University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Faculty Publications: Department of Teaching,
Learning and Teacher Education

Department of Teaching, Learning and Teacher
Education

2014

Spoken Spanish Language Development at the High School Level: A Mixed-Methods Study

Aleidine Kramer Moeller

University of Nebraska–Lincoln, amoeller2@unl.edu

Janine Theiler

Lincoln (Nebraska) Public Schools

Follow this and additional works at: <http://digitalcommons.unl.edu/teachlearnfacpub>

Part of the [Bilingual, Multilingual, and Multicultural Education Commons](#), [Curriculum and Instruction Commons](#), [Junior High, Intermediate, Middle School Education and Teaching Commons](#), and the [Secondary Education and Teaching Commons](#)

Moeller, Aleidine Kramer and Theiler, Janine, "Spoken Spanish Language Development at the High School Level: A Mixed-Methods Study" (2014). *Faculty Publications: Department of Teaching, Learning and Teacher Education*. 157.
<http://digitalcommons.unl.edu/teachlearnfacpub/157>

This Article is brought to you for free and open access by the Department of Teaching, Learning and Teacher Education at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications: Department of Teaching, Learning and Teacher Education by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Spoken Spanish Language Development at the High School Level: A Mixed-Methods Study

Aleidine J. Moeller¹ and Janine Theiler²

1. Edith S. Greer Professor of Foreign Language Education,
University of Nebraska-Lincoln

2. Spanish teacher, Lincoln Public Schools

Abstract

Communicative approaches to teaching language have emphasized the centrality of oral proficiency in the language acquisition process, but research investigating oral proficiency has been surprisingly limited, yielding an incomplete understanding of spoken language development. This study investigated the development of spoken language at the high school level over five consecutive years, involving more than 1,500 students representing 23 school districts. Quantitative Standards-Based Measure of Proficiency speaking scores and student-produced qualitative spoken samples ($n > 6,000$ samples) contributed to an understanding of the development of spoken language. Hierarchical linear modeling (HLM) revealed a consistent growth trajectory of spoken language development, and results indicated that 18–30% of the variance in student outcomes may be attributed to the teacher variable.

Keywords: classroom-based research, longitudinal study, mixed methods, oral language development

The rise of communicative language learning has led to widespread acceptance of communicative competence as a primary goal of language education and, as such, central to good classroom practice (Savignon, 1997). This approach to language instruction emphasizes the ability to communicate in a second language in real-life situations both inside and beyond the classroom. Instead of measuring language learning in terms of seat time, test scores, or number of credit hours, communicative skills are demonstrated through task-based communicative activities. As a result of this emphasis on oral communication, proficiency has emerged as central to communicative language learning and teaching. However, there is a lack of research at the classroom level that reveals what students are able to do with oral language after one, two, three, and four years of language

study. There is a paucity of research relating specifically to the development of spoken language at the secondary level (Tschirner & Heilenman, 1998). Although several studies have offered a glimpse of classroom-based proficiency ratings for the high school language learner (for examples, see Glisan & Foltz, 1998; Huebner & Jensen, 1992; Moeller & Reschke, 1993; Steinmeyer, 1984), the data have been strictly quantitative and have been conducted within educational systems with no consideration of related and potentially confounding factors, such as the teacher, during the exploration of oral language production.

Due to such limitations, as well as substantial differences in results, particularly at the beginning levels of language learning in secondary classrooms, this study explored students' progress toward proficiency over a period of 5 years using a combination of qualitative methods, including thematic coding and organization, to reveal overarching trends in oral spoken language, and quantitative methods, including the Standards-Based Measurement of Proficiency (STAMP) test, a teacher-independent, computer-mediated measure of oral language proficiency.¹ Purposefully integrating mixed methods offers "a very powerful mix" (Miles & Huberman, 1994, p. 42) that develops "a complex" picture of oral language development (Greene & Caracelli, 1997, p. 7). In choosing this integrative data design, the researchers' purpose was one of complementarity, a design element used to measure overlapping, but distinct, facets of a phenomenon under investigation (Caracelli & Greene, 1993). Results from one method—in this case, qualitative data—were used to enhance, illustrate, or clarify results from the other method—in this case, quantitative data (Greene & McClintock, 1985).

Quantitative research questions for this study investigated the growth trajectory of spoken Spanish over four consecutive years of high school Spanish learning. Quantitative questions also delved into the variance in spoken production scores that was attributable to

teacher differences or individual student differences. In addition, a qualitative analysis of students' speech samples was also carried out using a rubric that was developed to quantify particular aspects of the raw speech samples and thus create more detailed learner profiles and illustrate a range of language production. These data helped clarify and build upon the quantitative findings in order to establish a depth of understanding of spoken language production at specific intervals during language learning—specifically, at the end of years one, two, three, and four.

Literature Review

This overview of previous research addresses stated performance expectations, as well as contributions and limitations of existing studies investigating oral language production and proficiency. A brief overview of the value added of mixed methodology is also addressed.

Oral Language Production and Proficiency

What can students truly achieve with consecutive years of language study? According to the performance guidelines issued by ACTFL (1998), after four consecutive years of second language study, teachers should expect students to perform at the Intermediate Low level of language proficiency. While ACTFL provides this fairly clear-cut expectation for oral proficiency, the research literature at both the secondary and postsecondary levels presents a broader range of expectations for oral proficiency development.

Much of the research literature concerning oral proficiency has used ACTFL's Oral Proficiency Interview (OPI) to determine what students can achieve with consecutive years of language study. At the college level, the most recognized studies explored the first through fourth year of language study in German, French, and Russian. Magnan (1986), Dugan

(1988), Kaplan (1984), and Freed (1987) explored the development of French oral proficiency, indicating levels ranging from Novice High to Intermediate Mid after 1 year of study and Intermediate Mid to Advanced levels after 4 years of study. Thompson (1996) reported that one year of postsecondary Russian learning yielded a Novice Mid level of proficiency, and four consecutive years of study produced Intermediate High to Advanced levels of proficiency. Tschirner (1992, 1993) worked with German language learners and determined that 2 years of college-level German language study typically yielded Intermediate Mid levels of oral proficiency. More recently, Tschirner and Heilenman (1998) conducted a study of postsecondary learners after four semesters of German instruction, with 25% of the students scoring Novice High, 45% Intermediate Low, and 30% Intermediate Mid (p. 153). In addition, Tschirner and Heilenman detected no correlation between students' OPI outcome and the length of instruction, nor was there a correlation between the OPI score and student background variables. The range of proficiency outcomes among these studies and the lack of correlation between proficiency outcome and seat time may be surprising in light of the commonly held expectation that four consecutive semesters of college instruction, presumed to approximate 4 years of instruction at the high school level, should yield Intermediate Low levels of proficiency (ACTFL, 1998). Tschirner and Heilenman concurred: "Reaching the Intermediate Low level is a greater achievement than previously thought" (p. 154).

At the high school level, the most recognized studies explored primarily the second through fifth years of language study. Steinmeyer (1984) examined German OPIs ($n = 25$) administered after 2, 3, 4, and 5 years of instruction and discovered OPI ratings ranging from Novice Mid (second year) to Advanced (fifth year). Huebner and Jensen (1992) administered OPIs to French, German, and Spanish language learners after 2, 3, 4, and 5 years in

a secondary school. They found that beginners (second year) averaged Novice Mid to Novice High proficiency, and more advanced learners (fifth year in French and Spanish) averaged Advanced level of language proficiency. Moeller and Reschke (1993) presented one of the few studies addressing the first year of language learning, reporting the results of OPIs after 1 and 2 years of junior high German instruction. They found that first- and second-year German learners who averaged 240 minutes of instruction per week both achieved a mean proficiency level of Novice High.

