Spanish as a Heritage Language Assessment: Successes, Failures, Lessons Learned

Patricia MacGregor-Mendoza
New Mexico State University

Abstract
From its origins over three decades ago, interest in the field of Spanish as a heritage language (SHL) has grown and has produced a wealth of research. While our understanding of the sociolinguistic profile of Spanish heritage language learners has increased and we have advanced in our knowledge of the linguistic abilities and strategies Spanish heritage language learners bring to bear on specific language tasks, we are just beginning to apply this knowledge in meaningful ways for the purposes of assessment. The present paper describes the evaluation of the efficacy of the Spanish Placement Test (SPT) that has been used for over 15 years to evaluate students initiating their Spanish language study at New Mexico State University (NMSU). The SPT is intended to distinguish between students who would be best served by either the SHL sequence or the Spanish as a Second Language sequence and, further, to suggest which course within the appropriate sequence would best allow their skills to grow. An examination of the SPT was warranted as the population for which the SPT was originally designed did not appear to match that of NMSU's population of SHL learners. Additionally, at first glance, the items on the SPT did not appear to be a good fit with the goals of the courses in the SHL sequence. The present paper discusses the findings of our evaluation of the SPT in light of its ability to assess the skills of learners of Spanish as a heritage language and place them accurately in the sequence of SHL courses.

Introduction
While questions regarding how to teach Spanish to learners who have already acquired knowledge of the language at home had been raised as early as the 1930s, focused interest in Spanish as a heritage language (SHL) did not begin to coalesce into a discipline until the 1970s (Valdés, 1997). Prior to that era, the classroom presence of students who had learned Spanish in the home was largely met with disdain and criticism since teachers regarded the lexical choices, morphology, and syntax of SHL learners to be inferior to those displayed by an ideal(ized) educated speaker from a foreign country, and teachers expended a great deal of effort to polish, correct and eradicate what they perceived to be bad linguistic habits.[1] Thus, according to Spanish practitioners' early views of SHL learners, the assessment of these learners' abilities was neither relevant nor necessary. Rather, the prevailing sentiment revealed that Spanish teachers of that era viewed SHL learners as possessing few linguistic skills of any value, and what skills they might possess were deemed to be imperfect, inaccurate, and in need of forced correction.

One of the scholarly pioneers of the field, Dr. Guadalupe Valdés, countered these assertions. She affirmed that SHL learners arrive at the classroom already bearing a wealth of linguistic and cultural knowledge (Valdés Fallis, 1973) that had been acquired in natural surroundings. Drawing from the goals of English language arts classes for native speakers of English, she

argued that SHL courses should be oriented toward providing SHL learners with literacy skills in Spanish. Valdés further maintained that the Spanish teaching profession was overall unprepared to teach SHL courses due to its long-standing tradition of emphasizing Spanish as a foreign language, its tendency to measure linguistic performance based on a prescriptive (and often subjective) norm, and its ignorance of characteristics of Spanish spoken within the U.S. (Valdés Fallis, 1976, Valdés-Fallis, 1978). Valdés encouraged using the assessment of SHL learners' abilities as the foundation for SHL curriculum development, noted the lack of adequate measures of learner competence, and highlighted the theoretical and practical pitfalls that faced assessment developers:

> To speak of teaching is to speak of levels of instruction, of measured achievement, of adequate placement, and ultimately of effective testing. At this point, however, we are aware that tests which adequately measure communicative or receptive competence are not available. Thus in order to place a student in a section for second dialect learning we might not have the means at our disposal for determining how the individual might operate in specified sociolinguistic situations in the two dialects in question. Nor might we be able to determine the amount of receptive knowledge possessed by a student with regards to the standard dialect. We might be reduced again to the discrete point approach and the perhaps irrelevant contrasts between items which may tell us nothing about the student's actual competence in either the standard or the non-standard dialect. With the Spanish-speaking bilingual this presents an added problem because the use of such instruments presupposes a reading ability in Spanish by the person being examined. Such an ability, as we know, is very frequently not part of the experience of the bilingual Mexican-American (Valdés Fallis 1976, pp. 21-22).

Valdés (1997) continued to challenge the Spanish teaching profession when two decades later, despite the hard-won changes in attitude toward SHL learners that had fostered awareness of their needs and produced new textbooks and teaching practices oriented toward them, efforts still appeared to be unfocused and had not yet leapt beyond the classroom to providing a much needed theoretical foundations for SHL teaching practices. Valdés identified crucial connections between theories of language learning, pedagogical approaches and teaching practices with regard to SHL learners, outlined a series of learning objectives to which SHL programs should aspire, encouraged SHL practitioners to conscientiously ground their pedagogical practices in research, and suggested an agenda for future research that would capture the characteristics of the broad range of bilingual abilities displayed by SHL learners, thus allowing teachers to both identify and expand upon SHL learners' wide-ranging competencies (Valdés 1992, 1997, 2001).

Particularly in the last decade, researchers have risen to Valdés' challenge, enriching the SHL field with research that has provided greater insight into the linguistic skills and sociolinguistic profiles of SHL learners. Because of these efforts we have increased our understanding of SHL

learners' oral production (Achugar, 2003; Fairclough & Mrak, 2003), writing strategies (Schwartz, 2003; Spicer-Escalante, 2005), and have greater knowledge of SHL learners' sociolinguistic backgrounds (Carreira, 2003; Beaudrie & Ducar, 2005; Alarcón, 2010). There have also been a growing number of important contributions regarding various aspects of the grammatical competency of SHL learners (e.g., Montrul, 2002, 2004, 2005, 2006, 2008; Iverson, 2009), and important differences between SHL learners and learners of Spanish as a second language have been identified (Montrul, Foote, & Perpiñán, 2008; Montrul & Perpiñán, 2011; Potowski, Jegerski, & Morgan-Short, 2009; Bowles, 2011). Efforts are underway to discover how SHL learners process input (see Carreira & Potowski, 2011 for summaries). Such enhanced and detailed knowledge will not only allow us to design instruments to better capture the linguistic strengths of the SHL learners entering our programs, it will also aid us in more effectively directing our efforts to meet their linguistic needs (e.g. Montrul & Bowles, 2010; Carreira & Potowski, 2011).

What is still severely lacking in the present literature are sufficient studies on SHL assessment in general, and placement in particular, that meet the needs of SHL learners and SHL programs, as was called for by Valdés some 35 years ago. Fairclough (2006) and her colleagues (Fairclough, Belpoliti & Bermejo, 2010) have laid key groundwork in this area, examining learner responses to linguistic tasks and critically applying knowledge of both SHL learner characteristics and principles of testing design and analysis. The current special issue of the Heritage Language Journal, dedicated to examining the accumulation of knowledge to date regarding Spanish heritage language placement, represents a culmination of cross-disciplinary efforts in sociolinguistics, language acquisition, and assessment. This synthesis of knowledge is of utmost importance to developing assessment practices that are both scientifically sound as well as linguistically and culturally relevant to the population being assessed (Bachman & Palmer, 1996; McNamara, 2000; Hughes, 2003).

