the project run more smoothly.

4. **Familiarize instructors with the program and the project.**

Instructor preparation is crucial to the implementation process. Not only do instructors need to be familiar with how the program functions, but they also need an understanding of the rationale for the program in order to respond adequately to students' questions. It may be useful to acquaint new instructors with the history of the project and, if possible, to involve them in revising questions. In this manner, instructors will feel more a part of the lessons and will be able to offer useful suggestions for future improvements.

5. **Integrate the out-of-class work into class.**

Integrating the lessons into the coursework is a substantial challenge. Including the project on the course description and assigning it a percentage of the grade are only the beginning. Conducting student orientations helps make student users more aware of what they are expected to do and allows them to begin to establish a comfort level with the lessons. Beyond the need for these preliminary arrangements, our experience points to one key reality of the use of multimedia, especially for out-of-class work: learners interact variably with the program. Some students may do the lessons quickly with no difficulties, some may listen several times to the segments and consult numerous help features, others may begin but not finish, and others may not do the lessons at all.

Because the use of multimedia is a fragmentary experience that differs from one learner to another, spending time in class to recreate and unify the experience for the class as a whole is indispensible [4]. Thus, follow-up activities such as putting events in order, identifying characters and their roles, and summarizing, as well as providing traditional matching and dictation quizzes, need to be incorporated. Exams can also include questions that focus on the lesson segments.

6. Include evaluation.

Evaluation is an essential component of the process. The information it provides supports the subsequent development and implementation cycles and brings new questions into focus. Both phases of this project experienced difficulties with two overlapping concerns: participation and data gathering. Actual student participation in the lessons, which cannot be documented due to the lack of computer tracking, is not clearly reflected in the data collected. In addition to using the computer logs now available in Libra, enhancements of class integration techniques, improved lab attendant oversight, and increased access to videodisc stations to accommodate student schedules better may result in greater participation and better data in future phases. It may also be, however, that not all learners, particularly those at the first and second year levels, are motivated enough to do multimedia lessons outside of class. Investigating the use of multimedia with more advanced, more motivated students might result in increased participation and more complete data [5].

A fundamental issue revealed in this project is how little is actually known about what learners do with multimedia. The limited number of strategies reported and the number of students who listed no strategies suggest a lack of awareness of how to learn with multimedia. While much work has been done in identifying strategies in communication, research needs to identify those that are appropriate for computer-mediated learning. In addition to computer tracking of students' interactions with computers, researchers need to collect retrospective protocols to gain a better understanding of the strategies used. This approach can be incorporated in future cycles of evaluation through the random selection of a subset of participants to monitor more closely via appointments for the lessons and for follow-up interviews.

Other areas of further research suggested by this study include the effect of multimedia on vocabulary learning and the impact of learning styles and feedback on computer-mediated learning.

As the exploration of multimedia continues, projects such as the one undertaken at CUA will demonstrate the feasibility of creating lessons tailored to the context of an individual department. The process is not a simple one, however, as it involves ongoing cycles of development, implementation, and evaluation. The experience of one project provides important feedback to further this process and is also crucial in developing a knowledge

base in this rapidly expanding field. The generally positive outcomes of this project, both in terms of student learning and student attitude, serve to further motivate continued refinement and use of the lessons at CUA. The questions raised suggest areas for much-needed future research, in particular, the need to understand better how learners make use of resources afforded by the computer in order to manage their own learning. It is through continued cycles of this development-implementation-evaluation process that teachers and students will gain the ability to use the tools that technology provides more knowledgeably and effectively.

ACKNOWLEDGEMENTS

This project would have been impossible without the support and assistance of a number of people. Robert Fischer and Mike Farris, the developers of Libra, have been ever-willing to offer their guidance and insight. I would especially like to thank the rest of the team, media directors, Chris Higgins and Ed Dixon, and graduate students, Charmaine McMahon, Chris Schaefer, and John Crowley, who, throughout the project, have devoted time and energy while maintaining enthusiasm and good humor. Finally, I wish to express my appreciation to the Modern Language Department Chairs Mario Rojas and George Gingras for their vision of the role technology can play in language learning.

NOTES

1 For information on the Libra authoring system, contact Dr. Robert Fischer, Department of Modern Languages, Southwest Texas State University, San Marcos, Texas 78666, or see http://www.libra.swt.edu.
2 An IBM compatible version of Libra is currently being developed.
3 The most recent version of Libra, 1.2.9, has a tracking log.
4 Gilberte Furstenberg, author of the simulation program *A la rencontre de Philippe*, expressed this view in response to my question at the Tech Talk session at the Northeast Conference on the Teaching of Foreign Languages, April, 1997.
5 Judith Liskin-Gasparro offered this consideration at the University of Texas Spanish Second Language Acquisition Symposium, October 1997.