Glisan and Foltz (1998) conducted a statistical analysis of language competency of second- and fourth-year high school learners of Spanish. Fifty students from two schools participated, and OPI ratings for second-year students ranged from Novice Mid to Novice High. The mean oral proficiency rating for fourth-year students approached Intermediate Low, but 30% of the participants in fourth-year Spanish did not attain Intermediate Low levels, again emphasizing that attaining Intermediate Low after four consecutive years of study may be more difficult than previously thought (Glisan & Foltz, 1998, p. 9). Norris and Pfeiffer (2003) examined results of 100 SOPI (Simulated Oral Proficiency Interview) tests administered to all levels of instruction in a university German department and concluded that the "recommended proficiency standards may underestimate the potential and actual achievement of German language learners" (p. 572).

When analyzing this range of proficiency outcomes for students with similar temporal sequences of instructional experiences, it is important to recognize what Magnan (1986) referred to as bands of proficiency that overlap from one year/level to the next. These ranges of proficiency within levels of language learning and the year-spanning bands of proficiency remind educators of the individualized nature according to which students proceed along the continuum of language proficiency.

The Need for More Than Solely Oral Proficiency Scores

The aforementioned studies offered only proficiency scores as a means of exploring oral language production, and these overarching proficiency scores are indeed valuable in understanding the general growth of the learner; however, such scores are not sufficient in and of themselves to improve instruction and student learning. Kunnan and Jang (2011) and Long (2011) argued for a forward-looking approach in testing that incorporates diagnostic feedback into achievement and proficiency testing. Alderson, Clapham, and Wall (1995) pointed out that such diagnostic feedback should “identify those areas in which a student needs further help ... whether a student needs particular help with one of the four major language skills” (p. 12). Thus while proficiency scores are helpful, what one may learn from and subsequently accomplish with a proficiency score is quite limited. A holistic proficiency assessment as well as a diagnostic skills-based assessment are both needed in order to better understand the development of language production. Both a trajectory of language growth as well as a more detailed review of skills acquired over time can provide valuable feedback for the teacher and student to improve learning and instruction. Unlike previous studies, this longitudinal study allowed for tracking of language skills as well as development of oral language production.

The Need for Mixed-Methods Research Design

Mixed-methods studies involve the integration and mixing of two strands of quantitative and qualitative data “for the purposes of breadth and depth of understanding and corroboration” (Johnson, Dnwiegbuzie, & Turner, 2007, p. 123) the sum of which is greater than either approach alone. Creswell and Plano Clark (2011) defined a mixed-methods convergent design as a specific mixed-methods model that enables the researcher to “compare and contrast

quantitative statistical results with qualitative findings or to validate or expand quantitative results with qualitative data” (p. 62).

In this convergent data transformation mixed-methods design, researchers converted qualitative data into quantitative data, engaged in statistical data analyses, and returned to the original data set in order to facilitate increased understanding (see Bachman, Lynch, & Mason, 1995, for examples from the language assessment field; also see Caracelli & Greene, 1993; Creswell & Plano Clark, 2011; Sandelowski, Voils, & Knafl, 2009). A convergent mixed-methods design with data transformation (Creswell & Plano-Clark, 2011) allowed the researchers to quantitatively show the trajectory of language learning over time, while the qualitative data—that is, student-produced oral language—provided language samples over time.

Thus, based on the review of prior related research, it is clear that (1) existent oral language development studies have been largely limited to short-term quantitative studies conducted primarily at the postsecondary level with small samples, (2) there is a need for more detailed descriptions of language development beyond numerical proficiency scores, and (3) a mixed-methods approach provides a deeper and more comprehensive understanding of the development of spoken language for the beginning high school language learner.

The specific research questions were:

1. What is the four-year growth trajectory of students’ spoken Spanish skills?
2. What percentage of the variance in student speaking scores can be attributed to the respective Spanish teachers and to differences among the individual students?
3. What kinds of language can students produce at different stages and levels of language study?
4. How do the qualitative data from spoken samples enhance or provide a more complete picture of the quantitative results of spoken language development?

Methods

Study Design

Researchers adopted a convergent data transformation mixed-methods design for this study. The raw data consisted of (1) quantitative holistic oral production scores assigned by an external independent rater, and (2) qualitative speaking samples produced by students during the administration of an online teacher-independent test (STAMP). Hierarchical linear modeling (HLM) of quantitative data produced an overarching holistic depiction of spoken language development. Concurrently, analysis of student speech samples led to the development of a qualitative rubric, which then allowed for further quantification of samples (see Appendix A). HLM analysis of the resulting quantified detail-specific data addressed the underlying nature of oral language development. Finally, in an effort to concretely represent statistical findings in this study, researchers capitalized on the clarity afforded through authentic exemplars and descriptions situated in the original qualitative data set and rubric. This final step resulted in the production of learner profiles, consisting of qualitative raw speech samples and qualitative rubric descriptors interwoven with quantitative holistic results from the STAMP test and quantitative detail-specific results (from the rubric). The overarching procedures might be depicted as: {[QUAN + (QUAL → QUAN)] + qual}. Figure 1 is a visual representation of the analysis process and also depicts the progression utilized to concretely and comprehensively present the findings.

Participants

A purposive sample of teachers was recruited for the study, allowing researchers to follow the same students and teachers for several years in an attempt to reduce the impact of external variables often associated with conventional cohort studies. Between 2005 and 2010, researchers recruited 21 teacher-participants and their 1,544

individual students. These 21 teachers represented 23 urban and rural school districts in the state of Nebraska, including 19 public and four parochial schools of varying sizes.

In each year of the study, the sample grew in diversity and size as new language teachers were recruited and new students were added to the original first-year cohort. In addition, a number of participants were lost to attrition as teachers or students left the program or moved out of their school district. As a result of longitudinal tiered recruitment and attrition, the final dataset included all students at all levels of Spanish, some of whom were followed for multiple consecutive years and some of whom participated for only a portion of their Spanish learning experience. Data on the total set of participants are presented in Table 1.

Measures

Data were gathered using STAMP. This online adaptive assessment tool served as the source of both quantitative and qualitative data for this study. STAMP, a statistically validated, realia-based, and textbook-independent assessment, produces a comprehensive score for proficiency in reading, writing, and speaking (Avant Assessment, 2008). A factor integral in choosing the STAMP test for this study was its online archival capability that allowed researchers to access the approximately 6,000 speaking samples produced by student-participants during the assessment (see Profiles in Appendices B and C for examples of qualitative samples). Appendix B reveals the language progression of the same student over 3 years of language study. The speaking prompts as they appeared in the STAMP test are provided as well as a qualitative analysis of the student samples based on the rubric in Appendix A. Appendix C provides a snapshot of student variation in speech production among students enrolled in second-, third-, and fourth-year Spanish. This table makes transparent the wide deviations among student speech products enrolled in the same level of language study.

Figure 1. Research Procedures for Convergent Transformation Mixed-Methods Study. (Based upon design and mapping principles as addressed in Creswell and Plano Clark, 2011).

Table 1. Overview of Study Sample

	2005	2006	2007	2008	2009	2010
Total number of student-participants	106	394	527	484	331	295
1st-year students	80	270	186	49	173	166
2nd-year students	13	123	307	261	71	86
3rd-year students	13	1	34	158	37	17
4th-year students	NA	NA	NA	16	51	26

There were five dependent variables in the study: holistic speaking score, pronunciation, grammar, vocabulary, and fluency. The holistic score for speaking production was assigned at the end of each academic year as students engaged in the STAMP test and received a score from external, trained raters (Avant Assessment, 2008). The remaining indexes (pronunciation, grammar, vocabulary, fluency) were

assigned during the transformation of the qualitative data.