The work in parallel fields examining the effects of language and culture on testing of English Language Learners (Solano-Flores & Trumbull, 2003; Trumbull & Solano-Flores, 2011) underscores how a learner's language, culture, and lived experiences combine to influence their responses to items in a testing situation. Moreover, as cautioned by these researchers, inaccurate assumptions that teachers and test developers make either about the language of the test items or the language of the population of individuals being examined can greatly distort outcomes. Highlighting this hidden potential for failure, Trumbull and Solano-Flores (2011), state:

> Although most educators are aware of the fact that assessment depends on language to a large extent, not many are aware of the subtle ways in which this dependence operates. Wording, discourse style, and the use of idiomatic expressions and the structural complexity of sentences are among the many aspects of language that may increase the "linguistic demands" of assessment tasks and test items. These linguistic demands shape the ways in which students understand and interpret assessments and the ways in which they respond to them. … The first step toward more valid and equitable testing

practice is for educators and test developers to recognize how irrelevant
language demands may wrongly affect the scores obtained by students on tests
or their responses to questions in classroom assessment activities. (pp. 40-41)

Thus, both knowledge of the community of learners and sensitivity to the language used in the creation of the items is critical for the development of meaningful and valid measures of language ability. With regard to measures assessing ability for the purposes of placement, it is therefore vital that these measures be developed locally, reflect the cultural and linguistic characteristics of the local variety of spoken Spanish, serve the needs of the community of SHL learners, as well as respond to the institution's curricular requirements. In the case of NMSU, we have deviated from this model of practice over the years.

Background of NMSU and our Use of the SPT
The cultural and linguistic characteristics of NMSU's SHL population are varied due to the diversity of linguistic contexts in which SHL learners live. Mirroring Valdés' (2001) broad definition, the SHL learners in our area are individuals who were raised in a home where Spanish was spoken, speak Spanish themselves or may only be able to understand it, and are to some extent bilingual in Spanish and English. Most SHL learners that populate NMSU's SHL courses come primarily from both urban and rural locations in Southern New Mexico, although some may have arrived from other areas of the state as well as from Mexico. NMSU's home county of Doña Ana consists of 206,419 residents, over 65% of whom are Latino in origin (U.S. Census Bureau, 2011). Many can trace their family roots back prior to the Treaty of Guadalupe Hidalgo, while others are more recently arrived (Bills, Hernández Chávez, & Hudson, 1995). With some exceptions, SHL learners in our region have acquired their language skills orally in a home environment and normally have not had formal instruction in or about the Spanish language. Their linguistic abilities in Spanish have been nurtured primarily through family gatherings, cultural celebrations, music, *telenovelas*, and movie marathons featuring classic iconic figures of Mexican cinema.

The members of the local SHL learner population that retain a higher degree of fluency in Spanish are typically residents of Southern New Mexico's *colonias (settlements)*. These small, unincorporated villages appear alongside the highways and are home to primarily Latinos of differing immigrational generation and status. Due to their similarity in culture, language, and background, and because of the limited economic and educational resources available to them, these communities develop a web of tight-knit relationships (Norman, Parcher & Lam, 2004).

Previous research has pointed to a number of factors that encourage the maintenance of a minority language or, inversely, promote the shift to a majority language. In general, the retention of minority languages is encouraged in communities of speakers that share ties to a similar homeland, that settle in close proximity to one another either in rural environments or cultural enclaves in urban environments, and that demonstrate similarity in their educational, religious, and economic characteristics. Contrastively, residents that settle in more diverse urban areas often have a higher probability of interpersonal contact with a wider variety of

interlocutors, and may experience more pressure to assimilate to a majority language and lifestyle, and thus exhibit a greater tendency to shift to a majority language (see Milroy & Milroy, 1997; Romaine, 1995). Thus, the ethnically dense population of Latinos in the *colonias*, combined with the rural setting and limited physical, educational, and economic mobility that many share, provide favorable conditions for the maintenance of Spanish amongst *colonia* residents.

In contrast to the SHL learners in our area that come from the *colonias*, those that originate from the 88,732 residents of the city of Las Cruces (U.S. Census Bureau, 2011), as a group exhibit less overall facility with Spanish due to their more frequent opportunities for contact with English speakers than their peers in the rural *colonias*. Nonetheless, for all SHL learners, their exposure to and facility with literacy skills in Spanish varies, but is considerably less than their exposure to Spanish in oral form (MacGregor-Mendoza 1998, 2000, 2005).

The Spanish abilities of all SHL learners are often subject to societal scrutiny. The Spanish spoken by SHL learners, particularly those who demonstrate lesser facility with pronunciation and/or a tendency to incorporate English words or calques, is often criticized as "pocho" or "mocho" (Galindo, 1996; Carrasco & Rigelhaupt, 2003). As such, many SHL learners have internalized the criticism they've heard from family and friends over the years and as a result hold negative attitudes about and lack confidence in their abilities, a topic which needs to be addressed in the SHL classroom (Beaudrie & Ducar, 2005; Correa, 2011).

Given the diversity of linguistic experiences and, by extension, linguistic skills and abilities, the curricular challenges faced by the Spanish as a Heritage Language Program at NMSU is multifaceted. It is necessary to: 1) distinguish between students who have acquired Spanish through family heritage either in the US or abroad from those who learned the language solely in a US classroom setting; 2) identify the skills and abilities of the students in both categories; and 3) accurately place students in courses within our program of study in accordance with the skills that students already demonstrate.

In order to accomplish these tasks, since 1993 NMSU has employed the SPT which was originally developed by Teschner and his colleagues at a neighboring institution, the University of Texas at El Paso (UTEP) about 40 miles to our south (Rodríguez Pino & Villa, 1994). While in geographic proximity, the two campuses serve distinct populations. El Paso, a city of more than 600,000 residents, sits across the border from Ciudad Juárez, México, its sister city of nearly 1.5 million (Consejo Nacional de Población, n.d.). As Teschner (1995) explains, the cultural and economic ties shared by these two large metropolises and the constant flow of tourists, workers and students back and forth between them serves to continually reinforce and revitalize the presence and use of Spanish throughout El Paso. Teschner (1995) estimated at the time his article was written that more than 80% of El Paso residents spoke Spanish at home, and at least that amount exercise their oral skills in both Spanish and English on a regular basis outside of the home. Moreover, it is common for individuals to live on the Juárez side of the border and cross to attend a public school or UTEP; it is also normal for professionals to live in

El Paso and cross the border to work in Juárez (Witkowsky, 2003). This constant movement across the border continually strengthens the linguistic and cultural connections between the two cities. These connections are perhaps even more pronounced at present, given that the persistent violence occurring in Juárez prompted nearly 56,000 *juarenses* to leave that city and relocate across the border in El Paso between 2007 and 2009 (Velázquez Vargas & Martínez Toyes, 2010).

Study
The differences in lifestyle and, by extension, linguistic experiences of the SHL populations of NMSU and UTEP were part of the motivations for the current examination of the SPT. We question whether the content of a test originally designed for a population more than 20 years ago and 40 miles away, bears relevance for the local SHL learner population and whether or not the results of the SPT render information of value for the teachers of SHL learners. An additional motivation concerned the test's content since the items as a whole did not appear to represent the curricular objectives of the courses into which the SPT was placing SHL learners. In a previous paper (MacGregor-Mendoza, 2011), the SPT was examined from a design standpoint, and its structure and content critically evaluated against the criteria for validity as established by the most recent edition of the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1999, [hereafter AERA, APA & NCME, 1999]). A more thorough description of the SPT, its history, adoption and formatting modifications made at NMSU, and of the flaws related to validity concerns are also discussed in greater detail are provided in MacGregor-Mendoza (2011).

In the present paper, we will focus our attention on the value of the SPT in fulfilling the three responsibilities listed above and summarized below, with particular regard to SHL learners. Overall, in the present paper we seek answers regarding the most basic functions for which the test was designed. In short, we ask:

1. How well are we doing in terms of distinguishing between our SHL learner and non-SHL learner populations?
2. How well are we able to identify the relevant skills of our SHL learners?
3. How well are we placing students into our SHL programs?

In order to answer these questions, we examine the results collected from 4,764 takers of the electronic SPT over a 12-month period. It should be cautioned, however, that these data are unfiltered. That is to say, given that this test is available online, individuals taking the test may or may not be doing so in order to obtain an estimate of their placement in our program.

