

February 2018

Does the test work? Evaluating a web-based language placement test

Avizia Long
Texas Tech University

Sun-Young Shin
Indiana University - Bloomington

Kimberly Geeslin
Indiana University - Bloomington

Erik Willis
Indiana University - Bloomington

Follow this and additional works at: https://scholarworks.sjsu.edu/world_lang_pub

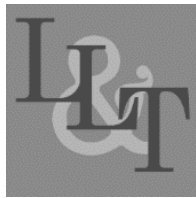


Part of the [Artificial Intelligence and Robotics Commons](#), [Computational Linguistics Commons](#), [Educational Assessment, Evaluation, and Research Commons](#), [First and Second Language Acquisition Commons](#), and the [Spanish Linguistics Commons](#)

Recommended Citation

Avizia Long, Sun-Young Shin, Kimberly Geeslin, and Erik Willis. "Does the test work? Evaluating a web-based language placement test" *Language Learning & Technology* (2018): 137-156. <https://doi.org/10125/44585>

This Article is brought to you for free and open access by the World Languages and Literatures at SJSU ScholarWorks. It has been accepted for inclusion in Faculty Publications by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.



Does the test work? Evaluating a web-based language placement test

Avizia Y. Long, Texas Tech University

Sun-Young Shin, Indiana University

Kimberly Geeslin, Indiana University

Erik W. Willis, Indiana University

Abstract

In response to the need for examples of test validation from which everyday language programs can benefit, this paper reports on a study that used Bachman's (2005) assessment use argument (AUA) framework to examine evidence to support claims made about the intended interpretations and uses of scores based on a new web-based Spanish language placement test. The test, which consisted of 100 items distributed across five item types (sound discrimination, grammar, listening comprehension, reading comprehension, and vocabulary), was tested with 2,201 incoming first-year and transfer students at a large, Midwestern public university. Analyses of internal consistency and validity revealed the test to be reliable and valid with regard to its functionality, the content covered on the exam, and the consistency with which placement decisions could be made. Findings are discussed in light of the AUA model developed for the placement test, and practical suggestions for university-level language program instructors and testing administrators are outlined.

Keywords: *Assessment, Testing*

Language(s) Learned in this Study: *Spanish*

APA Citation: Long, A. Y., Shin, S.-Y., Geeslin, K., & Willis, E. W. (2018). Does the test work? Evaluating a web-based language placement test. *Language Learning & Technology*, 22(1), 137–156.
<https://dx.doi.org/10125/44585>

Introduction

Test developers, institutions of learning, and instructors alike are increasingly turning to the web as a medium for testing language skills. Web-based language testing offers several benefits, including automatic scoring and autonomy in testing location and time for test takers. These benefits have prompted the spread of the web as a tool for language testing from low-stakes testing situations—originally advised by Roever (2001)—to medium-stakes testing contexts (Shin, 2012). One such medium-stakes testing context that has witnessed an increased use of the web as a means for test delivery is the university-level foreign language (FL) placement exam. Particularly within the institutional context of the present study, there is a push to change the administration of FL placement exams from paper- to web-based formats, given the increased flexibility of delivering a web-based test to larger numbers of students. Despite the aforementioned benefits of using web-based tests for placement purposes, there has been little research thus far investigating the reliability and validity of online FL placement tests. Additionally, the move from paper- to web-based placement testing oftentimes implicates the assistance of individuals such as classroom instructors and language program administrators who may lack the necessary expertise in language testing and evaluation to assess the functionality of a placement test. This all-too-common situation has created a need for examples of test validation from which everyday FL programs can benefit. To that end, the primary goal of the present study is to offer an example method to evaluate a

web-based FL placement test geared toward everyday FL programs. Specifically, the present investigation shows how to use Bachman's (2005) assessment use argument framework to examine evidence to support claims made about the intended interpretations and uses of scores based on a new web-based Spanish language placement test. In addition to contributing to the discussion of web-based language assessment for medium-stakes testing contexts, our example for evaluating the functionality and use of a FL placement test in a web-based setting that is offered is comprehensive and replicable. As will be shown, this method has quantitative aspects that do not require advanced knowledge of statistics, as well as qualitative components illustrating that statistics are not (and should not be) the only means for test evaluation.

Background

Web-Based Language Testing

Web-based language testing (WBLT) uses the Internet as a platform for test development and delivery; test input and questions are written in the HTML located on a server and test takers respond to the test items using web browsers such as Internet Explorer, Firefox, Chrome, or Safari (Shin, 2012). Recently, WBLT has been embraced more by language researchers and teachers as a teaching and testing tool because it has the potential to greatly enhance logistical efficiency and flexibility (Ockey, 2009). Test developers can easily upload and update test contents, and test takers can take the test at the place and time of their convenience. Test takers' responses on the test are scored immediately, and scores are reported to all stakeholders more quickly. Various item and test score statistics are available on demand, providing useful information for test developers and users to interpret test scores and revise the test when necessary.