Qualitative Data Analysis and Transformation (QUAL→QUAN)

In this study, a research team consisting of 10 members—six Spanish language experts and four language researchers—developed the

rubric that was based solely on student oral produced samples and that guided the transformation of STAMP speaking samples into a series of detail-specific quantitative proficiency-related scores. This speaking rubric was an integral tool for this study, as it allowed for an in-depth analysis of specific attributes of student speaking (pronunciation, grammar, vocabulary, and fluency). Development of the speaking rubric involved three steps of qualitative analysis of the student speaking samples as outlined below.

Step 1: Identify Emergent Themes

The six Spanish language experts conducted a qualitative analysis of 30 speech samples that were representative of all four levels of Spanish instruction. Each reviewer independently sorted all 30 samples into overarching categories (exceeds expectations, meets expectations, or in progress) of speaking quality. This process was conducted without knowledge of the STAMP proficiency score to avoid a bias in sorting. The research team met to discuss the factors that each individual considered when sorting the samples, and a pattern of shared attributes began to emerge. Researchers thus determined a need for a common coding system to represent the thematic attributes specific to student-produced speaking samples.

Step 2: Refine Emergent Themes

The members of the research team subsequently independently analyzed 30 additional student samples, with each team member producing codes to describe specific attributes of student speaking. The team again assembled, discussed the independently produced descriptors, combined similar descriptors, and eliminated redundancies. Members refined the emergent themes into four specific thematic categories that were common across speaking quality levels but that differed in degree and complexity: pronunciation, grammar, vocabulary, and fluency.

Step 3: Define Refined Themes

Finally, the members of the research team

worked to establish agreed-upon descriptive terminology within each of the four categories (pronunciation, grammar, vocabulary, fluency) for each level of quality (exceeds expectations, meets expectations, and in progress). This was done in the same manner as the previous two phases of rubric development, each of the six language experts working independently followed by sharing and discussion.

The final version of the rubric can be found in Appendix A. This rubric was subsequently applied to each of the 6,000 qualitative samples that were archived during the STAMP testing process. Students receiving a score of 3 in each category were described as exceeding expectations, a score of 2 indicated that students met the expectations, and a score of 1 indicated that students were progressing toward meeting the expectations. For example, a student scoring 3 in all categories (pronunciation, grammar, vocabulary, fluency) might be described as one who could create speaking samples that demonstrated logically developed ideas; correct word order; and speaking style appropriate to the task, text type, and speaking venue. Such a student could maintain control of sentence structure and show appropriate use of definite and indefinite articles, pronouns, verbs, number and gender agreement, possessive adjectives, and prepositions. Minimal errors might occur in the student's speaking samples, but these errors would not interfere with comprehensibility. The student would demonstrate creative use of vocabulary, which is beyond basic requirements, and appropriately use idiomatic expressions. The overall response of a student who exceeded expectations in foreign language speaking was creative and comprehensible, revealed minimal evidence of interference from the first language, made use of appropriate rejoinders, and provided information beyond the basic requirements. A narrative depicting a student meeting expectations or in progress might be developed through similar application of descriptors found in the speaking rubric.

Before scoring any samples, the six Spanish language experts worked cooperatively to

identify exemplars that might help guide the rating process. They next independently rated 30 additional samples to test for interrater reliability, which was established at 0.83. The language experts subsequently scored all 6,000 student samples according to the speaking rubric. This rubric-based scoring procedure transformed the qualitative speaking samples into quantitative data specific to the four skill areas addressed in the rubric.

Quantitative Data Analysis

The type of design underlying this study generally involved the concurrent but separate analysis of data related to the same phenomenon, with the results being merged during the interpretation phase of research (Creswell & Plano Clark, 2011). With the qualitative data transformed into quantitative scores (QUAL → QUAN), the next step for researchers consisted of an in-depth statistical analysis of all available quantitative data (QUAN+QUAL=[QUAL → QUAN]). HLM and descriptive analyses were conducted with all quantitative data.

In this study, researchers desired to make repeated measures representing Spanish learner growth while accounting for the nested learning structure; thus HLM was adopted, as it captures measurement occasions within a nested structure. In this study, these measurement occasions (lower-level or level 1) were nested within students (higher-level or level 2). These students (level 2) were then nested within teachers (the highest level, or level 3). Proc Glimmix in Statistical Analysis Software (SAS 9.2) facilitated the HLM analyses. For each dependent variable, researchers established the best-fit model through a series of unconditional and conditional models. For all models, restricted maximum likelihood was used for any missing data under the assumption of missing at random and with Satterthwaite approximation for degrees of freedom.

Researchers employed HLM to (1) determine growth in speaking production across 4 years of Spanish language learning at the high

school level, (2) predict change for each of the speaking variables, and (3) account for variance in student scores as a function of the classroom teachers and individual students. This longitudinal collection of data from students gave rise to a three-level HLM to describe the change in speaking for those enrolled in Spanish. This model can be depicted as:

$$(\text{STAMP Speaking})_{ijk} = (\alpha_{000} + U_{0ij} + V_{0i}) + (\beta_{000}) * (\text{Level Spanish}) + e_{ijk}$$

In this equation, α_{000} represents the fixed intercept, U_{0ij} represents the intercept difference or deviation for a specific teacher, V_{0i} represents the intercept difference or deviation for a specific student, β_{000} represents the slope, and e_{ijk} represents the error. In this model, “Level Spanish” represents time, spanning from 1 (representing the first year of Spanish study) to 4 (representing the fourth year of Spanish study). The slope describes the change in speaking score performance over time (levels of Spanish).

To build a model in HLM, researchers began with a basic, or empty, model, which aimed to reveal variance in the absence of specific predictors. In this case, the empty model focused on spoken language development independent of time as a predictive variable. A three-level empty model (random intercept only) was fitted for each dependent variable (STAMP speaking, pronunciation, grammar, vocabulary, and fluency). The three levels in this model represented the teacher (level 3), the student (level 2), and measurement occasion (level 1).

Results

The descriptive statistical results, one based on the STAMP ratings and the second based on the rubric, are depicted in Tables 2 and 3. The difference in sample size between Tables 2 and 3 is due to a small sample of students whose STAMP data were deemed not ratable by the STAMP raters; however, in spite of the poor recording quality, raters on the research team were still able to analyze these data.

Table 2. Mean and Standard Deviation of STAMP Holistic Speaking Scores

	<i>n</i>	$\bar{x}$	<i>sd</i>
1st-year Spanish	827	1.63	0.76
2nd-year Spanish	756	1.94	0.80
3rd-year Spanish	250	2.91	0.72
4th-year Spanish	93	3.25	0.56

As shown in Table 2, the STAMP holistic speaking score mean increased with each year of instruction, with the third year of study representing the largest gain in speaking proficiency. These consistent growth trends were not completely consistent when looking at the underlying details of student speaking represented in the speaking rubric variables of pronunciation, grammar, vocabulary, and fluency, reported in Table 3. While there remained a consistent increase in mean scores from the first through third year of study for all variables (0.37–0.55 increase), there was a decrease in all of the detail-specific variables in the fourth year of study (0.07–0.14 decrease).

Longitudinal collection of data from students gave rise to a three-level HLM to describe the change in speaking for those enrolled in Spanish. Table 4 reveals the variance in student speaking outcomes that was attributable to students and to teachers. Residual within-teacher

variance in Table 4 points to classroom teachers being accountable for 17.9% of the variance for STAMP holistic speaking, 22.5% of the variance for pronunciation scores, 24.1% of the variance for grammar scores, 29.4% of the variance for vocabulary scores, and 26.7% of the variance for fluency scores. Residual within-student variance also indicated that individual students were accountable for variance in scores, and the variance attributable to students was similar to the variance attributable to teachers. Residual in-student variance was calculated by referring to Table 4 and subtracting residual in-teacher from residual in-student in-teacher. Per this calculation, differences among individual students accounted for 16.4% of the STAMP speaking score variance, 21.6% of the variance in pronunciation scores, 24.2% of the variance for grammar scores, 18.8% of the variance in vocabulary scores, and 26.5% of the variance in fluency scores.