Results
Identifying Groups of Learners
First and foremost, a central goal of the SPT is to adequately differentiate SHL learners from non-SHL learners, so that students with different experiences in learning Spanish can be placed

in courses dedicated to meeting their needs. Working with Valdés' (2001) definition of SHL learners mentioned previously, our aim is to distinguish learners who learned Spanish primarily in informal home environments from those who have no previous knowledge of Spanish or whose knowledge of the language was initiated exclusively in the classroom. Such a task should be undertaken by first examining the context of the learners' linguistic experiences.

Teschner (1990) notes that originally the SPT was administered to all UTEP students wishing to enter Spanish programs *unless* either they were an L2 speaker and affirmed that they either had no previous (or recent) formal study of Spanish *or* they were an SHL learner that had not yet developed literacy skills in Spanish. Students exempted from testing were assigned to appropriate courses within the L2 or SHL sequences on the basis of their answers to a series of background questions. For those that did take the SPT, Teschner relied on 10 "Native Speaker Indicators" embedded in the test to distinguish between the two groups of learners. These "Native Speaker Indicators" represent colloquial terms or expressions from a Mexican-influenced variety of Spanish (e.g., "greñas" ("hair"); "agüitada" ("bummed out"); "no se me pegan" ("I just can't remember them")) that students would be unlikely to have acquired in a classroom setting. Even Teschner, however, recognized that these items on their own were insufficient to classify students who scored six items correctly and thus were considered "borderline" (1990, p. 817). Teschner reports that personal interviews were used to finalize classification in borderline cases.

While NMSU previously administered the SPT in a similar fashion, we have departed from the original guidelines. The SPT was initially administered at NMSU in pencil-and-paper format from 1993 to 2005 by faculty teaching Spanish and followed up with personal interviews in "borderline" cases. However, the administration of the exam was transferred to an online format in 2005 and is no longer under the control of the language teaching faculty. In the current automatized administration of the SPT, the Native Speaker Indicators serve as gatekeepers. Test takers that correctly answer four or more of these 10 items are identified as potential SHL learners and are directed the remaining 90 items of the original test. Correctly answering three or less of the Native Speaker Indicators directs the examinee to a series of items consisting of the first 40 elementary items of the original test. The decisions both to use the Native Speaker Indicators as gatekeepers and to give some students a 40-question test and others a 90-question test was made without the knowledge of faculty teaching Spanish.

Of the 4,764 individuals that took the SPT in our data set, 1,610, or 33%, were identified by means of the Native Speaker Indicators as individuals who had acquired Spanish outside a formal classroom setting. This result under-represents the Spanish-speaking linguistic background of the overall population from which many NMSU students originate. There are several possible explanations of this result: 1) fewer SHL learners are accessing the test than non-SHL learners, which would mean that they are either not aware of the test or are opting not to take it; 2) the population of students wanting to take Spanish courses is skewed away from SHL learners, by natural means (e.g., more L2 speakers want to take Spanish) or artificial means (e.g., advisors are not encouraging SHL learners to take the test). A third alternative is that the

students taking the SPT are SHL learners and non-SHL learners in proportions that mirror the demographics of the population at large, and the 10 Native Speaker Indicators are not providing an accurate initial assessment of these students' abilities.

As correctly suspected by Teschner (1990), the sole reliance on a 10-item pre-test for the determination of abilities is questionable for a number of reasons. First, given the limited literacy skills that we already anticipate from our SHL learner population, having this pre-test delivered in a text-based format would seem to already set up for failure our attempts to capture SHL learners whose skills may be what Valdés Fallis (1976) identified as receptive. Second, although the background of many of the SHL learners entering our courses is based in Mexico, there is no objective, empirical evidence that these specific expressions would, either individually or collectively, serve as accurate identifiers of SHL membership. Third, given that these items are, in the SPT's current administration, deployed and employed without taking into consideration any accompanying background information that would have more relevant bearing on the determination of SHL membership, the opportunity for measurement error is increased. Fourth, as the opportunity to personally interview students was eliminated in the transition to an online delivery of the SPT, we have less of an independent means of accurately placing SHL learners, or of confirming either the utility of the Native Speaker Indicators overall or of whether the lowered threshold of four Native Speaker Indicators that is being applied to identify SHL learners is valid.

Lastly, the lax security associated with the current administration of SPT (anyone in the world can take it at any time, repeatedly, anonymously, and even on behalf of someone else) and the fact that it is sometimes delivered in group fashion (during advising fairs or occasionally as whole class cohorts at the beginning of the semester) raise additional concerns. Given that the first 10 items provide a gateway to an easier or harder test, more savvy students might discover from repeated iterations or from others' experiences that if they answer these questions incorrectly on purpose they could potentially be recommended for a higher level course in the non-native track by being directed to the 50-item non-native test and thus answer fewer and easier items. Each of the above noted concerns indicate that the use of the 10 Native Speaker Indicators for the purposes of identifying SHL membership is a questionable practice from a procedural, cultural, statistical, and ethical point of view. Moreover, as using the Indicators in this fashion runs the risk of under-identifying the population of students who could benefit from SHL courses, our ability to accurately distinguish SHL learners from non-SHL learners is compromised.

Identifying SHL Learners' Skills
The second question we ask regarding the use of the SPT follows from the first. That is, once we have identified potential candidates, are we able to accurately place SHL learners identified by the pre-test into our courses? In order to answer that question, we need to begin by first examining what the course-level expectations are as well as the content of the questions. Table A1 (see Appendix) provides a brief description of the basic courses in the SHL program. As can be seen, these objectives clearly reflect the linguistic characteristics of the population that resides

in southern New Mexico. For example, the listening comprehension skills of the SHL learners in our area are expected to exceed those of their abilities in speaking and literacy, at least for the first three courses in the sequence. Thus, students entering Spanish 113 (Beginning Spanish for Native Speakers) are expected to have little to no formal grammatical knowledge or literacy skills in Spanish and may be reluctant or unable to engage in or sustain conversations entirely in Spanish. To enter into the higher level courses in the sequence, SHL learners are expected to have had previous academic exposure to Spanish. This exposure may be from a family member teaching them to read and write, previous experiences in U.S. bilingual programs, previous schooling in Mexico (usually limited to primary school) or no more than two years of Spanish at the high school level (students with three years of study at the high school level may begin coursework at the 300 level), or previous completion of Spanish 113.

Students entering Spanish 213 (Spanish for Native Speakers I)   are expected to have some grammatical knowledge, nascent abilities in reading and writing in Spanish, and vocabulary that may still be limited in scope and register. Students entering Spanish 214 are expected to demonstrate existing but still developing abilities in listening, speaking, reading and writing, and an expanding range of vocabulary. Spanish 214 (Spanish for Native Speakers II) is the last in the basic language sequence that is required for some majors. Students who plan to take -courses beyond the 214 level may be acquiring additional skills for personal reasons or are working on a major or minor in the language.

Our program does continue beyond the basic sequence of coursework to provide separate options for SHL learners and L2 speakers in our grammar review courses. Students entering Spanish 312 (Grammar for Native Speakers of Spanish) are expected to demonstrate a solid foundation of skills in all four areas; they have knowledge of grammar and vocabulary that allows them to engage in complex activities and interactions to further their linguistic skills. Beyond the 312 level, all students interested in furthering their studies take a composition course. However, in Spanish 312 and other higher-level courses, the curricular division between SHL learners and L2 speakers fades as the curriculum is increasingly focused on themes related to culture, literature, and linguistics (see Table A1 in the Appendix).

