Developing an Online Placement Exam for Spanish Heritage Speakers and L2 Students

Kim Potowski, MaryAnn Parada, and Kara Morgan-Short
The University of Illinois at Chicago

Abstract
This paper reports on the development and piloting of an adaptive, online placement exam that will be administered to L2 and heritage learners of Spanish at the University of Illinois at Chicago. Particular attention is given to the structure of the exam and the linguistic strategies employed to distinguish heritage speakers from L2 learners. Results obtained from three pilot phases are presented along with a description of modifications made during each stage. Further challenges and steps are outlined.

1. Introduction
Many publications have made clear the multiple reasons why separate language courses for heritage speakers, different from those offered to second language (L2) learners, should be available (e.g., American Association of Teachers of Spanish and Portuguese, 2000; Valdés, 1997; Webb & Miller, 2000). Heritage speakers acquire Spanish in naturalistic contexts, typically developing strong levels of oral/aural language while their literacy and metalinguistic knowledge tends to be weak. L2 learners, in contrast, usually present the opposite profile, with strong explicit knowledge of grammar but poor ability to utilize this knowledge to produce/understand fluent discourse. We can understand in general terms the differences between the strengths and needs of these two kinds of students by considering the differences between a student of English as a Second Language (ESL) and a native English-speaking student. Introductory ESL courses focus on developing basic communicative competence (Short, 2000) while the field of English language arts works with students on literacy, genre studies, and other forms of expression (International Reading Association and National Council of Teachers of English, 1996). In our case, the L2 Spanish learner is like the ESL student and is best served with a basic Spanish course. Heritage speakers, however, particularly those who have an ACTFL level of at least Intermediate Mid, are likely better served by a modified native language arts curriculum than by an L2 Spanish curriculum (Potowski & Carreira, 2004).

Given the widely accepted arguments for separate heritage speaker courses, many high schools, colleges, and universities offer them. According to Beaudrie (in press) 40% of college-level Spanish programs nationwide offer specialized heritage speaker courses, although this figure varies considerably by region. For example, 51% of programs in the Northeast (Beaudrie, in press) and 67% of those in California (Valdés, Fishman, Chávez, & Pérez, 2006; Carreira, 2011) have such courses. Each year new programs are created, and there are at least a dozen commercially available textbooks for college Spanish heritage courses (Center for Applied Linguistics, 2012).

However, despite a long history of interest in Spanish heritage language assessment (e.g., Otheguy & Toro, 2000; Teschner, 1983; Valdés, 1989; Ziegler, 1981), to date there have been few published placement exams for Spanish heritage speakers. A recent survey of California

colleges and universities has found that only about 11% of heritage programs report using an exam developed specifically for heritage speakers (Valdés et al., 2006, p. 191). Of the few exams available to the public, most have been created for a K-12 audience; for instance, the *Prueba de Ubicación para Hispanohablantes* [Placement Exam for Spanish Speakers] (Otheguy & García, 1996). At the university level, the only heritage speaker exam commercially available and used at various institutions is the PASS (Parisi assessment system for Spanish) developed in 1983 by Parisi and Teschner at the University of Texas, El Paso. It is described as "a 140-item blind multiple-choice proficiency test" (Teschner, 1983) and claims to "sort heritage learners from other types of false beginners" (González-Pino & Pino, 2000, p. 27). We will return to the important differences between placement versus proficiency exams in a later section, but for now, we note that Parisi and Teschner's (1984) exam is actually a proficiency test used for placement purposes. In addition, MacGregor-Mendoza (2011) mentions some potential concerns that institutions might have with this particular exam, including the lack of documentation available on justifying the selection of item topics and on the development of individual items. Interestingly, though, the question set used to separate natives[1] from non-natives in this exam tests knowledge of colloquial words and expressions, an area that we will show ended up being highly useful in discriminating between the two groups in our own test analysis. Teschner (1983) reports that of the 14 items designed to separate native/heritage speakers from L2 learners, the former group routinely answered 11 to 14 questions correctly while the latter never surpassed four, regardless of how many years a student had studied Spanish. Although a small number of first and 1.5 generation students take the placement exam each year at our institution, it primarily serves second and third generation heritage speakers.

Some institutions place heritage learners using ACTFL (American Council on the Teaching of Foreign Languages) proficiency guidelines and the OPI (Oral Proficiency Interview) or other related measures, such as the widely marketed Spanish WebCAPE (n.d.) standardized exam developed by Brigham Young University. The use of such guidelines and assessments has been criticized, however, as ill suited for heritage learners because they are designed with the foreign language learner in mind and around the classroom learning experience (e.g., Valdés, 1989). Unfortunately, as Fairclough (2006) explains, despite calls for such by organizations such as the Center for Applied Linguistics (CAL), virtually no alternative guidelines exist to assist with heritage language assessment, leaving heritage language programs to develop tools based on local curriculum and resources.[2] (p. 596). In addition to in-house designed paper-and-pencil exams, other traditional methods of identifying and placing heritage students have included background questionnaires, interviews, and self-placement, or a combination of these (Fairclough, 2006; González-Pino & Pino, 2000). Such methods, however, are becoming less realistic as heritage speaking populations grow, resources diminish, and increased efficiency and effectiveness in placement testing is desired.

Thus, there is no easy solution to the challenge faced by postsecondary departments with both a basic language track and a heritage track. Students first must be identified for placement in the appropriate track. However, such identification is not sufficient; heritage speakers, like L2 learners, vary broadly in terms of proficiency, ranging from having limited receptive abilities in the language to fully productive skills (Valdés, 1997; Carreira, 2003), and they require proper assessment for placement into the correct level.

This paper offers details of one department's solution to these concerns, detailing the development of an online adaptive placement exam that, based solely on linguistic criteria, distinguishes heritage speakers from second language learners and further places all students into the appropriate level of each track.

2. Background

The University of Illinois at Chicago (UIC) is characterized by an exceptionally diverse student body, recently earning the U.S. News ranking of the sixteenth most ethnically diverse postsecondary institution in the country for the 2009-2010 school year. Approximately 15% of UIC students are Latino and come from a variety of backgrounds, though they are primarily of Mexican descent. Of the approximately 1,200 students who took the heritage speaker placement exam between 2005-2011, 83% were Mexican[3] and, more specifically, 69% were second generation Mexican (born in the U.S. or arrived before age 6). Another 5% are Puerto Rican and 5% are Guatemalan;, smaller groups include Ecuadorians, Salvadorans, Colombians, and Peruvians. Due to their intense early exposure to Spanish in family and community settings, our heritage student population exhibits very strong levels of oral proficiency, typically between Advanced Mid and Superior on the ACTFL OPI scale (readers are directed to Breiner-Sanders et al., 2000 for a more complete description of these levels of oral proficiency). Thus, our heritage speaker program is designed for students with this level of Spanish proficiency, as they form the vast majority (approximately 95%) of our heritage speaker population.

Currently, UIC's Spanish language program enrolls approximately 1,700 students in two tracks: the Basic Language Program (BLP, 1,500 students per semester) and the Heritage Language Program (HLP, 200 students per semester). Most Colleges on campus have a foreign language requirement consisting of successful completion of a fourth-semester course. The BLP consists of Spanish 101, 102, 103 and 104, with 104 counting as the fourth-semester course satisfying the foreign language requirement. BLP students who place directly into Spanish 104 need only pass that course to fulfill their language requirement; those who place into 103 must pass 103 and 104, and so on. Thus, the maximum number of courses that a true beginner BLP student needs to complete is four. The HLP consists of Spanish 113 and 114, with 114 being the fourth-semester course satisfying the foreign language requirement. Thus, the maximum number of courses that a heritage speaker[4] needs to complete is two.