Table 3. Mean and Standard Deviation of Speaking Attribute Scores

	<i>n</i>	Pronunciation		Grammar		Vocabulary		Fluency	
		$\bar{x}$	<i>sd</i>	$\bar{x}$	<i>sd</i>	$\bar{x}$	<i>sd</i>	$\bar{x}$	<i>sd</i>
1st-year Spanish	830	1.68	0.48	1.58	0.52	1.54	0.50	1.48	0.51
2nd-year Spanish	756	1.82	0.45	1.77	0.52	1.72	0.52	1.71	0.52
3rd-year Spanish	253	2.05	0.44	2.07	0.50	2.02	0.50	2.03	0.48
4th-year Spanish	93	1.91	0.39	2.00	0.44	1.91	0.42	1.94	0.49

Table 4. Random Effects for Empty Models and Intraclass Correlation

		STAMP speaking	Pronunciation	Grammar	Vocabulary	Fluency
Random intercept variance at student level (level 2)		0.145	0.053	0.076	0.057	0.083
Random intercept variance at teacher level (level 3)		0.158	0.055	0.076	0.089	0.083
Residual variance (level 1)		0.580	0.137	0.163	0.157	0.146
ICC	Residual in-student	0.343	0.441	0.483	0.482	0.532
	in-teacher					
	Residual in-teacher	0.179	0.225	0.241	0.294	0.267
	Student in-teacher	0.522	0.509	0.50	0.610	0.500

Interestingly, including the random slope at the student level did not improve the model fit for STAMP speaking variables. Thus, the baseline model for student speaking variables did not include random slope. Table 5 represents the baseline model for the four speaking variables of interest in this study. The conditional model revealed the trajectory for student growth in speaking production over four years of language study. According to the conditional model, students were predicted to produce a STAMP holistic speaking proficiency score of 1.43 after the first year of study, and this score would increase by 0.58 with each additional year of study. When considering pronunciation, students would score 1.68 after one year of study. The conditional model suggested that students would score 1.57 (meets expectations) on the grammatical accuracy measure and would continue to improve by 0.13 1

points with each subsequent year of study. The vocabulary score would be 1.51 following

one year of study, and this score would increase by 0.13 with each additional year of study. For fluency, one year of study was predicted to yield a score of 1.51, and additional years would increase this by 0.13.

It is noteworthy to merge the descriptive and HLM results and consider the similarities and differences. Figure 2 presents composite graphs to indicate how the HLM model fit the data. These graphs indicate that both the aggregate mean results and the HLM results exhibited a similar pattern. Of note, however, is the deviation in this pattern of growth during the fourth year of study according to descriptive statistics.

While the quantitative analysis and findings alone produced an in-depth look into the development of student speaking proficiency, researchers elected to return to the qualitative data to provide an enhanced representation of the quantitative results. This step represented an effort to provide a concrete representation

Table 5. Parameter Estimates and Model Fit Statistics for Final Conditional Models

Parameters	STAMP Speaking			Pronunciation			Grammar			Vocabulary			Fluency		
	Estimate	SE	p value	Estimate	SE	p value	Estimate	SE	p value	Estimate	SE	p value	Estimate	SE	p value
Fixed effects															
Intercept	1.43	0.08	< 0.0001	1.68	0.07	< 0.0001	1.57	0.06	< 0.0001	1.51	0.06	< 0.0001	1.51	0.06	< 0.0001
Level of Spanish	0.58	0.02	< 0.0001	0.07	0.01	< 0.0001	0.13	0.01	< 0.0001	0.13	0.01	< 0.0001	0.14	0.01	< 0.0001
Random effects															
Residual variance	0.32	0.02		0.14	0.01		0.16	0.01		0.15	0.01		0.15	0.01	
Intercept variance BWT	0.12	0.04		0.05	0.02		0.06	0.02		0.07	0.03		0.07	0.02	
Intercept variance BWS	0.20	0.02		0.05	0.01		0.07	0.01		0.05	0.01		0.07	0.01	
Model fit															
AIC	4203.85			2272.55			2637.00			2451.93			2518.88		
BIC	4206.98			2275.68			2640.14			2455.06			2522.02		
Number of parameters	7			7			7			7			7		

SE = standard error

BWT = between teachers

BWS = between students

AIC = Akaike information criterion

BIC = Bayesian information criterion

Figure 2. HLM and Descriptive Composite Charts

of the quantitatively indicated growth of spoken language development for the high school learner. This merging of results, samples, and insights was represented by profile summaries of language learning after 1, 2, 3, and 4 years of study. These profiles provide the reader with a detailed, rich, and real-world understanding of second language speaking development. The profile for first-year learners of Spanish is provided in Table 6, while the profiles for second-,

third-, and fourth-year learners appear in Appendix B.

The speaking sample in Table 6 provides a concrete example of the oral skills and abilities of a student who received the mean score for first-year Spanish development at the high school level. Table 7 illustrates the level of skill for two outliers and makes transparent the variation in students' first-year oral production. These outliers revealed what was possible in

Table 6. First-Year Spanish Speaking Profile

Example Prompt	Select one of the pictured rooms and describe at least three common or typical activities that occur in this room on a daily basis. Be as detailed as possible and speak using complete sentences.
Authentic Student Product	<i>Mi madre y yo comemos ... con mi hermano, mi hermana, mi tía, mi tío y ... mi padre. Mi madre cocina en la cocina con naranjas, pollo, bistec y cereales... . La canter es blanco con negro canters. La glase es blanco.</i>
Pronunciation $\bar{x} = 1.68$	Exemplar score = 2
Qualitative Rubric Descriptors Largely incomprehensible Anglicized accent, intonation, and word stress. Began to sound accurate. Could use appropriate word stress. Could imitate target language sounds. English sounds were occasionally used.	Analysis of Student Sample An L1 interference accent was evident but still comprehensible. The pronunciation was beginning to sound accurate, and some target language Anglicized accent, intonation, sounds were imitated. The intonation, word stress, and accent were anglicized and demonstrated L1 interference.
Grammar $\bar{x} = 1.58$	Exemplar score = 1
Qualitative Rubric Descriptors Misuse of verbs Number and gender agreement misapplied Definite and indefinite articles absent or confused Lack of subject/verb agreement Word order was anglicized and message was obscured Errors interfered with meaning	Qualitative Analysis of Student Sample The grammar structure demonstrated mostly appropriate use of verbs. There was no evidence of gender or number agreement or correct definite or indefinite article usage. The evidence of incorrect definite article usage and gender agreement occurred with the L1 interference of vocabulary. For example, " <i>La canter es blanco...</i> " and " <i>La glase es blanco.</i> " Based on the context of the response, the word <i>canter</i> refers to the English word "counter," and the word <i>glase</i> refers to the English word "glass." Regardless of these vocabulary errors, the definite articles were incorrect as well as the gender agreement with the word <i>blanco</i> based on the articles given. Evidence of number disagreement and adjective placement was present in <i>negro canters</i> . The L1 interference can explain both the vocabulary error as well as the adjective placement error. Basic subject/verb agreement and appropriate pronoun use was evidenced. The word order at times was anglicized, but the message was communicated. The errors made did not interfere with overall meaning in the response.

Vocabulary $\bar{x}=1.54$

Exemplar score=1

Qualitative Rubric Descriptors

Limited/common/basic
vocabulary
Predominant single-word
utterances
Less than minimum word.
requirements
Much repetition of select
words

Qualitative Analysis of Student sample

The vocabulary was limited, and there was repetition of select words.
The minimum word requirements were not met as only two common
daily activities that occur in the room were mentioned. The limited
scope of vocabulary and L1 interference caused the sample to be
disconnected and needed to be decoded.