The responses of the 1,610 SPT examinees identified as potential SHL learners and the recommendations generated by the resulting scores provide a rough overview of the SPT's ability to place students. We would expect that the recommendations generated would suggest that, based on prior experience in placing students by interviews, that a roughly equal number of students be placed into each of the various course levels. However, a chi-square analysis, performed to check the distribution of placement recommendations for this group of test takers, is skewed toward learners with advanced skills. Of the 1,610 test takers, 324 were placed in Spanish 113, 296 were placed in Spanish 213 and 798 and 192 were placed in Spanish 214 and Spanish 312 respectively. The outcome of the analysis was significant, $(3, N = 1610) = 542.2$, $p < .0001$, indicating that the frequency of placement in each level was not equal and was not due to chance. Given the problematic issues previously identified with respect to the administration of the test (i.e. lack of personal interviews, lack of consideration of personal background

information, lack of security regarding the test) the assumption that the distribution of placement scores is due to the accuracy of the test items correctly identifying skills of potential learners is questionable.

Further questions are raised when we examine the remaining 90 items on the SPT that the 1,610 examinees were required to complete. The items are all multiple choice, with the number of options ranging from two to five, and assess the following lexical and grammatical features (see Table A2 in the Appendix for a sample item from each category):

1-40   basic grammar (*ser*, *estar*, subject-verb and adjective-noun agreement, prepositions, *gustar*, command forms)
41-50   basic vocabulary (simple nouns and verbs)
51-70   intermediate vocabulary
71-80   orthography (focusing on **s** and **z**, **g** and **j, b** and **v**, etc.)
81-90   formal grammar (identifying syntactic features)

An overall examination of each item's difficulty points to additional weaknesses in the test. The item difficulty, or p-value, represents the proportion of test takers that answer a question successfully. A very high p-value indicates that a high proportion of test takers were able to answer a question correctly. For example, items with a p-value of .85 indicate that the information being tested by that item is so widely known that 85% of the test takers were able to answer it correctly. This result then would diminish the utility of the item for the purposes of placement since its ability to distinguish between levels of ability is limited (Kehoe, 1995a, 1995b). At the other extreme, items with a low p-value indicate that the information being tested in the item is successfully answered by only a select few test takers, indicating that the question is very difficult, perhaps excessively so, which also is of limited value for the purposes of placement.

Ideally, p-values seek to determine how well an item can differentiate between test takers of varying abilities. The optimal p-value indicates that the item's ability to discriminate between different levels of test takers is at its statistical maximum, striking a balance between how likely a test taker is to answer an item correctly by chance (i.e., guessing) versus how likely s/he is to answer the item incorrectly. For an item with five options, the test takers would have a 1/5 or .20 probability of selecting the correct answer by chance and a 4/5 or .80 probability of failing in their attempt. Thus, the optimal p-value for an item is calculated by dividing the probability of answering an item incorrectly by two, and adding to that figure the probability of selecting the correct option (Lord, 1952). For an item with five options, the optimal p-value is .60 given that $(.80 \div 2) + .20 = .60$. Using the same formula, an item with four options would have an optimal p-value of .625 (often rounded to .63) and for an item with two options, the optimal p-value would be .75.

Nevertheless, item difficulty alone cannot be used to determine whether or not an individual test item or a test overall is valid or acceptable. Most issues of validity should be addressed during

the design phase of a test (Bachman & Palmer, 1996; AERA, APA & NCME, 1999; McNamara, 2000; Hughes, 2003); statistics such as the p-value are initial tools that test designers can use to gauge whether an item is meeting the designer's expectations of usefulness toward the goal of the instrument. Thus, in this examination of the SPT, we are using an overview of the item difficulty as a guideline to determine whether we observe a necessary progression in difficulty that would match the increasing skill levels of the test takers.

A summary of the ranges of p-values for the items is presented in Table 1. From the first column of this table we can observe that 23 items, derived from the basic grammar and vocabulary sections, can be eliminated because they are too transparent to be of use for determining skill levels among the local SHL population. At the other extreme, 35 items in the last two columns exceed the optimal range for p-values. The 21 items in the last category are good candidates for elimination given that they appear to be exceedingly difficult and therefore are just as ineffective in determining ability levels as the excessively transparent items. The 14 items in the second to last column should undergo elimination or modifications to improve their usefulness. From this global observation, we can clearly identify over 50% of the items which are problematic in assessing the skills of the SHL learners in our area and are thus of limited value in accomplishing their intended task.

Table 1

Frequency of Items within Given Ranges of Item Difficulty

		Item Difficulty				
		.85-.99	.75-.84	.60-.74	.45-.59	< .44
1-40	basic grammar	14	8	8	5	5
41-50	basic vocabulary	9	1			
51-70	Intermed. & adv. vocabulary		2	5	6	7
71-80	orthography		4	1	3	2
81-90	formal grammar		1	2		7

In addition to examining item difficulty, a complementary global measure useful in assessments that are designed to distinguish between levels of ability is the discrimination index (D). The discrimination index is used as a guideline to assess an item's discriminatory power. The D value of an item denotes the item's ability to distinguish between high scorers (the upper 27% of test takers) and low scorers (the lower 27% of test takers). The D value is calculated for each item by first subtracting the count of correct answers of the low scorers from the count of correct answers from the high scorers. That result is then divided by the sum of the total number of high and low test takers divided by 2. For example, the 1,610 test takers had 435 high scorers and 435 low scorers. On item 1, a total of 415 high scorers and a total of 384 low scorers chose the correct

answer for that item. The D value for item 1 is thus calculated as (415-384)/((435+435)/2) which renders a result of .07, indicating that students across all ability levels scored similarly on the item, thus it is of little value in distinguishing between students of different levels of ability.

Ebel and Frisbie (1986) offer guidelines for the use of D values. These researchers advise that items with D values lower than zero should be eliminated from the test since they indicate that low scorers were more likely than high scorers to choose the correct answer. Likewise, Ebel and Frisbie recommend that items with a D value of less than .20 either be discarded or subjected to intense scrutiny, as they provide limited ability to determine differences among learners, even those at the opposite ends of the skill range.

As can be observed in Table 2, all test items fall in the bottom two categories of Ebel and Frisbie's (1986) range of D values. This result further suggests that the use of these items in a placement test for the current population of SHL learners bears closer examination if not full elimination. Overall, no item on the test revealed a D value higher than .11, adding to the concerns that the items do not adequately meet the challenge of distinguishing levels of ability among the current SHL learner population.

Table 2

Frequency Items within Given Ranges of Discrimination Index Ranking by Question Category

Category	Excellent >.39	Good .30-.39	Mediocre .20-.29	Poor .00-.20	Worse <.00
1-40 basic grammar				38	2
41-50 basic vocab.				10	
51-70 Intermed. & adv. vocab				11	9
71-80 orthography				6	4
81-90 formal grammar				6	4

Examining a few representative questions and observing patterns of responses made by SHL learners allows us to further explore the factors underlying the difficulty of each item. Table 3 presents a summary of some test items, including the options provided for each item. The expected correct answer is indicated by an asterisk. For each item, the p and D values are provided as well as a count (N) of the test takers that selected each option or omitted the question altogether.