Each track has its own paper-and-pencil placement exam. The placement exam for the BLP was developed sometime prior to 1999 and has never undergone a formal evaluation. It consists of 100 multiple-choice items and is required of all incoming students (who did not learn Spanish outside of a classroom) before registration.

Heritage speakers self-identify as such, using the criteria listed on the Department website that read as follows: "You should take the Spanish for bilinguals placement test if you learned Spanish in a natural, non-academic environment (at home, during residence abroad, etc.)." Students who feel they fit these criteria are instructed to take the heritage placement exam. This exam consists of three parts: background questions (including country of origin, immigrant generation, time spent in Spanish-speaking areas, and prior schooling in Spanish), a written essay

of 18-20 lines in response to one of three prompts, and a short translation from English to Spanish that contains hypothetical sentences with compound verb tenses (e.g., "If Pepe had studied, he would have gotten an A on the test"). In general, heritage speakers who read these criteria on the Department website understand that they are supposed to take the heritage speaker exam, so this procedure of separating heritage speakers from L2 learners has been relatively adequate. However, approximately 10 heritage speakers each semester either mistakenly think that they do not fit the criteria, or they do not read the criteria (some are not properly directed to the website by their academic advisors) and therefore they enroll in BLP courses. BLP instructors are informed every fall during orientation that if they sense that a particular student has strong heritage speaker proficiency, they must tactfully and privately inform her/him that a brief appointment with the HLP program director is required.

Thus, Spanish heritage speakers who wish to study Spanish and, importantly, who possess levels of Spanish proficiency that fit well in the HLP,[5] are not permitted to take BLP courses; they are required to enroll in the HLP. This is the same logic that would prohibit a highly proficient/native speaker of English from taking an introductory ESL course. We will return in a later section to the question of the relatively few heritage speakers on our campus who do not possess sufficient proficiency in Spanish to be successful in our heritage speaker program.

Heretofore, the L2 and heritage placement exams have been administered via paper and pencil at the campus' Office of Testing Services. Students self-identify as heritage speakers or as L2 learners and each group takes different tests. The L2 exams, which are entirely multiple choice, are scored at the Office of Testing Services, while the heritage exams are sent for evaluation within the Department. In a handful of cases each semester, L2 and HS students test out of the fourth semester course (L2 or HS) and go directly to a fifth semester course, where they all study together.[6]

An array of factors combined to form what Davidson and Lynch (2002, p. 77) would term our collective "mandate" to redesign our placement test. Most important was that we had two separate placement exams and students had to self-identify in order to take the correct one; we sought to create one exam that would separate heritage speakers from L2 learners. Other factors included enrollment growth, reduction in campus resources, technological advancements, the awkwardness of BLP instructors having to request that heritage speaker students meet with the HLP program director (and the ensuing scheduling difficulties during the two-week drop/add period), and legal concerns that might arise from a misconception that placement in the HLP is based on ethnicity rather than linguistic characteristics.[7] We desired an assessment tool that would achieve the following goals:

- Separate L2 learners and heritage speakers based purely on linguistic criteria.
- Eliminate paper, as well as appointment scheduling in the Office of Testing Services, by online administration.
- Comply with standards of validity and reliability.

In short, we sought a "universal" exam for all Spanish learners (i.e., all students on campus wanting to take Spanish) that, regardless of time and location, would be available by means of testing software also capable of documenting and periodically generating reports on student results.

3. Test Design

3.1 Principles of Test Development

Placement testing contrasts in important ways with other test types (see Table 1).

Table 1

Test Types (based on Davidson & Lynch, 2002)

	Purpose
Achievement	To test what was learned
Proficiency	To test what is known (with no reference to duration, quality, etc. of learning)
Placement	To test for the purpose of putting somebody in a particular course sequence or level
Diagnostic	To determine areas of need
Aptitude	To determine ability to learn

However, there is considerable overlap among these test types. For example, many language departments administer a proficiency test, create a scale of what scores equal what courses, and use the results for placement purposes. An achievement test can also show diagnostically what areas are weak. Rather than using a proficiency test for placement purposes, we aimed to develop a genuine placement exam that is linked to the course content taught in our two tracks.

Although placement tests serve a particular purpose, how they should be developed overlaps with other test types. Davidson and Lynch (2002) argue that all test development should be:

1. **Iterative** – consisting of cycles of feedback-laden improvement over time
2. **Consensus-based** – test should result from dialogue and debate among a group of educators and interested parties
3. **Specification-driven** – be constructed using an efficient generative recipe that fosters dialogue and discovery at a higher, more abstract level than that achieved by analysis of a simple item or task

The authors explain that ongoing evaluations of reliability and validity at the item and task levels naturally form part of a test design fitting this description, but caution developers that a test is never truly in final form; instead, a working version should be periodically reviewed to avoid what they call the "set-in-stone" problem (p. 64). We have made an effort to adhere to these

principles during the development of our exam and to keep an open mind regarding ongoing needs for improvements in the future. The next sections will review the steps followed in the creation of our placement exam, and will include information on the general organization of the test, item writing, and statistical analyses.

3.2 Exam Structure

By way of small-group consensus, we began the development process by sketching a diagram depicting the desired structure of our exam. The successful separation of heritage speakers from L2 examinees required an adaptive element whereby following an initial set of test items to be taken by all students, they would, in accordance with their performance, be automatically directed to new and different blocks of questions until their placement level is determined. Figure 1 illustrates the structure and flow of the exam, with Test B as the entry point for all examinees. Depending on the score obtained on Test B items, a student is then directed either up to Test C, where heritage/L2 separation occurs, or down to Test A, to be further tested for placement into first- or second-semester L2 Spanish. From Test C, L2 and heritage students are directed to their respective final sets of questions, i.e., Test D-L2 or D-HS, where they are further assessed for placement into third- or fourth-semester L2 Spanish or first- or second-semester bilingual Spanish in the case of heritage learners. Scores on Tests D-L2 and D-HS would also determine whether a student qualifies to place out of the basic language program altogether.

Figure 1. Exam Flowchart
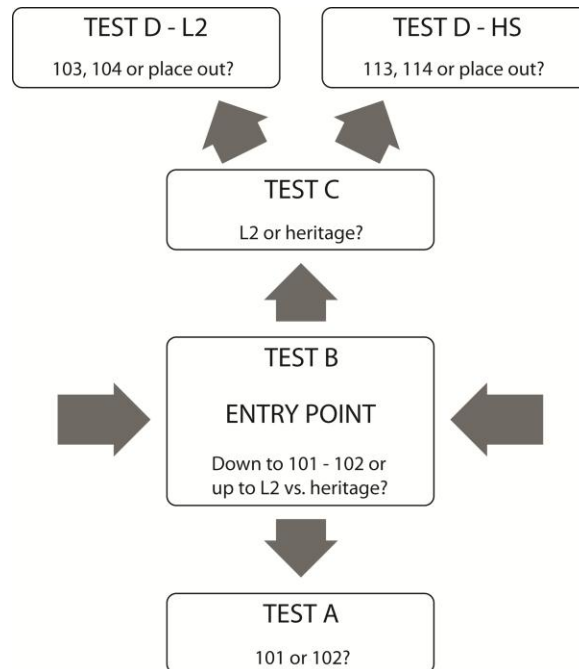
3.3 Exam Content
After establishing the structure of the exam, we turned our attention to item writing. First, the testing center provided an item analysis of the current L2 placement exam, based on several years of accumulated item reports. A small number of items were retained from this exam, provided they had an acceptable difficulty index (around .50) and had been shown to discriminate very well between lower and upper groups (with discrimination indices of .25 and higher). Decisions on the content of items for the L2 test blocks (including Test B) were made based on a review of several widely used classroom syllabi and textbooks. Observations were compiled regarding typical vocabulary themes and grammatical concepts corresponding to each level. These summaries (see Table 2 for an example) then guided us in the creation of level-appropriate multiple-choice items, a process that contributed to the content validity of the test. For example, since the purpose of Test A is to separate true L2 beginners from those having basic knowledge, i.e., that obtained in a semester-long college course or a year-long high school course, it is comprised solely of items based on the usual content of a first-semester L2 college course. The number of correct items on Test A would then determine whether a student places into Spanish 1 or Spanish 2.