Fluency $\bar{x}=1.48$

Exemplar score=1

Qualitative Rubric Descriptors

Required information lacking
Slow, hesitant speech
Excessive and long pauses
Inappropriate responses to
questions and prompts
L1 interfered with
comprehensibility Gave
required information but
no more

Qualitative Analysis of Student sample

Only the required information was given. The speech was slow,
hesitant at times, indicating the internal thought process. There
were frequent and longer pauses and hesitations. The sample
responded appropriately to the prompt with the exception of the
missing daily activity. There was some L1 interference in word order
and anglicized grammar, but this did not interfere with the overall
meaning.

Table 7. First-Year Spanish Student Sample Variation

Category	$\bar{x}$	Sample Score	
Pronunciation	1.68	1	<i>Me gusta es shrimp porque es deliciosoooo. Me gusta ... mmm ... me gusta fidero fijas as es delicioso ... me gusta tacos delicioso.</i>
Grammar	1.58	2	
Vocabulary	1.54	1	
Fluency	1.48	1	
Category	$\bar{x}$	Sample Score	
Pronunciation	1.68	2	<i>Mi hermano tiene diecinueve años. Mi hermano es eh, mas o menos alto. Um ... nosotros um ... amamos nadamos ... nadar. Nosotros amamos nadar. Mi hermano es muy inteligente y tímido. Mi madre es amable.</i>
Grammar	1.58	3	
Vocabulary	1.54	3	<i>Mi madre es baja ... muy baja. Um ...um. ...</i>
Fluency	1.48	2	

comparison to the mean student profile. Second-, third-, and fourth-year examples are found in Appendix C.

Discussion and Implications

This study explored the development of oral Spanish language production based on 6,000 archived speech samples over 4 years of high school language instruction. The findings provided a holistic profile of student growth, descriptive details underlying that growth, the identification of factors related to that growth, student profiles representing annual mean growth outcome, and student profiles representing growth variation within years of Spanish language learning.

HLM revealed growth expectations across four consecutive years of language learning. According to the most recently published ACTFL Performance Descriptors for Language Learners (2012), teachers can expect students to produce language at the Intermediate Low

level after four consecutive years of high school language learning. While STAMP test results are related to, rather than equivalent to, the ACTFL Proficiency Guidelines (Avant Assessment, 2008), it is nevertheless interesting to note findings in light of ACTFL expectations for proficiency outcomes. According to the HLM analysis for this study, 4 years of language learning resulted in a STAMP score of 3.17, a score that relates to the ACTFL proficiency level of Novice High (3=Novice High). This contrasts with ACTFL's stated expectation that learners reach Intermediate Low after 4 years of study. HLM analysis indicated an outcome of 2.01 after two consecutive years of study, which relates to the ACTFL's Novice Mid level of proficiency (2=Novice Mid). Again, this is slightly lower than what is stated by ACTFL's (1998) expressed Novice High expectation for the student with two consecutive years of language learning. Based on these findings, it is recommended that additional research be directed at exploration into proficiency outcomes at the high school level.

The data further revealed a large increase in mean spoken language production scores between the second and third year of language study. Given the frequency with which 2 years of high school study of a second language are required for college entrance, this statistically indicated growth may be attributed to a generalized change in the language learner population. That is, the jump in mean scores may indicate that students who chose to discontinue their study after 2 years are the same students who struggled most with learning a language. If this is the case, early identification of struggling students, analysis of specific areas of difficulty, and the provision of extra support via defensible methodological principles as outlined by Long (2011, p. 387) may encourage these students to continue their language studies.

The oral language development profiles in this study depict a mean expectation for student oral production throughout the high school language learning experience. Although these outcome profiles provide valuable information about students' general patterns of progress toward proficiency and also offer exemplars to increase awareness of variability among language learners, there is a potential danger in overstressing mean levels of student oral language production when setting expectations for student outcomes. This is supported in the literature on interlanguage that confirms the nonlinear and unevenly paced increases and decreases in student language development as each learner constructs his or her own language system (Selinker, 1972). When the most widely recognized assertions regarding oral production expectations are based on mean findings, it is easy to overlook the widespread potential for much greater growth with oral language production as seen in Table 7, depicting the variation among individual students in the sample. This is further underscored by the individual student variability depicted by HLM analysis in Table 4. The variability may serve as motivation to strive for

more than the "mean expectation"; clearly, there is potential for students to attain the upper levels of each range for spoken language production. To maximize spoken language development, teacher-practitioners, teacher-educators, and researchers must establish a better understanding of both what is possible as well as how to attain such possibilities.

The analysis of the quantitative holistic STAMP scores afforded a more general overview of students' spoken language production, while the more detailed analysis of the archived speech samples using the qualitative rubric yielded a profile of specific component skills and abilities. Taken together, the combination of both holistic and detail-specific data yielded a more comprehensive depiction of oral language production than would have been revealed with either source of data in isolation. Kunnan and Jang (2011) and Long (2011) stressed the importance of incorporating diagnostic feedback into achievement and proficiency testing. This study may serve as an example of how results from an oral language production assessment (STAMP) may be analyzed in multiple manners in order to yield both a holistic understanding of students' speech and information that can be used to diagnose strengths and weaknesses so as to provide additional support for students who do not yet meet expectations and challenge students who exceed them in an effort to maximize opportunities for growth in oral language production for all learners.

ACTFL's recent research priority initiative addressing high-leverage teaching practices attested to the critical role of the teacher in the language learning process (Glisan & Donato, 2012). According to the HLM results of the current study, teachers accounted for 18 to 30% of the variance in students' spoken language production scores. Although data on teachers' use of high-leverage teaching practices were not collected as part of the current large-scale study, Cohen, Weaver, and Li (1996) noted that direct teaching of strategies before, during, and after each speaking task resulted in improved

performance. They concluded, "If instructors systematically introduce and reinforce strategies that can help students speak the target language more effectively, their students may well improve their performance on language tasks" (p. 29).

Similarly, Rossiter, Derwing, Manimtim, and Thomson (2010) emphasized that teachers need to be aware of the types of activities in which students participate and the ways in which such activities support the development of oral proficiency. Their analysis of the student and teacher editions of a number of textbooks indicated that very few included a full range of activities that were designed to enhance oral fluency. Furthermore, they noted that the most frequent activity, free production, would have alone been insufficient to help students develop the skills they need to produce unscripted speech addressing a range of topics in real-life settings. Other activities that may support students' progress toward higher levels of proficiency, such as rehearsal and repetition, consciousness-raising, and use of discourse markers, were underrepresented in the set of textbooks that were analyzed. The authors stressed that the teacher must know which activities are most beneficial in helping students to develop oral fluency and how to use those activities in order to enhance students' progress toward more native-like use of language.

In addition, the integration of technology into instruction can play a critical role in students' development of oral communication skills. Payne and Whitney (2002) conducted a quasi-experimental study in which they tested whether synchronous chatting in the second language might indirectly improve students' oral proficiency because it promotes the development of the same mechanism that underlies spontaneous conversational speech. Fifty-eight students from a third-semester Spanish course were split into groups who either participated in the computer-mediated intervention or received more conventional instruction. Students in the experimental group showed greater

gains in skills in both writing and speaking than those in the control group. Similarly, Lee (2014), also in this issue, addressed the role of technology in providing students with personalized and meaningful opportunities to improve their oral skills.