Table 3

Summary of Representative Test Items, Their p and D Values, And the Frequency of Test Takers That Selected the Given Options for Each Item

Item	Option A	Option B	Option C	Option D	Omitted
1. Ella y yo __ español *She and I __ Spanish.* (p .93; D .07)	hablas *(you) speak* N= 117	hablo *(I) speak* N=18	hablan *(they) speak* N=24	*hablamos *(we) speak* N=1507	N=44
11. ¿Ellos los conocen a ustedes? *Do they know you (pl)?* (p .38; D .03)	*nos conocen *they know us* N=616	los conocen *they know them* N=239	nos conocemos *we know each other* **N=505**	los conocemos *we know them* N=195	N=55
31. ¿Hay alguna persona aquí que __? *Is there anyone here who ___? (Formally, this structure requires the next verb to be in a subjunctive form.* (p .32; D .05)	puede ayudarnos *can (Pres. Indic) help us* N=675	llegó antes de la una *arrived (Pret. Indic.) before one?* N=204	*sepa italiano *knows (Pres. Subj) italian* N=526	ha visto el accidente *has seen (Pres. Perf. Indic) the accident* N=147	N=58
77. Ella no es amiga de __. *She's not the friend of ___* (p .58; D .00)	*nadie *anybody* N=943	nadien *anybody (colloquial variant)* N=623			N=44

In examining these questions further, we discover several issues. First, many of the first 40 items, including item 1., examining basic grammar, such as the use of *ser, estar*, *gustar*, simple subject-verb and noun-adjective agreement, and simple vocabulary; none of these aid us in discerning levels of ability for a population who acquired Spanish in an authentic setting. This finding is consistent with Montrul (2004), who found evidence that SHL learners exhibit a high mastery of many grammatical features that are early acquired by monolingual Spanish speakers such as subject-verb agreement and the use of clitics. Gender assignment, also early acquired, has likewise been observed to be less accurate than in monolingual Spanish speakers, but is relatively stable in SHLs' spontaneous oral production, although SHL learner performance diminishes on written tasks (Montrul, 2005). Second, we see other items like 11 and 31 that privilege prescriptive grammatical interpretations and do not place value on the broad scope of

socio-pragmatic linguistic skills that SHL learners display. SHL learners' preference for the pragmatically correct *nos conocemos* (we know each other) option in Item 11 competes with the grammatically expected, yet less comfortable, form *nos conocen* (they know us). This situation is also reflected in their choice of the more plausible option *puede ayudarnos* (can help us) in item 31 rather than the highly implausible, yet grammatically correct, *sepa italiano* (knows Italian). Likewise, as can be seen in Item 77, test takers favor both the formal option *nadie* (nobody/anybody) and the colloquial variant *nadien* (nobody/anybody), reflecting the high frequency of both terms in the speech of the local SHL community; their means of distinguishing between the two options is limited given the orally based, naturalistic manner in which they acquired their skills as well as the overall lack of exposure to formal grammatical norms.

Lastly, these and other questions exhibit additional flaws in their design such as an inconsistency in the number of answer options provided (they range arbitrarily from two to five), the inclusion of multiple grammatical concepts in the same question that confound the intention of the item, question stems that are overtly misleading, orthographic errors in the answer options, and so on. (see MacGregor-Mendoza, 2011). In all, these issues and others call into question the value of the SPT as a placement instrument for our SHL learner population.

Considering all of the above evidence aids us in answering our second question, which sought confirmation of the SPT's ability to discriminate between SHL learners of different abilities and successfully match those abilities with courses in our SHL curriculum. Clearly, the test does not receive a passing grade. Overall, the response patterns of the 1,610 SHL learners examined here indicate that items on the SPT have limited to no ability to accurately assess the knowledge and skills of our local SHL population. As such, we could not find evidence that the items demonstrate a progression of abilities such that students' differing skill levels could be discerned for the purposes of placement.

Additionally, a brief examination of the content of some items reveals several items that are either inappropriate for or inapplicable to the SHL learner population in our area. Moreover, rather than adequately explore the linguistic abilities that SHL learners possess, the SPT instead holds them accountable for prescriptive knowledge they would not normally have access to. Even more disconcerting is that some items appear to penalize SHL learners for their naturally acquired linguistic knowledge and inability to discern the educated standard from forms commonly used in the local SHL community. In the end, the items included in the test do not adequately detect the expected funds of knowledge (Moll, Amanti, Neff and González, 1992) that the different course levels in our SHL program seek to build upon.

Accuracy in Placement
The answer to our third question, regarding how well we are doing in placing our SHL learners, should by now be evident. For our SHL learner population, the SPT is working poorly at best and, more likely, not at all. This is a disturbing result as it implies that not only are we not aiding our SHL learners in confirming their skills and finding an appropriate place in our program in which they can grow, we are likely doing them harm.

Instructors of SHL learners readily acknowledge that the challenges in teaching Spanish to SHL learners have as much to do with the nuances of the language as they do with changing the deficit mindset that many SHL learners have internalized regarding their own competencies (Galindo, 1996; Carrasco & Rigelhaupt, 2003; Beaudrie & Ducar, 2005; Correa, 2011). Having SHL learners undergo a lengthy, cognitively challenging, perhaps unduly uncomfortable process that puts their linguistic skills under formal scrutiny may negatively impact their performance due to the language demands of the task (Trumbull & Solano-Flores, 2011). Additionally, because many items on the SPT are narrowly focused on the display of formal grammar and vocabulary knowledge, SHL learners will find little confirmation of their linguistic knowledge, which may serve to further undermine their linguistic sense of self. Lastly, given that an ethical principle underlying any scientific activity involving human beings is to commit to keeping the participants free from injury or mistreatment, whether it be physical or psychological, requiring SHL learners to complete a test that in the end has no bearing on a suitable course of study is unethical.

As outlined above, the problems concerning the lack of usefulness of the SPT stem from the mismatch between the linguistic expectations underlying the content of the instrument, the linguistic practices of our local SHL learner population, and the lack of alignment between the SPT and our SHL programmatic goals. Additional concerns include various design problems associated with the composition of several items as well as administrative issues related to the removal of language teaching faculty from the supervision, administration of the SPT, and the follow up of the results.

Recommendations and Final Thoughts
Our findings here have outlined the steps used to place SHL learners in our sequence of courses, but we have also uncovered how easily the process can get off track and how even the best of intentions can fall short of their promise. From the analysis conducted here as well as the growing body of research exploring the foundation of SHL learners' linguistic skills, we offer the following suggestions to individuals intending to develop or adopt placement tests for heritage language learners (HLLs) in general.

*1.   **Placement measures applied to populations of HLLs need to be home grown.***
That is to say, the instrument needs to be developed with local interests in mind. Placement measures serve the practical function of matching the varied characteristics of a student group with the appropriate place in a known academic context. Therefore, the item content should reflect, from a linguistic standpoint, what is typical of the population most likely to take the test. This means that the language of the items needs to mirror and be authentic to the local context of which the HLLs form a part. Items written in language that appears to be contrived solely to illustrate formal elements of grammar may prove inadequate for tapping into the linguistic skills appropriate for the placement of HLLs.

In addition to understanding the current context of HLLs, we must consider the manner in which their linguistic skills were likely acquired. Given HLLs' naturalistic experience of acquisition, inherent pragmatic principles will likely guide their responses more than explicit principles of grammar. Consequently, test items should not ignore the pragmatic notions that will likely influence HLLs' responses nor, for the purposes of placement, put pragmatic-based choices in opposition to syntactic-based choices among the answers. Similarly, placement tests shouldn't simply hold HLLs accountable for a series of grammatical norms that likely have not been part of their experiences prior to taking the test, since doing so ignores and invalidates the skills they do have. Finally, in addition to concerns of linguistic authenticity, the items should be able to effectively match the linguistic skills displayed by the test takers with the expectations of the program's different levels so that the test may fulfill its goal of accurately placing students in an the appropriate course of study.

2. *HLL placement measures need to be informed by current research.*
Given that the purpose of a placement is to discern levels of skill, items created for that task need to focus on areas that show variation across HLLs. Accordingly, test developers need to avoid including items that display aspects of grammar that are early acquired and thus stable across a wide variety of HLLs, and instead focus on areas where the evidence in the research on HLLs has indicated greater variability.