Table 2

Example of Level-appropriate Item Content, Spanish L2 Course

Student must show sufficient knowledge of these concepts/domains to be placed out of Spanish 1 and into Spanish 2	Grammar	Vocabulary
	▪ present indicative -- regular & frequent irregular verbs ▪ basic *ser/estar* difference ▪ gender/# agreement ▪ DO pronouns	▪ greetings ▪ numbers 0-100 ▪ telling time ▪ weather, basic leisure activities ▪ describing people ▪ colors ▪ family

Test B, the entry point for all examinees and the mechanism for separating Spanish 1 and 2 students from the rest, was likewise made to include only items reflecting content typically mastered in second-semester L2 Spanish. Because of the relative simplicity and straightforwardness of the items in Test B, we did not anticipate them being problematic for our heritage speakers (see Table 3 for example items from Test B).

Table 3

Example Items from Test B

¿Bailar? ¡Me ....!   (a) fascina   (b) fascino (c) fascinar
Algo típico que hacen los niños es..... (a) jugar al escondite   (b) navegar un barco   (c) buscar empleo
Me gustan ....   (a) el verano   (b) correr y nadar   (c) los plátanos

Following the same logic, Test D-L2 was made to include only items reflecting third-semester material, the total number correct determining placement into Spanish 3 or 4, or the granting of "place-out" status. Section Four will review the results of three piloting phases and will evaluate to what degree each test block fulfills its intended function.

The creation of items for the blocks intended for heritage students, Tests C and D-HS, followed a different route. Item types to be included in Test C (where L2/heritage separation occurs) were determined based on differences in the linguistic profiles of these two groups. For instance, one section of items targeted metalinguistic knowledge in requesting the matching of verb forms with verb tense labels. It was expected that L2 learners would possess greater familiarity with verbal terminology and would therefore more accurately identify examples of each tense. Additionally, it was predicted that they would outperform heritage learners on the correction of errors involving spelling and accent placement. Alternatively, another section was dedicated to colloquial lexical items and phrases, some dialect-neutral and others Mexican to reflect our local population. We predicted that such informal vocabulary would be recognized by even low proficiency heritage speakers, yet remain unfamiliar to high proficiency L2s. We will see below that only the vocabulary items discriminated reliably between heritage speakers and L2 learners.[8]

Table 4

Summary of Types of Items on Test C

Items 1-12	Informal vocabulary
Items 13-22	Spelling, accent placement
Items 23-30	Gerund/infinitive, prepositions, coloquial morpho-syntax
Items 31-38	Verbal tense identification

The content of Test D-HS, meant to separate heritage speakers into the two courses Heritage 1 and Heritage 2, included vocabulary, accent mark placement, and spelling and grammatical errors commonly resulting from aural learning and/or contact with English (such as direct object marking and *gustar*-type verbal constructions). There were also two brief reading passages on which students were to answer three comprehension questions. Following the format of the other

test blocks, all items for Tests C and D-HS were presented in a multiple-choice format with three to four response options. In addition, as in the former exam and separate from the discrete item portion, heritage students submit a several-paragraph writing sample on an assigned topic (as described in Section 2). These topics and skills typically form part of heritage speaker curricula and, recalling the important fact that this was to be a placement test, not a general proficiency test, we aimed to link students' results with our course content.

Table 5

Summary of Types of Items on Test D-HS

Items 1-8	Vocabulary
Items 9-24	Spelling, accent placement
Items 25-36	Definite articles, verbal expressions
Items 37-47	Reading comprehension

4. Pilot Testing

4.1 First Phase

A first pilot of our exam was administered in the spring of 2009 to a total of 461 students dispersed across all four levels of the basic language program and the two levels of the heritage language program. All students were given the opportunity to complete the test at home for an extra credit assignment and were expressly instructed to do so without the aid of any individual, textbook, notes, internet pages, etc. They were assured that performance on the test would have no bearing on their class grades and that their identity would remain anonymous to all involved parties. The test was administered in a paper-and-pencil format after which scores were tallied and organized for analysis.

4.1.1 Results and Discussion

The data reported for this and all pilot phases is based on test item analyses rather than subject analyses, meaning that what is measured across levels is the difference in group performance on the individual items of each subtest. Table 4 summarizes the mean scores for each class level on the test blocks administered. Even though accuracy scores mostly ascended along with class level, these basic figures indicated that a greater gap between pertinent levels on all test blocks was needed. For example, recall that the intended purpose of Test B (the entry point) was to separate beginning levels (Spanish 1 and 2) from all others; nonetheless the results show that Spanish 2 patterns very much like Spanish 3 and 4. A second issue from this first phase of piloting is that in general, we wanted to see higher scoring test-wide, particularly for the upper level students on each test block, who are expected to be familiar with the item material.

Table 6

Mean Percentage Scores per Test Block (Pilot #1)

	Level	N	A (10 items)	B (20 items)	C (35 items)	D-L2 (30 items	D-HS (48 items)
			M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
BLP	Spanish 1	55	77 (16.3)	41 (15.5)	44 (8.4)	--	--
	Spanish 2	100	72 (17.8)	54 (14.6)	49 (11.3)	--	--
	Spanish 3	100	--	55 (19.2)	49 (14.7)	41 (15.3)	--
	Spanish 4	100	--	56 (19.4)	52 (11.2)	47 (16.6)	--
HLP	Heritage 1	42	--	83 (11.0)	69 (13.1)	--	60 (14.0)
	Heritage 2	64	--	84 (15.2)	72 (12.8)	--	64 (13.4)

In addition to examining accuracy scores, we ran an item-by-item analysis to assess items according to: (1) their difficulty index, which measures the proportion of examinees who got the item right, (2) their item-test correlation, which shows whether the scoring on an individual test item correlates with overall test scores and, similarly, whether overall test scores correlate to scoring on a particular test item and whether generally low scoring students are scoring low on the same item. Based on these statistics, test items were identified using the following criteria: (a) difficulty indices above .80 or below .40 were flagged as too difficult or too easy, respectively, and (b) Pearson correlation values below .20 were noted as too low. By keeping the criteria as generous as possible, we were able to retain in the mix some of the more difficult questions as well as some of the easier items. Problematic items were either eliminated or revised and additional items were created following the format and topics that had proven most successful in discriminating levels.

Test C, which merits special attention in this paper, initially consisted of four parts (as discussed briefly in 3.3). Table 5 summarizes Test C mean scores by section. A breakdown of scores per section revealed that the section testing colloquial vocabulary knowledge (C.1) was by far the most effective in distinguishing L2 learners from heritage students. Table 6 displays several examples of scores on these items. We therefore decided to retain only these vocabulary items in Test C and increased the item count from 12 to 20.

Table 7

Mean Percentage Scores on Test C by Section (Pilot #1)

	Level	C.1	C.2	C.3	C.4
BLP	Spanish 1	29	66	41	40
	Spanish 2	30	66	44	60
	Spanish 3	31	64	44	59
	Spanish 4	31	69	47	68
HLP	Heritage 1	84	76	55	54
	Heritage 2	85	77	59	60

Table 8

Examples of Test C.1 (Informal Vocabulary)

Susana no acabó su trabajo porque le dio (ganas, enojada, flojera).
  Susana didn't finish her work because she got (desire, angry, lazy).
No le cuentes nada a mi tía porque es muy (perezosa, chismosa, callada).
  Don't tell my aunt anything because she's very (lazy, gossipy, quiet).
Mi sobrino no camina todavía, apenas aprendió a (gatear, cosechar, incubar).
  My nephew doesn't walk yet, he barely learned how to (crawl, harvest, incubate).