Thus while prior research has investigated a wide range of factors (e.g., instructional approaches, learner motivation, high-leverage learning strategies) that contribute to the development of student proficiency in the language classroom (e.g., Cohen et al., 1996; Long, 2011; Payne & Whitney, 2002), the impact of these factors on instructional effectiveness is not clear and does not necessarily demonstrate whether these factors are key to producing the optimal result. Instead of focusing on one single factor and relating it to student success in isolation of other possible factors, it may behoove researchers to consider approaching the situation from a different angle. There is scant research available that identifies the most successful students and teachers and *then* reveals the factors that are responsible for these most salient, positive results. In the quest for optimal spoken language development for all learners, it would be advisable to study those teachers who seem to have identified a set of teaching strategies that seem to be most beneficial.

The mixed-methods approach to research design for this study allowed the researchers to capitalize on the strengths of both quantitative and qualitative data sources and analyses in understanding a single phenomenon (Creswell & Plano Clark, 2011). Data convergence generated a breadth and depth of understanding that would not have been possible had researchers been limited to only quantitative or qualitative data. Qualitative data yielded an enhanced understanding of the development of oral language production throughout the high school learning experience, whereas the quantitative data presented a clear and concrete trajectory of oral language development. Findings based on both qualitative and quantitative data comprehensively depicted the trajectory of growth in oral language production as well as the nature

of that growth, and student speech samples provided concrete representations of quantitative findings.

Although the strengths of the approach adopted in this study were many, there were also limitations. In a study that uses true convergent mixed-methods design with data transformation, one would ideally work with two separate sets of data; this study derived both qualitative and quantitative data from the same data set. The HLM statistical approach may also be considered a limitation in this study. HLM took full advantage of available data, as it did not force the omission of data for those who chose not to continue with the study of Spanish. Missing data, however, were still an issue. HLM estimated coefficients for students for whom there were missing data, but there was concern about estimates based on only one or two data points (Raudenbush & Bryk, 2002). In addition, as noted in this article, these data represented a nested structure. While this is a strength of HLM, it should also be noted that one might further nest the data both conceptually and statistically. As such, it is possible that some of the teacher-attributable variance, as indicated in this study, may rather be school-attributable variance. Further studies are needed that delve into differences at expanded levels of nesting.

Conclusions

If language educators are to strive for communicative competence for all language learners, it is critical that we continue to develop a clear, deep, and accurate understanding of their oral language development and then link that development to student, teacher, language, school, community, and other variables. This study revealed consistent growth in oral language development over four years of classroom language learning, highlighted the variability in growth attributable to individual students and teachers, and provided profiles of mean language production as well as variants at the end of each academic year. These profiles

served as exemplars of what students could do with oral language.

This study illustrated the added value of mixed-methods research designs for the exploration of language development and, in general, for the field of second language acquisition. A melding of qualitative and quantitative data as illustrated in this study can enrich an understanding of the language learning process and be of value in identifying the variables that contribute to language development in the classroom. Amassing the cumulative results of studies such as these will bring educators ever closer to realizing the expressed goal of widespread communicative competence for second language learners.

Notes

1. The STAMP literature states that STAMP levels are "related to" (<http://www.avantassessment.com/stamp>) and "defined by" (http://www.bellarmine.edu/docs/default-source/foreign_languages_docs/STAMP_StudentGuide_Apr08.aspx) the proficiency levels and sublevels that are described in the ACTFL Proficiency Guidelines (e.g., Novice High, Intermediate Mid). Some STAMP results are reported using those designations; however, they are not equivalent to official ACTFL ratings.

References

- ACTFL. (1998). *ACTFL guidelines for K-12 learners*. ERIC Document Access Number ED 426593.
- ACTFL. (2012). *ACTFL guidelines for K-12 learners: 2012 Edition*. Retrieved March 12, 2014, from <http://www.actfl.org>
- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language testing construction and evaluation*. Cambridge, UK: Cambridge University Press.
- Bachman, L. F., Lynch, B. K., & Mason, M. (1995). Investigating variability in tasks and rater judgments in a performance test of Foreign Language Annals _ VOL. 47, NO. 2 227 foreign language speaking. *Language Testing*, 12, 238-257.

- Avant Assessment. (2008). *STAMP reporting guide*. Retrieved March 12, 2014, from [http:// avantassessment.com](http://avantassessment.com)
- Caracelli, V. J., & Greene, J. C. (1993). Data analysis strategies for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 15, 195–207.
- Cohen, A. D., Weaver, S. J., & Li, T.-Y. (1996). The impact of strategies-based instruction on speaking a foreign language. In A. D. Cohen (Ed.), *Strategies in learning and using a second language* (pp. 107–156). London: Longman.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed-methods research*. Thousand Oaks, CA: Sage.
- Dugan, J. S. (1988). *Standardized tests as an alternative to the oral interview*. Paper presented at the Seventh Annual Eastern Michigan University Conference on Languages for Business and the Professions. ERIC Document Reproduction Service No. ED 304 923.
- Freed, B. (1987). Preliminary impressions of the effects of a proficiency-based language requirement. *Foreign Language Annals*, 20, 139–146.
- Glisan, E. W., & Donato, R. (2012). Guest editors' message. *Foreign Language Annals*, 45, 3–7.
- Glisan, E. W., & Foltz, D. A. (1998). Assessing students' oral proficiency in an outcome-based curriculum: Student performance and teacher intuitions. *Modern Language Journal*, 82, 1–18.
- Greene, J. C., & Caracelli, V. J. (1997). Defining and describing the paradigm issue in mixed-method evaluation. *New Directions for Evaluation*, 74, 5–17.
- Greene, J. C., & McClintock, C. (1985). Triangulation in evaluation: Design and analysis issues. *Evaluation Review*, 9, 523–545.
- Huebner, T., & Jensen, A. (1992). A study of foreign language proficiency-based testing in secondary schools. *Foreign Language Annals*, 25, 105–115.
- Johnson, R. B., Dnwguegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed-Methods Research*, 1, 112–133.
- Kaplan, I. (1984). Oral proficiency testing and the language curriculum: Two experiments in curricular design for conversation courses. *Foreign Language Annals*, 15, 491–498.
- Kunnan, A. J., & Jang, E. E. (2011). Diagnostic feedback in language assessment. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 610–627). West Sussex, UK: Wiley-Blackwell.
- Lee, L. (2014). Digital news stories: Building language learners' content knowledge and speaking skills. *Foreign Language Annals*, 47, doi: 10.1111/flan.12084.
- Long, M. H. (2011). Methodological principles for language teaching. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 373–394). West Sussex, UK: Wiley-Blackwell.
- Magnan, S. S. (1986). Assessing speaking proficiency in the undergraduate curriculum: Data from French. *Foreign Language Annals*, 19, 429–438.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.
- Moeller, A., & Reschke, C. (1993). A second look at grading and classroom performance: Report of a research study. *Modern Language Journal*, 77, 163–169.
- Norris, J. M., & Pfeiffer, P. C. (2003). Exploring the uses and usefulness of ACTFL oral proficiency ratings and standards in college foreign language departments. *Foreign Language Annals*, 36, 572–581.
- Payne, P. J., & Whitney, P. J. (2002). Developing L2 oral proficiency through synchronous CMC: Output, working memory, and interlanguage development. *CALICO Journal*, 20, 7–32.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage.
- Rossiter, M. J., Derwing, T. M., Manimtim, L. G., & Thomson, R. I. (2010). Oral fluency: The neglected component in the communicative language classroom. *Canadian Modern Language Review*, 66, 583–606.