For SHL learners in particular, we can use what scholars have begun to observe about their skills and abilities as a guide for developing items for placement that have the potential for being more meaningful. For example, we can avoid including items such as subject-verb agreement (Montrul, 2004; Iverson, 2009), the use of accusative and dative clitics (Montrul, 2004), as well as the interpretation of unaccusative and unergative verbs (Montrul, 2005, 2006). Acquisition of these items is less dependent on having received consistent input and many SHL learners demonstrate stability in producing and interpreting these forms from an early age (Montrul, 2010). Similarly, given SHL learners' lesser command of metalinguistic skills in comparison to L2 learners (Bowles, 2011), and the resistance SHL learners have demonstrated to mastering certain orthographic skills even after direct instruction (Beaudrie, 2007), formal writing tasks may be of limited usefulness for SHL learner placement. Contrastively, semantic interpretations of the preterit, imperfect, and subjunctive which are later acquired appear to be more promising as they demonstrate more variability in mastery, possibly the result of the linguistic conditions under which SHL learners learn Spanish (Montrul, 2009, 2011; Potowski, Jegerski & Morgan-Short, 2009; Montrul & Perpiñán, 2011). Similarly, SHL learners' gender agreement skills have been shown to vary according both to proficiency level and whether the task required a written or oral response (Montrul, Foote, & Perpiñán, 2008), thus examining this area may also yield benefits for placement. Overall, the items included in a placement test for any population of HLLs need to be guided by current research in order to effectively evaluate the HLLs' skills in the most efficient way possible (see also Fairclough, this issue).

*3. **HLL placement measures need to be mindful of the linguistic and cultural issues at play that can confound results.***
Our linguistic and cultural experiences shape the way we approach linguistic tasks, and any perceived incongruity between prior experiences and the task at hand increases the linguistic demands of the situation and may negatively affect our performance (Trumbull & Solano-Flores, 2011). For HLLs, who have frequently learned their heritage language orally and often have limited experience with its written form, taking a test in a text-based format already raises the bar and may, by itself, have a negative impact. Focusing on linguistic knowledge that is largely inaccessible, confusing, or inapplicable to the local HLL population only serves to increase this cognitive burden.

In order to obtain information that more directly pertains to the linguistic knowledge of HLLs, placement tests should attempt to minimize the extraneous effects of the testing condition. As suggested above, using authentic language and focusing in on areas that will efficiently demonstrate variability are important first steps; however, they are not enough. Efforts need to be made to incorporate items that better capture the oral, aural, and pragmatic knowledge possessed by HLLs to more adequately and accurately assess their skills for the purposes of placement. Text-only test items are largely inconsistent with HLLs' acquisition of their skills and therefore are inadequate as sole measures of their abilities. Given the broad accessibility of computers and software, test items can easily be developed to integrate visual and spoken stimuli, and evoke non-verbal responses such as touching a screen to select an answer.

Moreover, because of the complexity and diversity of experiences and skills observed in HLL populations, it is unlikely that any one test will be adequate to accurately place HLLs in a sequence of courses. Thus, while some value may be placed on knowing whether an SHL learner has knowledge that *nadie* is more broadly accepted by a wider variety of speakers than the *nadien* form, which might also be heard in the SHL community, that knowledge alone is insufficient to determine where an SHL learner should be placed. Instead, HLL placement must be complemented by a background survey that reveals details regarding the context of the HLLs' acquisition of the heritage language (Fairclough, Belpolti, & Bermejo, 2010) as well as face-to-face interviews (Teschner, 1990). Thus, HLL placement must be undertaken from a comprehensive approach in order to ensure a more precise system of placement.

In sum, an initial assessment of HLLs' skills should seek to mitigate the effects of the testing conditions by integrating elements with a familiar look and feel. Test items should replicate a variety of contexts and should incorporate situations and questions that reflect more of the familiar, natural activities that HLLs are used to and should represent less of a conventional testing regimen. Doing so more closely imitates elements of the HLLs' customary environment, and may help reduce the stress involved in testing and more directly access HLLs' linguistic knowledge.

*4. **HLL placement measures need to be grounded in established test development principles.***
Tests are often developed based on their author's personal knowledge, experiences, and expectations (Cizek, Fitzgerald, & Rachor, 1995, Bachman & Palmer, 1996; McMillan, 2001) and, once administered and graded, are rarely examined for errors, including ambiguities of the questions, miscues, and so on that could have a bearing on outcomes (McMillan, 2003). Moreover, tests that are derived from external, often commercial, sources are treated as "sacrosanct" and are rarely questioned as potential sources of error (Fairbairn & Fox, 2009).

Any test that purports to evaluate the skills of students and uses these results to render a decision about the academic path students should take should warrant close scrutiny. No test should be adopted on blind faith without knowledge of how that test was put together, the well-formedness of the test items, and their applicability to the local population of HLLs. Even though the SPT we have been using for nearly two decades was developed by knowledgeable scholars with similar intentions and within close proximity to our location, when subjected to closer inspection, it was revealed that the test was not useful in determining meaningful differences in our population of SHL learners. Thus, given the diversity of HLLs' linguistic experiences, we can't assume that a one-size-fits-all test will be suitable to meet the needs of the local HLL population or program.

Broadly accepted standards for the validity of assessments (AERA, APA & NCME, 1999) as well as the principles outlined in literature on language testing (Bachman & Palmer, 1996; McNamara, 2000; Hughes, 2003) should be used as guidelines for either developing an initial assessment or evaluating an existing one. Tests should be designed conscientiously to identify the skills necessary to evaluate learners, establish distinct levels of skills to facilitate placement, and verify the appropriate number of questions needed to determine these skills. The procedures are time-consuming, but are worthwhile endeavors as their practice yields an assessment that has greater potential for accomplishing the purposes for which it is intended without subjecting HLLs to undue harm (Fairclough, 2006; Fairclough, Belpoliti, & Bermejo, 2010).

*5. **HLL placement measures need to be administered responsibly.***
Accepting responsibility for designing or administering a test assumes an ethical obligation for monitoring the content and outcomes to ensure the test is fair and accurate (AERA, APA & NCME, 1999) and that the consequences of taking the test do not negatively impact students. Additionally, any placement assessment needs to be integrated into an infrastructure that more closely and continually monitors the outcomes of HLLs and their accurate placement to confirm the test's reliability and efficacy. Therefore, any placement test needs to be accompanied by regular and rigorous oversight by persons knowledgeable about the subject matter, experienced in assessment methods, and skilled in evaluating the effectiveness of the items and measuring the impact of any modifications in the manner in which the assessment is administered.

In addition to the validity concerns we uncovered concerning the SPT's design as well as the lack of fit between the SPT and our program, we discovered that the modifications in the procedures for administering the SPT have further accentuated the SPT's failure to accomplish its task. As noted above, placing 10 dubious items as gatekeepers and lowering the threshold of these items to identify SHL learners is questionable at best. Even more problematic is the

removal of language faculty from the process of reviewing the results of the SPT and having the opportunity to supplement the tests or even correct them with personal interviews, particularly in light of the discovery that the test alone is of little consequence in accurately placing SHL learners in our program. Finally, having the SPT freely accessible on the web not only raises security concerns but also diminishes our ability to focus in on the SHL learners in our program.