REFERENCES

Beauvois, M., & Eledge, J. (1996). Personality types and megabytes: Student attitudes toward computer-mediated communication in the language classroom. *CALICO Journal, 13* (2 & 3), 27-45.

Bush, M. (1997). Implementing technology for language learning. In M. D. Bush (Ed.), *Technology-enhanced language learning* (pp. 287-349). Lincolnwood, IL: National Textbook Co.

Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning*. Cambridge: Cambridge University Press.

Chun, D., & Plass, J. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal, 80* (2), 183-198.

Davis, J., & Lyman-Hagar, M. (1997). Computers and L2 reading: Student performance, student attitudes. *Foreign Language Annals, 30* (1), 58-72.

Fischer, R., & Farris, M. Libra [Computer software]. San Marcos, TX: Southwest Texas University.

Foelsche, O. et al. Annotext [Computer software]. Hanover, NH: Dartmouth College.

Fund for the Improvement of Postsecondary Education. (1994). *Annual report*. Washington, DC: FIPSE.

Hoven, D. (1997). Instructional design for multimedia: Towards a learner centered CELL model. In K. A. Murphy-Judy & R. Sanders (Eds.), *Nexus* [CALICO Monograph Vol. 4] (pp. 98-111). Durham, NC: Duke University Press.

Joiner, E. (1997). Teaching listening: How technology can help. In M. D. Bush (Ed.), *Technology-enhanced language learning* (pp. 77-120). Lincolnwood, IL: National Textbook Co.

Lyman-Hager, M., Stout, M., & Weinstock, M. (1995). *Glossing authentic language texts (GALT)* [Computer software]. State College, PA: Education Technology Services, The Pennsylvania State University.

Martínez, A., & Herren, D. (1998). Challenges and opportunities: Curriculum pressures in the technological present. In J. Harper, M. Lively, & M. Williams (Eds.), *The coming of age of the profession* (pp. 141-167). Boston: Heinle and Heinle ITP.

Masters-Wicks, K., Pistlewate, L., & Lewental, M. (1996). Developing interactive instructional software for language acquisition. *Foreign Language Annals, 29* (2), 217-222.

Meunier, L. (1997). Affective factors and cyberteaching: Implications for a postmodern pedagogy. In K. A. Murphy-Judy & R. Sanders (Eds.), *Nexus* [CALICO Monograph Vol. 4] (pp. 122-132). Durham: Duke University Press.

Nagata, N., & Swisher, M. (1995). A study of consciousness-raising by computer: The effect of metalinguistic feedback on second language learning. *Foreign Language Annals, 28* (3), 335-347.

Oxford, R. (1990). *Language learning strategies*. New York: Newbury House Publishers.

Pennington, M. (1996). The power of the computer in language education. In M. C. Pennington (Ed.), *The power of CALL* (pp. 1-14). Houston, TX: Athelstan.

Pusack, J., & Otto, S. (1997). Taking control of multimedia. In M. D. Bush (Ed.), *Technology-enhanced language learning* (pp. 1-46). Lincolnwood, IL: National Textbook Co.

Rubin, J. (1994). A review of second language listening comprehension research. *Modern Language Journal, 78* (2), 199-221.

Scott, V. (1990). Task-oriented creative writing with *Système-D*. *CALICO Journal, 7* (2), 58-67.

VanPatten, B., Marks, M., & Teschner, R. (1992). *Destinos: An Introduction to Spanish* [video]. Boston, MA: WGBH Annenberg/CPB Project.

APPENDIX—QUESTIONNAIRE SPANISH 102

This questionnaire has been developed to solicit your views about the *Destinos* Libra computer lessons you have just completed in order to better understand their effectiveness and how students can best make use of them. Your responses are needed and appreciated. Your name is requested here in order to give you **attendance credit** for Lección 33; your responses themselves will not affect your grade in any way.