We sought to ensure that the items ultimately selected for Test C were not highly limited to Mexican Spanish; otherwise they would not be valid for speakers of other Spanish dialects. It merits emphasizing, then, that placement instruments must be grounded in local student populations, which prohibits the use of one test for all populations of heritage speakers. In addition, if a test is truly a placement test according to the criteria in Table 1, it should be closely linked to the course content. We will return to this issue when discussing the written essay on part D-HS.

Sections C.2, which included items on spelling and accents, and C.3, which requested corrections in sentences containing grammatical errors "typical" of heritage learners, also returned important differences between the two groups, but in the opposite direction of our initial predictions. The data would seem to indicate that an L2 advantage in these areas is more likely to emerge at more advanced levels rather than in first- and second-year students. Section C.4, which tested the successful pairing of verb tense classifications with matching forms, produced unsystematic results in which fourth-semester L2 students outperformed heritage learners while all other L2 levels performed comparably or, in the case of Spanish 1, more poorly.

Test D-HS also included its share of difficulties. Item-test correlation values were low for a number of items that did not distinguish well between high and low scoring students. Other items emerged as either much too difficult or far too easy for both heritage levels. The reading passages, for example, proved to be overly simple, as both groups scored very high on comprehension questions. Items with acceptable item-test correlation and difficulty indices were retained, and reports on the frequency selection of distractors also aided us in determining which ones to retain and which to alter. At this stage, new questions were created that more closely reflected the newly implemented curriculum in our heritage track. That is, we sought to make this a true placement exam by utilizing topics that are taught in our heritage speaker program. These items included *ser*/*hacer* confusion, gerund/infinitive usage, rules of capitalization, specific cases of accent mark placement (e.g., *habló* vs. *hablo*), relative pronouns, imperfect subjunctive, and prepositions.

4.2 Second Phase
A second pilot of our exam was administered to students during the second week of the spring 2010 term, before they had received much instruction in the new course. This version reflected the changes outlined in the last section and others of a logistical nature: (a) class time was

reserved in all basic and heritage course sections to allow for a proctored environment in which all attending students could take the exam; (b) students recorded their answers on a Scantron to simplify subsequent scoring and analysis; and (c) students were limited to 60 minutes, though many finished in less time. A total of 1,183 students participated in this pilot phase. An increase in total items for each test block lengthened the pilot test, but, we hoped, would ultimately facilitate a larger bank of items, which the final online test could eventually use in alternation in order to increase test security.

4.2.1 Results and Discussion

Table 7 summarizes the mean scores on the test blocks taken by all class levels in our second pilot phase. Results for Test C (informal vocabulary) echo the positive findings of the first pilot and display an even larger gap between L2 and heritage student scores. We were also able to observe the more pronounced separation between the Spanish 2 and the Spanish 3/4 groups on Test B (the entry point) as well as between the two heritage levels on Test D-HS.

Unfortunately, the average score still seemed too low overall, particularly on the tests where placement is determined (A, D-L2, and D-HS) and an adequate mastery of concepts should be evident for the upper level (e.g., sufficient knowledge of Spanish 3-level material by those placing into Spanish 4 on Test D-L2). Likely contributors to the low averages were: (a) too difficult items; b) insincere test takers; (c) too little time for some students; and (d) accidental misnumberings on the Scantron sheet. A problem unique to D-L2 was that 10 very difficult final items intended only for placing out purposes were included in the overall analysis, thus decreasing the mean scores substantially.

Table 9

Mean Percentage Scores per Test Block (Pilot #2)

	Level	N	A (40 items)	B (40 items)	C (20 items)	D-L2 (40 items)	D-HS (30 items)
BLP			*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
	Spanish 1	117	35 (12.1)	30 (11.7)	--	--	--
	Spanish 2	301					
			52 (16.6)	45 (15.7)	--	--	--
	Spanish 3	293	--	50 (14.2)	27 (7.3)	35 (14.9)	--
	Spanish 4	358	--	59 (16.0)	30 (10.4)	39 (16.0)	--
HLP	Heritage 1	47	--	90 (10.8)	85 (12.8)	--	57 (15.8)
	Heritage 2	67	--	91 (9.9)	87 (10.8)	--	72 (16.1)

In order to confirm the above observations, we ran analyses of variance that examined whether students at various course levels were performing differently on the placement test. The results of a one-way repeated measures ANOVA (with class level as the within-item factor) for Test B show a significant difference among levels: $F(5, 195) = 259.63$, $p < .001$. Pairwise comparisons using a Bonferroni correction to maintain an alpha level of .05 revealed significant differences (p < .001) between Spanish 1 scores and all others. Spanish 2 item scores were also found to be significantly different from all other levels (p < .001), except Spanish 3 (p = .158). These results, which are visually presented in Figure 1, provide positive evidence that Test B, comprising the

entry point set of questions, is able to successfully separate first-semester L2 students from second-year L2s (Spanish 3 and 4) and heritage learners (Heritage 1 and 2). However, despite mean score differences in the predicted direction, Test B did not significantly distinguish between the key class pair of Spanish 2 and Spanish 3.

Figure 2. Distribution of Test B by Level
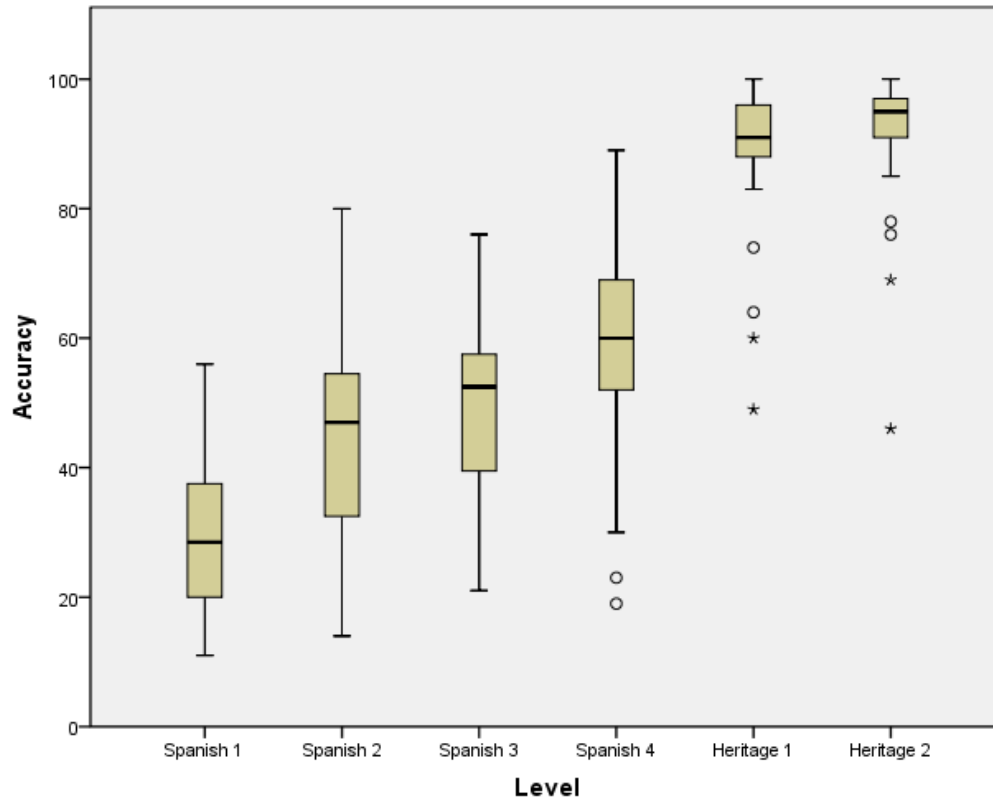
A one-way repeated measures ANOVA for Test C also provided evidence of significant differences in performance among the course levels: $F(3, 57) = 428.57$, $p < .001$. Follow-up pair wise comparisons between the heritage and L2 levels, Spanish 3 and Spanish 4 versus Heritage 1 and Heritage 2, confirmed that the mean scores were significantly different (p < .001), leading to the conclusion that Test C is indeed a highly effective tool for discriminating between the two types of learners (L2 and heritage). This exceedingly well-defined separation between the two class types is illustrated in Figure 2. Results from a paired t-test on the data for block D-HS also showed a significant difference (p < .001) between the two heritage levels (113 and 114), as schematically shown in Figure 3. In addition, similar paired analyses carried out for levels Spanish 1 and Spanish 2 on block A (p < .001) and levels Spanish 3 and Spanish 4 on D-L2 (p < .01) were also returned as statistically significant.