In sum, despite NMSU's long-standing legacy of promoting the linguistic abilities of SHL learners, we ourselves are guilty of not following through with the basic principles of testing ethics. Our actions may have ended up doing SHL learners more harm than good by requiring them to take a lengthy test that is inadequate to judge their linguistic skills and thus placing them inaccurately in our program. We have been even more remiss in removing Spanish teaching faculty from participating in placement decisions. While we don't think our program is alone in making mistakes regarding the placement of HLLs, we acknowledge that we fell short of the mark in several instances and offer these experiences as a caveat for others who are either currently using a placement test that they know little about or are contemplating designing their own test. Nonetheless, we recognize that this situation presents us with the opportunity to correct our mistakes and are currently working toward redesigning the SPT to make it more applicable to the local SHL learner population and to our SHL curriculum, as well as providing SHL learners the opportunity to display the full range of their linguistic and cultural talents. To further enhance these efforts, we will restore the involvement of Spanish teaching faculty in the placement process and will incorporate a means for these faculty to analyze placement decisions in order to provide greater oversight on our procedures. Overall, our desire is to share the responsibility for designing and maintaining both a product and a process that is effective in appropriately placing students, is culturally and linguistically relevant to our program and community of learners, and is responsive to changes in either.

References

Achugar, M. (2003). Academic registers in Spanish in the U.S.: A study of oral texts produced by bilingual speakers in a university graduate program. In A. Roca & M. C. Colombi (Eds.), *Mi lengua: Spanish as a heritage language in the United States* (pp. 213-230). Washington, DC: Georgetown University Press.

Alarcón, I. (2010). Advanced heritage learners of Spanish: A sociolinguistic profile for pedagogical purposes. *Foreign Language Annals, 43*(2), 269-288.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

Barker, M. E. (1972). *Español para el bilingüe*. Skokie, IL: National Textbook Company.

Beaudrie, S. (2007). La enseñanza del acento ortográfico en español y su relación con la percepción de la sílaba tónica. *Hispania, 90(4) 809-823.*

Beaudrie, S., & Ducar, C. (2005). Beginning level university programs: Creating a space for all heritage language learners. *Heritage Language Journal, 3*(1), 1-26. Available from http://www.heritagelanguages.org

Bills, G. D., Hernández Chávez, E., & Hudson, A. (1995). The geography of language shift: distance from the Mexican border and Spanish language claiming in the Southwestern U. S. *International Journal of the Sociology of Language, 114*, 9-27.

Bowles, M. (2011). Exploring the role of modality: L2-Heritage learners' interactions in the Spanish language classroom. *Heritage Language Journal, 8*(1), 30-65. Available from http://www.heritagelanguages.org

Carrasco, R. L., & Rigelhaupt, F. (2003). META: a model for the continued acquisition of Spanish by Spanish/English bilinguals in the United States. In A. Roca & M. C. Colombi (Eds.), *Mi lengua: Spanish as a heritage language in the United States* (pp. 170-197). Washington, DC: Georgetown University Press.

Carreira, M. 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., & Potowski, K. (2011). Commentary: Pedagogical implications of experimental SNS research. *Heritage Language Journal, 8(1)*, 134-151. Available from http://www.heritagelanguages.org

Cizek, G. J., Fitzgerald, S. M., & Rachor, R. E. (1995). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment, 3*(2), 159-179.

Consejo Nacional de Población (CONAPO) (n.d.) Proyecciones de la población de México 2005-2050: Municipales. Retrieved from http://www.conapo.gob.mx/00cifras/proy/municipales.xls

Correa, M. (2011). Advocating for critical pedagogical approaches to teaching Spanish as a heritage language: Some considerations. *Foreign Language Annals, 44*(2), 308-320.

Ebel, R. L. & Frisbie, D. A. (1986). *Essentials of education measurement*. Englewood Cliffs, NJ: Prentice Hall.

Fairbairn, S. B., & Fox, J. (2009). Inclusive achievement testing for linguistically and culturally diverse test takers: Essential considerations for test developers and decision makers. *Educational Measurement: Issues and Practice, 28*(1), 10-24.

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.

Fairclough, M. & Mrak, N. A. (2003). La enseñanza del español a los hispanohablantes bilingües y su efecto en la producción oral. In A. Roca & M. C. Colombi (Eds.), *Mi lengua: Spanish as a heritage language in the United States* (pp. 198-212). Washington, DC: Georgetown University Press.

Galindo, L. (1996). Language use and language attitudes: a study of border women. *The Bilingual Review/La revista bilingüe*, *21*(1), 5-17.

Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.

Iverson, M. (2009). Knowledge of noun-drop across various lexical and functional categories in heritage Spanish bilinguals. In J. Crawford, K. Otaki & M. Takahashi (Eds.) *Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition North America (GALANA 2008)* (pp. 98-106). Somerville, MA: Cascadilla Proceedings Project

Kehoe, J. (1995a). Writing multiple-choice test items. *Practical Assessment, Research & Evaluation*, 4(9). Retrieved from http://PAREonline.net/getvn.asp?v=4&n=9

Kehoe, J. (1995b). Basic item analysis for multiple-choice tests. *Practical Assessment, Research & Evaluation, 4*(10). Retrieved from http://pareonline.net/getvn.asp?v=4&n=10

Lord, F.M. (1952). The relationship of the reliability of multiple-choice test to the distribution of item difficulties," *Psychometrika, 18*, 181-194.

MacGregor-Mendoza, P. (1998). Language and the bilingual teacher: Use, attitudes, roles, *Southwest Journal of Linguistics, 17*(2), 83-99.

MacGregor-Mendoza, P. (2000). Aquí no se habla español: Stories of linguistic repression in Southwest schools. *Bilingual Research Journal, 24*(4): 333-345.

MacGregor-Mendoza, P. (2005). El Desplazamiento intergeneracional del español en los Estados Unidos: Una aproximación. In L. A. Ortiz López y M. Lacorte (Eds.), *Contactos y contextos lingüísticos: El español en los Estados Unidos y en contacto con otras lenguas* (pp. 287-300). Madrid: Lingüística Iberoamericana.

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 from the 13th Hispanic Linguistics Symposium* (pp. 150-160). Somerville, MA: Cascadilla Press.

McMillan, J. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and practice, 20*(1), 20-32.

McMillan, J. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and practice, 22*(4): 34-43.

McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.

Milroy, J. and Milroy, L. (1997). Varieties and variation. In F. Coulmas (Ed.), *The handbook of sociolinguistics,* pp. 47-64. Oxford, UK: Blackwell Publishers

Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory into Practice, 31*(2), 132-141.

Montrul, S. (2002). Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism: Language and Cognition, 5*, 39-68.

Montrul, S. (2004). Subject and object expression in Spanish heritage speakers: A case of morpho-syntactic convergence. *Bilingualism: Language and Cognition, 7*, 125-142.

Montrul, S. (2005). Second language acquisition and first language loss in adult early bilinguals: exploring some differences and similarities. *Second Language Research, 21*(3), 199-249.

Montrul, S. (2006). On the bilingual competence of Spanish heritage speakers: Syntax, lexical-semantics and processing. *International Journal of Bilingualism, 10*(1), 37–69.

Montrul, S. (2008). *Incomplete acquisition in bilingualism. Re-examining the age factor*. Amsterdam: John Benjamins.

Montrul, S. (2009). Knowledge of tense/aspect and mood in Spanish heritage grammars. *International Journal of Bilingualism*, *13*(2), 239-269.

Montrul, S. (2010). Current issues in heritage language acquisition. *Annual Review of Applied Linguistics, 30,* 3-23.

Montrul, S. (2011). Morphological errors in Spanish second language learners and heritage speakers. *Studies in Second Language Acquisition, 33*, 163-192.

Montrul, S. and Bowles, M. (2010). Is grammar instruction beneficial for heritage language learners? Dative case marking in Spanish. *Heritage Language Journal, 7*(1) 1-25. Available from http://www.heritagelanguages.org

Montrul, S., Foote, R., Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning, 58*(3): 503-553.