Name _____ **Instructor:** _____

YOUR EXPERIENCE USING THE LIBRA COMPUTER PROGRAM

Fill in the information as appropriate:

1. How much time did you spend doing the Libra lessons?

Lesson #	Time spent					
	Did not do this lesson	less than 15 min.	15-30 min.	30-45 min.	45-60 min.	more than 1 hr.
Lec. 22						
Lec. 30						
Lec. 33						

2. In doing the computer lessons, which of the following features did you use? How helpful were they?

Features	never used	some-times	often used	not helpful	helpful	very helpful	no opinion
Directions and sample questions in English							
Text map with synopsis							
On the video palette:							
play							
fast forward							
rewind							
pause/stop							
Script (not available in Lesson 30)							
Glossary (not available in Lesson 30)							
Feedback button							
Grammar exercises							

Please comment on your experiences using any of the above features:

3. How useful were the computer lessons in helping you develop skills in the following areas? In the last column, number them 1 to 5 to rank the areas according to which ones were most improved for you by doing the

computer lessons, with 1 being the **least** improvement up to **5** for the **greatest** improvement.

Skills	not useful	somewhat useful	very useful	no opinion	RANK 1-5
listening comprehension					
vocabulary learning					
cultural understanding					
situational language use					
grammar usage					

DO YOU AGREE OR DISAGREE?

For the following statements, indicate your view by circling **5** for **Strongly Agree**, **4** for **Agree**, **3** for **Neutral**, **2** for **Disagree**, and **1** for **Strongly Disagree.** Please clarify your response in the far right hand margin as needed.

Statement	SA	A	N	D	SD	Comments
1. I am generally comfortable working on computers.	5	4	3	2	1	
2. I had sufficient orientation to the Libra program before doing the Libra lessons on my own.	5	4	3	2	1	
3. I cannot follow the *Destinos* storyline just by watching the segments in the Libra lesson.	5	4	3	2	1	
4. The computer lessons are sufficiently reinforced in class activities.	5	4	3	2	1	
5. I often feel frustrated while doing the lesson.	5	4	3	2	1	
6. At the beginning of this semester, my attitude toward using computers to learn a foreign language was positive.	5	4	3	2	1	
7. After completing the Libra computer lessons, my attitude about using computers to learn a foreign language is positive.	5	4	3	2	1	
8. The directions and questions in Spanish are clear so I know what I need to do to answer.	5	4	3	2	1	
9. It is difficult to "navigate" from one part of the lesson to another.	5	4	3	2	1	
10. It is easy to quit the Libra lesson and open the grammar.	5	4	3	2	1	
11. Doing the computer lessons helps me do better on the tests.	5	4	3	2	1	
12. The lab attendants (work study students) are helpful to	5	4	3	2	1	

me when I use the Libra computer program.

13. It is easier to do dictations in class/on exams after completing the Libra lesson. 5 4 3 2 1

14. I would like to do more Libra computer lessons. 5 4 3 2 1

YOUR IDEAS, PLEASE

1. What was the main benefit you derived from doing the Libra computer lessons?

   What strategies did you use to maximize this benefit?

2. For Lesson 30, you saw the complete video in lab and you did the Libra lesson in the MLC. How would you compare these two experiences? Explain the advantages and disadvantages of each.

3. What did you like the most about doing the Libra computer lessons?

4. If you could improve one aspect of the Libra computer lesson, what would it be?

5. There are approximately 17 *Destinos* episodes per semester. In your opinion, what would be the ideal number of Libra computer lessons to do? How much time would you be willing to spend on each? Please explain.

Editorial Policy

TPFLE is devoted to the promotion of pedagogical and theoretical issues in the fields of Foreign/Second Language Education, Teaching of English as a Foreign/Second Language, and Applied Linguistics related to language learning and teaching. Manuscripts submitted for publication undergo blind evaluation by two referees selected from members of an editorial advisory board. The editor has the right to make editorial changes in any manuscript accepted for publication.

Submission Information

To be considered for publication, a manuscript should be typewritten according to APA style and format (see *TESOL Quarterly*) and be no more than 30 double-spaced pages in length (including references and appendices). Leave 1 1/2-inch top and left margins and 1-inch margins elsewhere. Include an abstract of not more than 200 words under the title of the manuscript. A separate page should contain the title of the manuscript and the author's name, address, e-mail, and phone number. The author's name should not appear elsewhere in the text. Refer to your previous publications in the third person. Submit three copies of the manuscript to —

The Editor of TPFLE
Sanchez Building 528
The University of Texas at Austin
Austin, TX 78712

Besides providing the three hard copies, the author(s) should submit the article on a Macintosh diskette (Microsoft Word 5.1 or later). Please set the document in 12-point type.

Subscription Information

TPFLE is published biannually in the Spring and Fall. The annual membership to TPFLE is $10.00, inclusive of postage, and an individual issue of this journal is available at $5.00. Make your checks payable to "The University of Texas at Austin."

FL 025681

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

REPRODUCTION BASIS

☑ This document is covered by a signed "Reproduction Release (Blanket) form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").