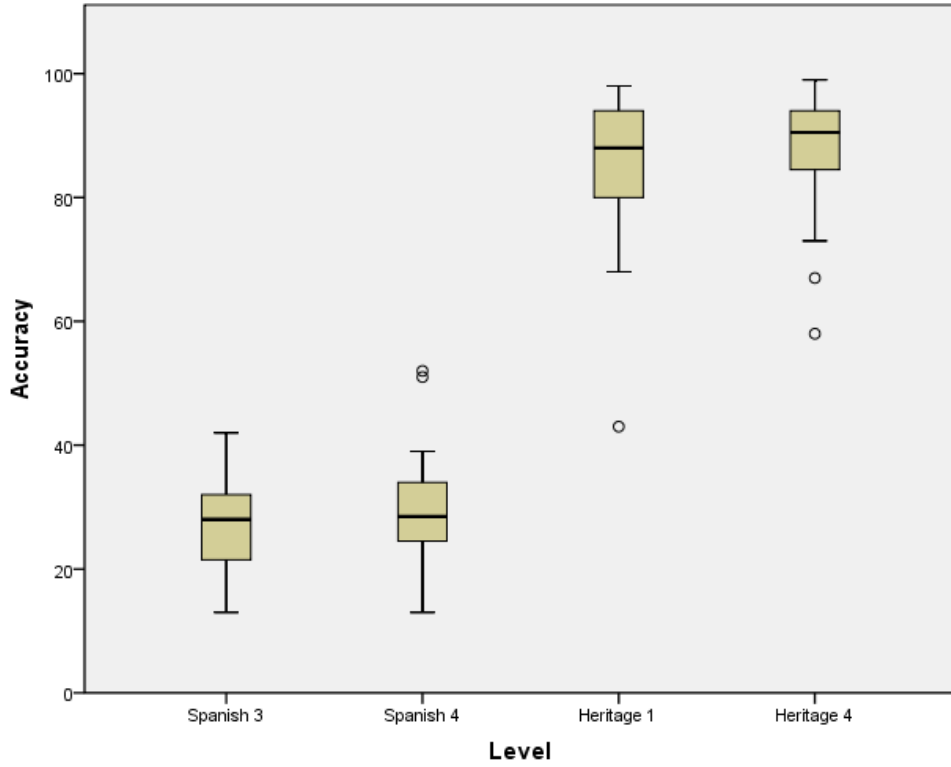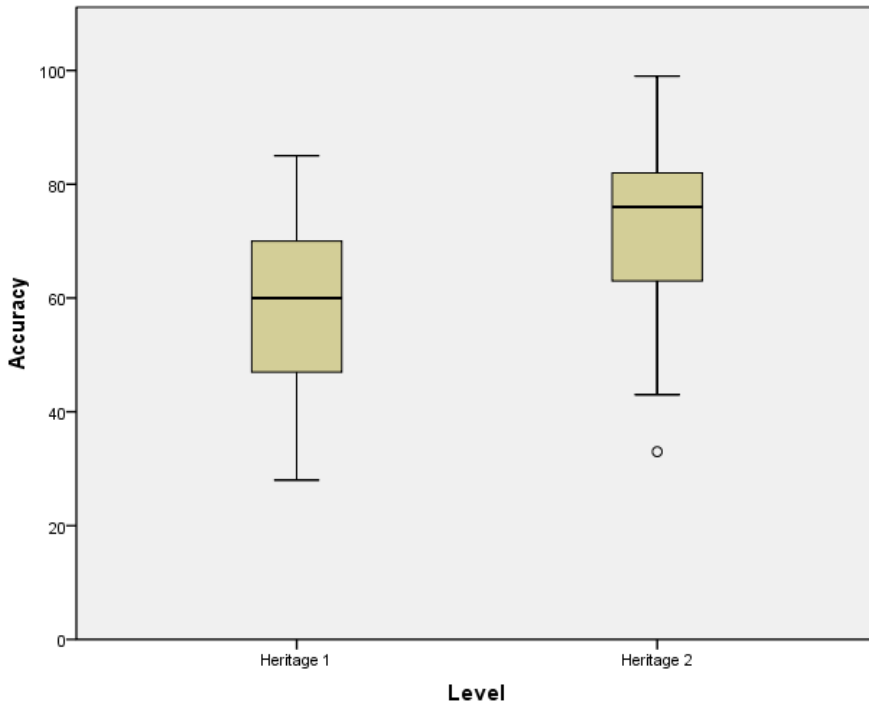*Figure 3. Distribution of Test C by Level*

Figure 4. Distribution of Test D-HS by Level

Despite the statistical significance found among levels, we believed the test required additional modification. Item analyses revealed once again which questions were problematic either due to their level of difficulty or correlation with the rest of the test. Using the same criteria as with the first pilot version (i.e., same item-test correlation value and difficulty index cutoffs), several items were eliminated while others were subjected to slight alterations in the target item or distractor set. The retention of only the most effective items reduced the item counts of Tests A, B and D-L2 to 20, 15 and 20, respectively. We also separated out the 10 "place-out" questions from the remainder of the items in D-L2 in order to improve the accuracy of results. Alterations to C and D-HS were minimal.

4.3 Third Phase

Having implemented the changes outlined in the last section, and in accordance with our goal for an iterative developmental process, we proceeded to a third pilot phase. This pilot took place during the summer months, thus causing limitations on the student population available for participation. In spite of the smaller participant pool, we succeeded in obtaining participants for all levels but Heritage 1, for which there is no course offered during the summer. This phase combined the methods of the first and second pilots in that students were given the test as a required assignment,[9] yet class time was not allotted for its administration. Students were instructed not to consult any aids while completing the test and were once again assured that the results would have no bearing on their class grade.

4.3.1 Results and Discussion

Table 8 shows the mean scores obtained by each level on pertinent test blocks. Important improvements can be observed in the scores of the upper level(s) on each test block. Once again, separate one-way repeated measures ANOVAs were run on the results of each test in order to confirm whether learners at different levels did indeed perform differently on these tests. On Test B, significant differences were again observed among course levels ($F(4, 56) = 61.76$, $p < .001$), as depicted in Figure 4. Especially noted were differences between the heritage student scores and those of all L2 levels. Pairwise comparisons using a Bonferroni adjustment showed a significant difference ($p < .001$) between each of the two lowest levels (Spanish 1/2) and Spanish 3, Spanish 4, and Heritage 2, except in the case of the Spanish 2-3 pair, which once again did not reach significance despite a 6% difference in mean scores. As predicted, the content and structure of these initial basic "entry" items do not present challenges to heritage learners in the least. Tests A and D-L2 were found once again to distinguish between the two levels tested with t-test significance values at $p < .001$ and $p < .05$ respectively.