- Sandelowski, M., Voils, C. J., & Knafl, G. (2009). On quantitizing. *Journal of Mixed- Methods Research*, 3, 208-222.
- Savignon, S. J. (1997). *Communicative competence: Theory and classroom practice* (2nd ed.). New York: McGraw-Hill.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10, 219- 231. 228 SUMMER 2014
- Steinmeyer, G. (1984). Oral proficiency interviews at Amherst Regional High School: Report of a pilot project. *Unterrichtspraxis*, 17, 330-334.
- Thompson, I. (1996). Assessing foreign skills: Data from Russian. *Modern Language Journal*, 80, 47-65.
- Tschirner, E. (1992). Oral proficiency baselines for first- and second-year college German. *Unterrichtspraxis*, 25, 10-14.
- Tschirner, E. (1993). Developing grammars: Acquired grammatical systems of beginning and intermediate students. Paper presented at the annual meeting of the American Association of Teachers of German, San Antonio, TX.
- Tschirner, E., & Heilenman, L. K. (1998). Reasonable expectations: Oral proficiency goals for intermediate-level students of German. *Modern Language Journal*, 82, 147-158.

Appendix A

Rubric Used to Evaluate Speech Samples

	Exceeds Expectations (3 points)	Meets Expectations (2 points)	Does Not Meet Expectations (1 point)
Pronunciation	Appropriate intonation Appropriate word stress	Comprehensible; some accent evident Beginning to sound accurate	Largely incomprehensible Anglicized accent, intonation, and word stress
Vocabulary	Imitated target language sounds with ease English sounds were rarely used	Could use appropriate word stress Could imitate target language sounds English sounds were occasionally used	
	Beyond common/basic vocabulary Creative use of vocabulary Idiomatic expressions	Common/basic vocabulary Attempted to use vocabulary creatively Attempted to use idiomatic expressions	Limited vocabulary Predominant single-word utterances Less than minimum word requirements
	Speaker exceeded minimum word requirements	Speaker met minimum word requirements Some repetition of select words	Much repetition of select words
	Appropriate use of verbs Number and gender agreement	Mostly appropriate use of verbs Some number and gender agreement	Misuse of verbs Number and gender agreement misapplied
Grammar Structure	Appropriate use of definite and indefinite articles Appropriate subject/verb agreement Correct word order	Could use appropriate definite and indefinite articles Basic subject/verb agreement Word order was anglicized but message was communicated	Definite and indefinite articles absent or confused Lack of subject/verb agreement Word order was anglicized and message was obscured

	Exceeds Expectations (3 points)	Meets Expectations (2 points)	Does Not Meet Expectations (1 point)
Fluency/Response	Few errors	Some errors; errors did not interfere with meaning	Errors interfered with meaning
	Errors did not interfere with meaning	Evidence of pronoun use	
	Pronouns used appropriately	Could self-correct	
	Self-corrected		
	Natural speed/ Appropriate pacing Appropriate pause and hesitations Appropriate use of rejoinders Responded appropriately to questions and prompts	Gave required information but no more Slower speech evident Frequent and longer pauses and hesitations Responded appropriately to most questions and prompts Evidence of L1 interference	Required information lacking Slow, hesitant speech Excessive and long pauses Inappropriate responses to questions and prompts L1 interfered with comprehensibility

Appendix B

Speaking Profiles

Second-Year Spanish Speaking Profile

Example Prompt

Your pen pal from Veracruz, Mexico, is coming to visit you for a few months. She called to ask what kind of clothing to pack. Do your best to help her decide by describing:

- the weather
- what kind of clothes are popular
- what she might need for different occasions (school, sports, parties)

Tell her you are happy she's coming.

Authentic Student Product

Hace sol y hace frio. Necesita las camisetas, las pandrillas, las zapatillas de tenis. Tambien necesita las pandrillas de cuarto para jugar los deportes. Estoy muy contenta.

Pronunciation

$\bar{x} = 1.82$

Exemplar Score = 2

Qualitative Rubric Description

Comprehensible; some accent evident
 Began to sound accurate
 Could use appropriate word stress
 Could imitate target language sounds
 English sounds were occasionally used

Qualitative Analysis of Student Sample

In sample 2, an accent was evident but still comprehensible. The pronunciation was beginning to sound accurate, and some target language sounds were imitated. However, the word *pandrillas* was mispronounced several times in the response. This indicated the interference by L1 and required meaning to be decoded. The appropriate word stress was applied to most words. English sounds were occasionally used.

Grammar

$\bar{x} = 1.77$

Exemplar Score = 2

Qualitative Rubric Description

Mostly appropriate use of verbs
 Some number and gender agreement
 Could use appropriate definite and indefinite articles
 Basic subject/verb agreement
 Word order was anglicized but message was communicated

Qualitative Analysis of Student Sample

The grammar structure demonstrated appropriate use of verbs. There was some evidence of gender and number agreement. There was correct definite article usage. There was evidence of basic subject/verb agreement. There were no grammar errors that required the meaning to be decoded. The word order was anglicized, but the overall message was comprehensible.

Some errors; errors did not
interfere with meaning
Evidence of pronoun use
Could self-correct

Vocabulary

$\bar{x} = 1.72$

Exemplar Score = 1

Qualitative Rubric Description

Common/basic vocabulary
Attempted to use vocabulary
creatively
Attempted to use idiomatic
expressions
Speaker met minimum word
requirements
Some repetition of select words

Qualitative Analysis of Student Sample

The vocabulary was common, but the sample did demonstrate the creative use of verb forms. The minimum word requirements were met in the response. The limited scope of vocabulary and repetitive use of vocabulary, especially mispronounced vocabulary, caused the sample to be disconnected and required the meaning to be decoded.

Fluency

$\bar{x} = 1.71$

Exemplar Score = 2

Qualitative Rubric Description

Gave required information but
no more
Slower speech evident
Frequent and longer pauses and
hesitations
Responded appropriately to
most questions and prompts
Evidence of L1 interference

Qualitative Analysis of Student Sample

Only the required information was given. The speech was slow, hesitant at times, indicating some L1 interference. There were frequent and longer pauses and hesitations. The sample responded appropriately to questions and the prompt with some L1 interference in word order.

Third-Year Spanish Speaking Profile

Example Prompt

You are living with a host family in Spain. You've called to let your host family know that you won't be home for dinner tonight because you are going out with some friends. No one is home, so leave a message on the answering machine. Make sure you include the following in your message: Greet your host family and let them know that you will not be able to be home for dinner tonight; apologize for missing dinner; explain in detail why you are missing dinner tonight (where you are going, who you are with, and what you will be doing there); conclude the message by letting them know what time you plan to return and any other details that might be important for them to know.

Authentic Student Product

Hola todos. Lo siento pero no voy a comer la cena con ustedes. Voy a comer con juanes y mi amiga Sofia y voy a comer el restaurante nuevo en la ciudad la comida de pañal. Voy a [pause] vamos a comer y luego vamos a ir al parque y luego voy a regresar a casa. Pienso que regresar a las cinco y um...es todo lo siento que no voy a comer con ustedes. Es muy triste. Adios.

Pronunciation

$\bar{x} = 2.05$

Exemplar Score = 2

Qualitative Rubric Description

Comprehensible; some accent evident
 Began to sound accurate
 Could use appropriate word stress
 Could imitate target language sounds
 English sounds were occasionally used
 Appropriate intonation
 Appropriate word stress
 Imitated target language sounds with ease
 English sounds were rarely used

Qualitative Analysis of Student Sample

In sample 3, some accent was evident but the response was still comprehensible. The pronunciation began to sound accurate and could imitate target language sounds. Appropriate word stress was used. English sounds were rarely used.

Grammar $\bar{x} = 2.07$

Exemplar Score = 3

Qualitative Rubric Description

Mostly appropriate use of verbs
 Some number and gender agreement
 Could use appropriate definite and indefinite articles
 Basic subject/verb agreement
 Word order was anglicized but message was communicated
 Some errors; errors did not interfere with meaning
 Evidence of pronoun use
 Could self-correct

Qualitative Analysis of Student Sample

The grammar structure demonstrated appropriate use of verbs. There was evidence of gender and number agreement. There was correct definite article usage. There was evidence of self-correction as the speaker paused to change "*voy a*" to "*vamos a*" to represent the first-person plural, indicating that he and his friends would be eating together. There was the absence of the preposition *en* in the phrase "*comer el restaurante nuevo*." The word order was sometimes anglicized, but the overall message was communicated. The grammar errors made did not interfere with meaning.