Montrul, S., & Perpiñán, S. (2011). Assessing differences and similarities between instructed L2 learners and heritage language learners in their knowledge of Spanish Tense-Aspect and Mood (TAM) Morphology. *The Heritage Language Journal, 8*(1). Available from http://www.heritagelanguages.org

New Mexico State University. (2008). 2008-2009 Undergraduate catalog. Las Cruces, NM: New Mexico State University.

Norman, L M., Parcher, J. W., & Lam, A. H. (2004). Monitoring colonias along the United States-Mexico border: U.S. Geological Survey Fact Sheet 2004-3070. Retrieved from http://egsc.usgs.gov/isb/pubs/factsheets/fs307004.html

Potowski, K., Jegerski, J., & Morgan-Short, K. (2009). The effects of instruction on subjunctive development among Spanish heritage language speakers. *Language Learning, 59*(3), 537-579.

Rodríguez Pino, C., & Villa, D. (1994). A student-centered Spanish for native speakers program: Theory, curriculum and outcome assessment. In C. Klee (Ed.), *Faces in a crowd: Individual learners in multisection programs. American Association of University Supervisors and Coordinators and Directors of Foreign Language Programs Issues in Language Program Direction* (pp. 355–373). Boston: Heinle & Heinle.

Romaine, S. (1995). *Bilingualism* (2nd Ed.) Oxford, UK: Blackwell Publishers.

Schwartz, A. M. (2003). ¡No me suena! Heritage Spanish speakers' writing strategies. In A. Roca & M. C. Colombi (Eds.), *Mi lengua: Spanish as a heritage language in the United States* (pp. 235-256). Washington, DC: Georgetown University Press.

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher, 32*(2), 3-13.

Spicer-Escalante, M. (2005). Writing in two languages/Living in two worlds: A rhetorical analysis of Mexican-American written discourse. In M. Farr (Ed.), *Latino language and literacy in ethnolinguistic Chicago* (pp. 217-246). Mahwah, New Jersey: Erlbaum.

Teschner, R. V. (1990). Spanish speakers semi- and residually native: After the placement test is over. *Hispania, 73*, 816-822.

Teschner, R. V. (1995). Beachheads, islands, and conduits: Spanish monolingualism and bilingualism in El Paso, Texas. *International Journal of the Sociology of Language*, *114*, 93-105.

Trumbull, E., & Solano-Flores, G. (2011). The role of language in assessment. In Basterra, M., Trumbull, E. & Solano-Flores, G. (eds.*), Cultural validity in assessment: Addressing linguistic and cultural diversity* (pp. 22-45). London: Routledge.

U.S. Census Bureau. (2011). Table QT-P10: Hispanic or Latino by Type [data table], Doña Ana County and Las Cruces City, New Mexico. 2010 Census Summary File 1. Washington, DC: Author. Available from http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml

Valdés, G. (1992). The role of the foreign language teaching profession in maintaining non-English languages in the United States. In H. Byrnes,(Ed.) *Languages for a Multicultural World in Transition.1992 Northeast Conference Reports* (pp. 29-71)*.* Chicago: National Textbook Company.

Valdés, G. (1997). The teaching of Spanish to bilingual Spanish-speaking students: Outstanding issues and unanswered questions. In C. Colombi and F. Alarcón (Eds.), La *enseñanza del español a hispanohablantes: Praxis y teoría* (pp. 93-101). Boston, MA: Houghton Mifflin.

Valdés, G. (2001). Heritage language students: Profiles and possibilities. In J. K. Peyton, D. A. Ranard, and S. McGinnis (Eds.), *Heritage languages in America: Preserving a national resource* (pp. 37-77). McHenry, IL: Delta Systems Co., Inc.

Valdés-Fallis, G. (1973). Spanish as a native language. *Hispania, 56* (4), 1041-1043.

Valdés Fallis, G. (1976). Pedagogical implications of teaching Spanish to the Spanish-speaking in the United States. In G. Valdés Fallis and R. García-Moya (Eds.) *Teaching Spanish to the Spanish speaking: Theory and practice*, (pp. 3-27). San Antonio, TX: Trinity University.

ValdésFallis, G. (1978). A comprehensive approach to the teaching of Spanish to bilingual Spanish-speaking students, *Modern Language Journal, 62*, 102-110.

Velázquez Vargas, M., & Martínez Toyes, W. L. (2010). Migración y violencia. *Observatorio del seguridad y convivencia ciudadanas Boletín Num. 4,* 54-60. Retrieved from http://observatoriodejuarez.org/dnn/portals/0/boletines/pdfs/Boletin_4_Obs.pdf

Witkowsky, K. (2003). El Paso's border students. *National Crosstalk, 11*(1)*,* 15-16. Retrieved from http://www.highereducation.org/crosstalk/ct0103/front.shtml

Appendix

Table A1

Catalog Descriptions of SHL Courses (New Mexico State University, 2008)

SPAN 113 Beginning Spanish for Native Speakers
Emphasis on listening comprehension and recognition of vocabulary for retrieval of the native language. Confidence building activities for developing oral skills.

SPAN 213 Spanish for Native Speakers I
Emphasis on development of native language reading skills. Covers speaking, writing and vocabulary activities to strengthen command of the language. For Spanish-speaking students only.

SPAN 214 Spanish for Native Speakers II
Emphasis on reading and writing with speaking activities for skill development. Discussion of problematic areas in grammar.

SPAN 312 Grammar for Native Speakers of Spanish
For students who have been exposed to Spanish at home or in the community. Review of grammatical concepts and analysis of both spoken and written Spanish.

Table A2

Sample Items from Sections of the SPT

1-40				
basic	¿Qué _____ comiendo?			
grammar	(What _____ eating?)			
	a. has	b. puedes	c. eres	d. estás
	a. have you	b. can you	c. are you	d. are you
			(attribute cop)	(progressiv cop.)
41-50				
basic	La televisión cuesta 1500 dólares. No la puedo comprar porque es muy ____.			
vocabulary	The television costs 1500 dollars. I can't buy it because it's very _____.			
	a. revista	b. barata	c. cara	d. moneda
	a, magazine	b. cheap	c. expensive	d. coin
51-70				
intermediate	La señora Corina está enferma de _____			
vocabulary	Mrs. Corina has a/an _____ illness.			
	a. un roble b. un fardo c. un pulmón d. la manga e. la acera			
	a. oak b.bundle c. lung d. sleeve e. sidewalk			

Table A2 (con't.)

71-80 orthography	El cocinero _____ un flan especial The cook _____ a special flan. a. hiso b. hizo a.made b. made (homophonic distractor)
81-90 formal grammar	¿Cuál de las palabras siguientes es un <u>pronombre demostrativo</u>? Which of the following words is a <u>demonstrative pronoun</u>? a. ella b. uno c. nuestro d. ése a. she b. one c. our d. that

Note

1. Barker's (1972) *Español para el bilingüe* illustrates this heavily prescriptive philosophy in the description of her text when she claims, *"...hemos formulado ejercicios de tipo correctivo que como su nombre lo indica, se han preparado para el alumno que ha hecho sus estudios en inglés, y que por eso tiene la tentación de emplear anglicismos, arcaísmos y otros vicios de dicción (iii)".* (…we have designed corrective exercises which, as their name implies, have been prepared for the student who has been schooled in English, and because of this is tempted to use anglicisms, archaisms and other linguistic vices (iii)). The approach is championed in the introduction by a colleague who writes, "…a certain dogmatism is necessary; since it is too late to refine the language gradually, one must resort to a greater than usual degree of force to strain out the impurities—those ingredients which make the product unmarketable on a large scale." (p. vi)