Differences between levels were also observed in Test C scores ($F(2, 38) = 359.41$, $p < .001$). Post hoc analyses once again confirmed a significant difference between the means of L2 and Heritage levels ($p < .001$), with between-group means even more divergent this time around. Thus, importantly, our results from three different pilot phases offer evidence in support of the use of colloquial vocabulary as an effective yet simple linguistic device for separating out heritage learners from even relatively advanced L2s. Owing to the lack of level Heritage 1 participants during the third phase, no comparisons could be made between the two heritage class levels on Test D-HS.

Table 10

Mean Percentage Scores Per Test Block (Pilot #3)

	Level	N	A (20 items)	B (15 items)	C (20 items)	D-L2 (20 items)	D-HS (30 items)
BLP			*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
	Spanish 1	52	46 (14.4)	37 (14.4)	--	--	--
	Spanish 2	49	68 (11.0)	51 (17.5)	--	--	--
	Spanish 3	71	--	57 (15.2)	31 (9.2)	37 (20.1)	--
	Spanish 4	37	--	69 (14.9)	33 (11.5)	44 (17.9)	--
HLP	Heritage 1	--	--	--	--	--	--
	Heritage 2	7	--	96 (6.4)	96 (8.2)	--	75 (10.2)

To further probe the ability of the current version of the placement exam to distinguish among the levels and, more importantly, to be able to place learners in the correct level, we ran discriminant analyses on Tests A, B, C and D-L2 from the third testing phase as well as on Test D-HS from the second testing phase. Discriminant analysis provides information about how well an independent variable (the scores on the test) predicts group assignment (the course level). The results from the discriminant analyses showed that the discriminant function for four of the five tests was significant (see Table 9). Thus, course level could be reliably predicted by the mean scores on an item for all tests except D-L2.

Table 11

Discriminant Function for Each Test

Test	Wilk's Lambda	Chi-square	df.	Sig.
A	.559	21.789	1	.000
B	.626	34.002	1	.000
C	.091	138.085	1	.000
D-L2	.964	1.361	1	.243
D-HS	.815	11.749	1	.001

More specifically, Test A correctly classified 85% of the Spanish 1 group item means and 80% of the Spanish 2 item means. Test B correctly classified 76.7% of the lower level (Spanish 1 and 2) item means and 73.3% of the upper level student (Spanish 3, 4, and Heritage) item averages. For Test C, 100% of original L2 and Heritage item means were classified correctly. Finally for Test D-L2, only 65% of the item mean scores and 55% of Spanish 4 scores were classified correctly, and the discriminant function indicated that course level could not be reliably predicted by this test. Although we did not have two levels of heritage students in the final testing phase, we ran a discriminant analysis for Test D-HS based on the phase 2 data, which indicated that that test correctly classified 63.3% of the Heritage 1 student item means and 66.7% of the Heritage 2 item scores. These analyses validate the conclusion that Test C reliably predicts L2 and heritage students as such with 100% accuracy. With the exception of Test D-L2, all other tests also

reliably assign item scores to the appropriate level, although Tests A and B seem to do so more accurately than Tests D.

Lastly, a final examination of the overall reliability and internal consistency of each test was conducted using Cronbach's alpha (Pilot 2 data were used for Tests C and D-HS). Generally, reliabilities of 0.7 or higher are considered sufficient, but others argue that in applied settings where important decisions are made based on assessment scores, a reliability of 0.9 is the acceptable minimum (Nunnally & Bernstein, 1994). The alphas for all test blocks but one exceeded .95 (Test A, α = .979; Test B, α = .984; Test C, α = .997; Test D-L2, α = .810; Test D-HS, α = .98). Thus, all of the tests met acceptable levels of reliability, and all but one test met the highest criterion for an acceptable reliability level.

4.3.2 Next Steps
Although the current version of the placement exam appears to be valid and reliable, as noted in the results and discussion of third pilot phase, it would undoubtedly benefit from further improvements . For example, attention should be directed to Test B in order for a significant difference in scoring to be achieved between Spanish 2 and Spanish 3. We predict that adjustments to existing items will improve reliability.

The pilot data collected from students enrolled in our program will aid in the establishment of cut-off points for placement in various levels. It has been suggested, however, that it would be of great benefit to conduct a final pilot in which the exam would be administered only to a selection of students at each level whom instructors can verify are well-placed (Jerry Larson, personal communication, 2009). We plan to do so in the near future.

A particular issue we must consider when establishing cut-offs is the reality of low proficiency heritage learners. Because our exam is designed to divide heritage and L2 learners utilizing purely linguistic criteria, it is quite possible that less proficient heritage learners will place into L2 Spanish, which is not ideal, but likely better than having them struggle in the heritage speaker course. The results of our second pilot phase indicated that no Heritage 2 student scored below 60 percent on Test C, but that six out of 47 (13%) Heritage 1 students did. While we recognize that a low Heritage 1 is not necessarily a high Spanish 4, we also acknowledge that our course offerings and curriculum cannot accommodate every type of heritage learner as appropriately as desired, particularly the 10 or so low proficiency heritage speakers per semester who take our classes. Since we were unable to collect data for Heritage 1 students in our third pilot phase, it is not yet known to what extent post-second-pilot changes might mitigate this problem.

In addition to the need to retest items from D-HS with both Heritage 1 and Heritage 2 students, it is important to note that, even if the final version of the exam proves to be effective in separating Heritage 1 from Heritage 2, we will still keep the written essay portion of the placement exam. We believe that the most accurate placement will reflect our program, which is centered on making academic arguments in writing. The disadvantage of continuing to spend time analyzing essays,[10] rather than relying wholly on multiple choice answers for placement, is, in our mind, far outweighed by at least three advantages. First, while Test D-HS ideally will accurately place students into Heritage 1 or Heritage 2, when it comes to placing students out of the HLP, we

believe that a more comprehensive measure of written academic argument is required. Second, we plan to continue using placement exam essays as entrance point data, with which we compare a similar essay that students produce at the end of the program in order to measure their growth. Finally, these essays provide a wealth of data that can help us notice overall changes in students' linguistic uses and writing and adjust our program content accordingly.

4.4 Selecting Testing Software
In our experience, the search for suitable testing software was a challenging aspect of the development process, particularly if there are budgetary restraints and limited access to on-site instructional technology expertise and support. In addition, when dealing with an undergraduate population of 16,000 or more students, many campus parties become involved in order to allow for a smooth transition from one testing process to another. This is an area that remains unresolved for us and must be finalized before our online exam can be launched. We have, however, performed an extensive survey of existing testing software options in an effort to locate one compatible with our specific requirements, which include the following:

- Internet capabilities
- Integration with university student authentication for sign-on
- Integration with university student portal for automated score posting
- Little formal programming training or skills required
- Range-based release of test blocks
- Randomization of questions within blocks
- Expedited grading
- Ability to generate periodic statistical reports
- Positive reputation among academic institutions

Two options that seem promising are (a) Questionmark's Perception (n.d.) online assessment software by and (b) the learning management system Moodle. Perception is used by numerous academic institutions nationwide, including at the University of Houston in the administration of their placement test for Spanish heritage learners (Fairclough et al., 2010). It is characterized by a wide-ranging set of functions requiring virtually no programming training. Moodle (Moodle Trust, n.d.) is open-source and, inasmuch as it is a learning management system (like Blackboard (n.d.)), it has capabilities beyond testing. It is quickly gaining momentum across the world and its users contribute to a wealth of help forums online. It also requires little training to manage.

4.5 Conclusions
In this article we have outlined the steps and challenges involved in the drafting of a Spanish placement examination that aims to serve the L2 and heritage student alike. The motivations for this project were varied and, in addition to the practicality of consolidating two tests into one, included the wish to streamline the placement process through (a) the elimination of a physical, proctored setting, (b) a reduction in time and costly resources required, (c) the elimination of the need for students to be directed to program information and subsequently identify themselves correctly as heritage speakers, and (d) a more efficient collection of statistical data. The implementation of our test will also serve to separate heritage speakers from L2 learners using language-based criteria only, avoiding any unfounded concerns of discrimination.