Vocabulary $\bar{x} = 2.02$

Exemplar Score = 3

Qualitative Rubric Description

Common/basic vocabulary
 Attempted to use vocabulary creatively
 Attempted to use idiomatic expressions
 Speaker met minimum word requirements
 Some repetition of select words

Qualitative Analysis of Student Sample

The vocabulary was common but the sample did demonstrate some attempt at the creative use of vocabulary. There was one example, "*la comida de pañal*," which did not fit the context, and the word *pañal* was incorrectly used. The context referred to a new restaurant, and the phrase "*la comida de pañal*" means "baby food." The minimum word requirements were met and exceeded. There was some repetitive use of select words.

Fluency $\bar{x} = 2.03$

Exemplar Score = 3

Qualitative Rubric Description

Gave required information
 Slower speech evident
 Frequent and longer pauses and hesitations
 Responded appropriately to most questions and prompts
 Evidence of L1 interference

Qualitative Analysis of Student Sample

All of the required information was given with some additional details. Slower speech was evident at times, indicating translation of thoughts and ideas. There were frequent and longer pauses and hesitations with English rejoinders such as "um." The sample responded appropriately to questions and the prompt with some L1 interference in word selection and anglicized grammar.

Fourth-Year Spanish Speaking Profile
Example Prompt

Your host family in Nicaragua has asked you about pets in the United States. Describe a pet that you have or would like to have. Explain how you care for the pet (what you feed them and how often you feed them, etc.) and describe some activities that you do with your pet. Talk about some common pets that families may have and include unusual or exotic pets that you know of. Conclude by asking them what pets they have or might like to have.

Authentic Student Product

Hola familia. No tengo una mascota pero...el perro es el es la mascota muy común en los estados unidos. Necesitas jugar con los perros y los das comida. Si podrías tener una mascota cual tienes.

Pronunciation $\bar{x} = 1.91$

Exemplar Score = 2

Qualitative Rubric Description

Comprehensible; some accent evident
 Began to sound accurate
 Could use appropriate word stress
 Could imitate target language sounds
 English sounds were occasionally used

Qualitative Analysis of Student Sample

Some accent was evident, but the pronunciation was still comprehensible. The pronunciation used appropriate intonation and word stress. The target language sounds were imitated, and English sounds were rarely used.

Grammar $\bar{x} = 2.00$

Exemplar Score = 1

Qualitative Rubric Description

Mostly appropriate use of verbs
 Some number and gender agreement
 Could use appropriate definite and indefinite articles
 Basic subject/verb agreement
 Word order was anglicized but message was communicated
 Some errors; errors did not interfere with meaning
 Evidence of pronoun use
 Could self-correct

Qualitative Analysis of Student Sample

The grammar structure demonstrated appropriate use of verbs. The conditional verb tense was used. There was correct definite and indefinite article usage with the exception of *uno mascota*. As *mascota* is singular and feminine, the article should have been *una*. This is an example of L1 interference. There was evidence of basic subject/verb agreement. The word order was correct with few errors in grammar. These errors did not interfere with meaning. There was evidence of incorrect direct object pronoun usage with “*los das comida*.” The speaker identified *los perros* as what needed to be replaced but did not use the indirect object pronoun *les* because *los perros* is the indirect object of that sentence. The sample did show evidence of self correction with a recast with “*el perro es el es la mascota*.” The speaker began with the singular masculine definite article *el*, then corrected as *mascota* requires the singular with feminine article. This demonstrated a higher level of language ability.

Vocabulary $\bar{x} = 2.01$

Exemplar Score = 2

Qualitative Rubric Description

Common/basic vocabulary
 Attempted to use vocabulary creatively
 Attempted to use idiomatic expressions
 Speaker met minimum word requirements
 Some repetition of select words

Qualitative Analysis of Student Sample

The vocabulary was beyond common, and the sample did demonstrate some attempt at the creative use of vocabulary and verb forms. There was one vocabulary word that was in error in “*la mascota muy común en los estados unidos*.” This was *muy*, as it is in a superlative phrase and must be *más*. This could be caused again by L1 interference. The minimum word requirements were met in this sample.

Fluency $\bar{x} = 1.94$

Exemplar Score = 2

Qualitative Rubric Description

Gave required information but
no more

Slower speech evident

Frequent and longer pauses and
hesitations

Responded appropriately to
most questions and prompts

Evidence of L1 interference

Qualitative Analysis of Student Sample

All of the required information was given. Slower speech was evident at times. The sample responded appropriately to questions and the prompt with little evidence of L1 interference.

Appendix C

Variation With Language Development

Second-Year Spanish Student Sample Variation

Category	$\bar{x}$	Sample Score	
Pronunciation	1.82	1	<i>Mi clase de español es muy grande y rojo. Yo estudiar mucho y libros es azul. Yo profesora y inteligente.</i>
Grammar	1.77	1	
Vocabulary	1.72	1	
Fluency	1.71	1	
Category	$\bar{x}$	Sample Score	
Pronunciation	1.82	2	<i>Tengo cuatros personas en mi familia. Pero me gustan dos personas mucho. Los personas esta mi madre y mi hermano. Mi madre se llama Deb. Deb tiene treinta y nueve años. Mi madre esta mi amiga mejor. Pero yo puedo hablar por muchas cosas. Mi madre le gustan esquiar acuatico y cocinar pasteles con mi. Mi hermano se llama Christopher. Christopher tiene diez y seis años. Christopher le gustan practicar los deportes futbol norteamericano y baloncesto y muchas cosas afuera. Christopher le gustan los chicas muchos y hablar con las chicas para el noche.</i>
Grammar	1.77	3	
Vocabulary	1.72	3	
Fluency	1.71	3	

Third-Year Spanish Student Sample Variation

Category	$\bar{x}$	Sample Score	
Pronunciation	2.05	1	<i>Para desayuno me gustaria jueves. Para almuerzo te me gustaria jamburguesa y papas fritas y para cena me gustaria carne.</i>
Grammar	2.07	2	
Vocabulary	2.02	1	
Fluency	2.03	1	

Category	$\bar{x}$	Sample Score	
Pronunciation	2.05	3	<i>Me levanto a las seis y media. Me ducho y me maquillo antes de ir al colegio. Me cepillo el pelo y me cepillo los dientes. Yo tomo el carro al colegio. Me encanta la clase de matematicas porque es muy interesante. No me gusta la clase de historia porque es muy difícil. Me gusta la clase de ingles porque es muy fácil. Me gusta la clase de biologia porque es muy divertido. Despues de clases yo practico deportes.</i>
Grammar	2.07	3	
Vocabulary	2.02	3	
Fluency	2.03	3	

**Fourth-Year Spanish Student
Sample Variation**

Category	$\bar{x}$	Sample Score	
Pronunciation	1.91	1	<i>Hoy en ese cuarto el un amigo mire el tele, dureme en la sofa y come comida y bebe un refresco. Muchos actividades son en el cuarto.</i>
Grammar	2.00	1	
Vocabulary	1.91	1	
Fluency	1.94	2	
Category	$\bar{x}$	Sample Score	
Pronunciation	1.91	3	<i>Aquí hay dos chicas en la niebla. Las dos chicas hicieron ángeles de la nieva. Ellas viven cerca de un bosque. También hay un pueblo aquí. Hace frío y bastante nublado.</i>
Grammar	2.00	3	
Vocabulary	1.91	3	
Fluency	1.94	3	