Students will find this test more practical, as a large majority commute to campus and will benefit from the ability to take the exam remotely. Due to the "low-stakes" nature of placement testing, we are not overly concerned with the absence of the traditional proctored environment. Sign-on requirements and question randomization are two features that will discourage dishonesty. But perhaps the greatest motivator for students to resist reliance on unapproved sources is that the end result of being placed into a course that is too difficult will only incur negative consequences at a personal level. Such a perspective is shared by Larson and Hendricks (2009), who stated that if a "test is taken for a low-stakes intent such as placement or diagnostic purposes, no proctoring is necessary," adding that "in these kinds of testing situations, there is no reason for individuals taking the test to give less than an honest effort, inasmuch as they would be the ones to suffer the consequences of any testing impropriety." (p. 5). The fact that there exists the possibility of placing out does increase the stakes of the exam to a degree; however, given the minimal number of students who are granted this status each year, these cases will be dealt with individually by verifying results through an in-person interview with program directors. With a web-based test, the BLP students will also be pleased to receive immediate feedback on their performance (heritage speakers must wait for their essays to be evaluated).

Finally, we have discovered that a language placement exam is not "one-size-fits-all" and that it must reflect the needs and composition of the heritage speaking population it is intended for, as well as the goals of the curriculum decided upon by the institution. This is not to say, however, that it cannot be adapted for use at other institutions where demographics and heritage language programs may differ. In particular, much of the informal vocabulary (Exam C) is likely useful at other campuses to separate L2 students from heritage speakers.

References

American Association of Teachers of Spanish and Portuguese. (2000). *Professional development series handbook for teachers K–16: Spanish for native speakers* (Vol. 1). Fort Worth, TX: Harcourt College.

Beaudrie, S. (in press). Research on university-based Spanish heritage language programs in the United States: The current state of affairs. In S. Beaudrie & M. Fairclough (Eds.), *Spanish as a heritage language in the U.S.: State of the field*. Washington, DC: Georgetown University Press.

Blackboard Inc. (n.d.) Blackboard Learn [education technology platform]. Washington, DC: Author. Available from http://www.blackboard.com

Breiner-Sanders, K., Lowe, P., Miles, J., & Swender, E. (2000). ACTFL proficiency guidelines – Speaking, revised 1999. *Foreign Language Annals, 3*(1), 13-18.

Carreira, M. (2003). Profiles of SNS students in the twenty-first century: Pedagogical implications of the changing demographics and social status of U.S. Hispanics. In A. Roca & M. C. Colombi (Eds.), *Mi lengua: Spanish as a heritage language in the United States* (pp. 51-77). Washington, DC: Georgetown University Press.

Carreira, M. (2011, January). *Heritage language teaching: Opportunities and challenges for community colleges*. Paper presented at the 126th Annual MLA Convention, Los Angeles, California.

Center for Applied Linguistics. (2012). *Heritage languages in America*. Retrieved from http://www.cal.org/heritage/index.html

Davidson, F., & Lynch, B. (2002). *Testcraft: A teacher's guide to writing and using language test specifications*. New Haven, CT: Yale University Press.

Fairclough, M. (2006). Language placement exams for heritage speakers of Spanish: Learning from students' mistakes. *Foreign Language Annals, 39*(4), 595-604.

Fairclough, M., Belpoliti, F., & Bermejo, E. (2010). Developing an electronic placement examination for heritage learners of Spanish: Challenges and payoffs. *Hispania, 93*(2), 273-291.

González Pino, B., & Pino, F. (2000). Serving the heritage speaker across a five-year program. *ADFL Bulletin, 32*(1), 27-35.

Larson, J., & Hendricks, H. (2009). The name assigned to the document by the author. This field may also contain sub-titles, series names, and report numbers.A context-based online diagnostic test of Spanish. *CALICO Journal, 26*(2), 309-323.

MacGregor-Mendoza, P. (2011). Probando, probando...analizando la utilidad de un examen de ubicación para hablantes de español como lengua de herencia. In L.A. Ortiz-López (Ed.), *Selected Proceedings of the 13th Hispanic Linguistics Symposium* (pp.150-160). Somerville, MA: Cascadilla Proceedings Project. Retrieved from http://www.lingref.com (Document #2483).

Moodle Trust. (n.d.). Moodle [online course management system]. Perth, Western Australia: Author. Available from http://www.moodle.com

International Reading Association and National Council of Teachers of English. (1996). *Standards for the English language arts*. Newark, DE; Urbana, IL: Authors.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Otheguy, R., & García, O. (1996). *Prueba de ubicación para hispanohablantes*. Boston: D.C. Heath.

Otheguy, R., & Toro, J. (2000). Tests for Spanish-for-native-speaker classes. *Spanish for Native Speakers: AATSP Professional Development Series, Volume 1*. New York: Harcourt College Publishers.

Parisi, G., & Teschner, R. V. (1983). *PASS: Parisi Assessment System for Spanish*. El Paso, TX: Vargas Printing Co.

Parisi, G., & Teschner, R. V. (1984). Unitary placement testing for Spanish: Can a single instrument adequately serve native speakers and non-natives alike? *Foreign Language Annals, 17*, 173–184.

Potowski, K. (2011). *Conversaciones escritas: Lectura y redacción en contexto* (1st Ed.) Hoboken, NJ: John Wiley & Sons.

Potowski, K., & Carreira, M. (2004). Towards teacher development and national standards for Spanish as a heritage language. *Foreign Language Annals, 37* (3), 421-431.

Questionmark. (n.d.). Questionmark Perception [assessment management software]. Norwalk, CT: Author. Available from http://www.questionmark.com

Short, D. (2000). *The ESL standards: Bridging the academic gap for English language learners*. Center for Applied Linguistics Digest. Retrieved from http://www.cal.org/resources/digest/0013ESLstandards.html

Teschner, R.V. (1983). Spanish placement for native speakers, non-native speakers and others. *ADFL Bulletin, 14*(3), 37-42.

Valdés, G. (1989). Teaching Spanish to Hispanic bilinguals: A look at oral proficiency testing and the proficiency movement. *Hispania, 72*(2), 392-401.

Valdés, G. (1997). The teaching of Spanish to bilingual Spanish-speaking students: Outstanding issues and unanswered questions. In M. C. Colombi and F. X. Alarcón (Eds.), *La enseñanza del español a hispanohablantes* (pp. 8-44). Boston, MA: Houghton Mifflin.

Valdés, G., & R. Figueroa. (1994). *Bilingualism and testing: A special case of bias*. Norwood, NJ: Ablex Publishing Corporation.

Valdés, G., Fishman J.A., Chavez, R.M, & Perez, W. (2006). *Towards the Development of Minority Language Resources: Lessons from the Case of California*. Multilingual Matters: London.

Webb, J., & Miller, B. (Eds.) (2000). *Teaching heritage language learners: Voices from the classroom*. Yonkers, NY: ACTFL.

webCAPE (Computer Adaptive Placement Exam), Spanish. (n.d.). Orem, UT: Perpetual Technology Group. Available from http://www.perpetualworks.com

Ziegler, J. (1981). Guidelines for the construction of Spanish placement examinations for the Spanish-dominant and Spanish-English bilingual. In G. Valdés, A. Lozano, & R. García-Moya (Eds.). *Teaching Spanish to the Hispanic bilingual: Issues, aims and methods* (pp. 211-214). New York: Teachers College Press.

Appendix
Final exam items, tests C and D-HS

Test C: To separate heritage speakers from L2 learners

1. Susana no acabó su trabajo porque le dio…
    a. ganas   b. enojada   c. flojera

2. Cuando Humberto no ganó el premio, le dio mucho…
    a. coraje   b. molesto   c. depresión

3. No le cuentes nada a mi tía porque es muy…
    a. perezosa   b. chismosa   c. callada

4. Esa señora es alta ejecutiva y gana _____ dinero.
    a. pluma   b. lleno   c. harto

5. Ese equipo nos metió….
    a. un peatón   b. una paliza   c. un traste

6. A ese sinvergüenza espero que le den una buena…
    a. empuje   b. cachetada   c. muestra

7. Rosa está emocionada porque esta noche sale con su…
    a. galán   b. alberca   c. relleno

8. El niño anda ___ , hay que bañarlo ya.
    a. fisquito   b. mugrosito   c. Lanudito

9. Ay, qué problemón. Esta vez realmente metiste…
    a. el diente   b. la pata   c. el fallo

10. Me dijeron que acababan ya ___   pero tengo 3 horas esperando.
     a. mero   b. obra   c. leído

11. Mi prima que tiene 8 meses de embarazada ya está bien…
    a. bochada   b. perejil   c. panzona

12. No puedes andar regalando dinero a cualquier ___que te lo pida.
    a. fulano   b. roscado   c. moneda

13. Esa película violenta no es apropiada para…
    a. chamacos   b. público   c. niñeras

14. Estaba estudiando cuando ese ruidazo me sacó de…

a. onda   b. casona   c. Tinta

15. No confíes en esa compañía porque siempre hace movidas…
    a. advertidas   b. chuecas   c. prietas

16. Por todo el hielo en la calle, mi hermano se cayó…
    a. de nalgas   b. de chidos   c. de maníes

17. Mi tío no pudo venir a la fiesta porque le tocaba…
    a. guardar   b. enlistar   c. chambear

18. No seas ___, préstame $20 y te los regreso el martes.
    a. liviano   b. adorado   c. gacho

19. Mi sobrino no camina todavía, apenas aprendió a….
    a. gatear   b. cosechar   c. incubar

20. No hay mal que por bien…
    a. no viene   b. no venga   c. vino

Test D-HS: To place students into Heritage 1 or Heritage 2
(Questions are based on a portion of our Heritage Program content)

1. ¿Qué _____ ella cuando termine de estudiar?
    a. va hacer   b. va ser   c. va a hacer

2. El profesor _____ tarde pero yo no lo ___ .
    a. llegó, espere   b. llego, espere   c. llegó, esperé

3. Bajo mejores circunstancias, lo _____ hecho a tiempo.
    a. vieran   b. hubieran   c. ubieran

4. El jefe _____ las demandas de los trabajadores.
    a. iba considerar   b. hiba considerar   c. iba a considerar

5. Ese plan _____ estudiado en todos sus detalles.
    a. hacido   b. a sido   c. ha sido

6. Ella está soñando con _____ a París con su novio.
    a. viajando   b. viajar   c. viajó

7. En el momento que _____ el líder del sindicato, todos se sentaron a hablar.
    a. llego   b. llegó   c. yego

8. Hay problemas _____ no existen soluciones perfectas.
   a. para las cuales   b. cuyas   c. para los cuales

9. Yo no ____ los consejos de ese infeliz, y tampoco quiero que mi amiga los ____.
   a. tome, considere   b. tomé, considere   c. tome, consideré

10. Esta semana me toca trabajar ___.
    a. Jueves y Viernes   b. jueves y viernes   c. weves y viernes

11. No sé _____ decidieron vender la casa.
    a. porque   b. por qué   c. por que

12. Ella ____ el tren para llegar al trabajo.
    a. ba tomar   b. va tomar   c. va a tomar

13. ¡ _____ baila ese niño!
    a. Como   b. Cómo   c. Comó

14. Hay muchas organizaciones para estudiantes ____ y otras para _____ en las
    universidades.
    a. Latinos, asiáticos   b. latinos, asiáticos   c. Latinos, Asiáticos

15. ____ enfrentan muchos problemas.
    a. Inmigrantes   b. A inmigrantes   c. Los inmigrantes

16. Cuando me pidió ayuda, le dije que ___, quería hacerlo.
    a. si   b. sí   c. sé

17. Estoy seguro que mi abuela me _____ para felicitarme en mi cumpleaños.
    a. llamara   b. yamara   c. llamará

18. Ayer yo ____ con mi profesor acerca de mi nota.
    a. hable   b. able   c. hablé

19. Es increíble que mi abuelo se ___ comprado un Blackberry.
    a. haiga   b. haya   c. alla

20. Mi tío me regaló $100 para que _____ libros.
    a. compraba   b. compré   c. comprara

21. Yo lo habría hecho si me _____ a tiempo.
    a. vieras llamado   b. hubieras llamado   c. habrías llamado

22. Marcos es muy problemático y no quiero entrar en pleitos con ___.
    a. el   b. él   c. ella

23. En realidad, yo no _____ si tu idea es buena.
    a. se   b. sabo   c. sé

24. Mi mamá sabe hablar _____, _____ y _____ .
    a. Inglés, Español y Francés   b. inglés, español y francés   c. Ingles, Español y Francés

25. Mi sobrina _____ inglés desde que nació.
    a. empezó hablar   b. empezó ablar   c. empezó a hablar

26. Sandra nunca _____ tarde a sus clases.
    a. a llegado   b. ha llegado   c. haya llegado

27. Se _____ muriendo mucha gente en la frontera.
    a. esté   b. esta   c. está

28. La autora dijo que se _____ que hacer algo para ayudar.
    a. tenia   b. tenía   c. tenió

29. Un día, mi sueño se _____.
    a. realizara   b. realizará   c. realisara

30. Siempre _____ cosas que nos molestan en la vida.
    a. va a ver   b. va a haber   c. van haber

Notes

1. Teschner (1984, p. 37) uses the term 'native' to refer to all students having grown up using Spanish, including those primarily educated in Spanish who migrated to the US as young adults or adolescents (i.e., first or 1.5 generation individuals), and those brought up in the U.S. largely educated in English (i.e., heritage speakers).

2. Placement instruments grounded in local heritage speaker student populations and particular curricula are in fact desirable, as we will argue ahead. However, designing reliable and valid assessment measures is difficult without the aid of psychometricians.

3. The data reported here is based on the ethnolinguistic group of students' mothers only. There are many students on campus of mixed Latino heritage.

4. That is, a heritage speaker who possesses the linguistic background to complete our heritage program successfully. As we will describe ahead, we occasionally encounter students who self-identify as heritage speakers and take the heritage speaker exam, but their written essay makes it clear that the student would be better served in the BLP, and they are advised accordingly.

5. Our HLP courses focus on academic reading and writing, and cover grammatical points that are often problematic for heritage speakers. Details about what is taught can be found the textbook *Conversaciones escritas* (Potowski, 2011), which was written largely based on knowledge acquired in years of working with students in this program.

6. An exception is the composition course, which continues to be separated into L2 - and HS sections.

7. It should be noted that we have had students who were not Latino successfully complete our heritage language program. They had acquired Spanish in naturalistic contexts, either through caregivers or living abroad for extended periods, and thus demonstrated linguistic profiles of heritage speakers.

8. While oral proficiency constitutes a principal difference between heritage speakers and L2 learners, we wished to avoid having to record and evaluate speech samples from potentially a thousand students per semester. Aural (listening) comprehension might be a fruitful area to examine for separation of these two student types, but the reliability may be questionable given that individuals often comprehend a great deal more than they can produce.

9. Participant consent was not required because the IRB considers this anonymous program data.

10. We have come to find that, while reading a 18-20-line essay, an experienced evaluator can determine placement in approximately 2-3 minutes per exam.