Additionally, WBLT has been known to lead to improved test measurement qualities including reliability and validity (Chapelle & Douglas, 2006). A large number of test takers' responses on true/false and multiple-choice formats can be instantly scored without any errors. Even productive responses are scored consistently once reliable scoring algorithms are developed and applied to test takers' responses (Bernstein, Van Moere, & Cheng, 2010; Carr & Xi, 2010). Further, inter- and intra-rater reliability in assessing test takers' written or spoken responses are not a concern in WBLT using an automated scoring system (Williamson, Bejar, & Sax, 2004). Authenticity can also be enhanced because various test formats are possible, including interactive and dynamic features of test input and question types (Chapelle & Douglas, 2006; Huff & Sireci, 2001). For example, computer technology makes it possible to include visual input for online listening tests, more closely reflecting language use in real-world tasks (Ockey, 2007; Wagner, 2010). Thus, WBLT is becoming more widely used in many high-stakes standardized language proficiency exams, such as the Test of English as a Foreign Language Internet-based Test (TOEFL iBT) and the Pearson Test of English Academic (PTE Academic), as well as in placement and screening tests used for medium-stakes decisions made in FL programs (Bardovi-Harlig & Shin, 2014; Elder & Randow, 2008).

Placement Testing

Previous research on placement testing has grown remarkably over the past several decades, reflecting the incorporation of languages other than English (e.g., Bernhardt, Rivera, & Kamil, 2004; Eda, Itomitsu, & Noda, 2008) and the discussion of several key issues and topics in the field of language testing. An important issue that has attracted the attention of language testing scholars is related to the nature of placement tests. As pointed out by Brown (1989), a central purpose of placement testing is to "[sort] students into relatively homogeneous language-ability groupings, sometimes within specific skill areas" (p. 65). Further, placement tests can be designed to spread out these groupings along a continuum from lower to higher scoring students or to distinguish students who have mastered specific learning objectives from those who have not (Green, 2012). The former type of test is called a norm-referenced test (NRT), in which a test taker's score is compared to other individuals who take the same test, whereas the latter type

of test is called a criterion-referenced test (CRT), in which scores are used to determine whether or not a test taker demonstrates mastery of a previously specified skill or content related to instruction. However, Brown (1989) points out that in actual use, aspects of both NRT and CRT are present in placement testing. Specifically, placement testing developers and administrators may want to determine what objectives have been learned or mastered while simultaneously spreading out students to identify ability or proficiency groups. In these cases, Brown (1989) calls for the use of norm- and criterion-based item analysis techniques for placement testing contexts, not only to determine how much the test spreads out students but also to examine the extent to which NRTs include content relevant to the given curriculum under study. It should also be noted that a NRT puts emphasis on discriminating among students relative to others, whereas a CRT is concerned with content mastery of a particular course (Brown & Hudson, 2002). Similarly, it is important to notice that score reliability for a CRT is not estimated in the same way as it is for a NRT because a CRT tends to produce little variance in test scores (Bachman, 2004). Thus, a placement test featuring both orientations should demonstrate that the test could be judged against various test qualities such as reliability and validity before it can be implemented for decision-making purposes, such as course placement.

An examination of previous language placement testing research demonstrates that issues of reliability and validity have received a fair amount of attention in the empirical literature. Example studies include those by Wall, Clapham, and Alderson (1994) and Fulcher (1997), in which not one but several approaches to assessing reliability (e.g., correlation coefficients, inter- and intra-rater reliability, Rasch data modeling, etc.) and validity (e.g., analyses of cut scores, principle components analysis, concurrent validity, content validity, feedback from students and instructors, etc.) are employed. These studies demonstrate the range of approaches that are appropriate for the development and assessment of language placement tests. Furthermore, these approaches have been widely adopted in more recent studies, which have served to strengthen methods of language placement testing development and evaluation (e.g., Bernhardt et al., 2004; Eda et al., 2008; Pardo-Ballester, 2010; Shin, 2008).

It is important to point out that the present study is not the first to investigate the presumed practicality and efficiency of web-based language placement testing in a university-level FL program. Bernhardt et al. (2004) examined the process of using the web to deliver language placement tests to incoming and transfer students (i.e., not yet enrolled) at Stanford University, showing that web-based placement testing could be conducted in a reliable and valid manner. Additionally, their findings revealed that online testing was particularly beneficial for testing administrators and instructors who made more effective decisions when given the time to contemplate a student's performance. For students, the convenience of taking the exam at a time and place of their choosing saved them an additional trip to campus. The present analysis improves and expands upon research such as Bernhardt et al. (2004) by gathering and presenting additional evidence in support of test validity—namely, content relevance, concurrent criterion relatedness, and consequences of test score uses. Furthermore, the current study shows the need for test validation procedures that are based on a logical framework of argumentation and supporting evidence to address the appropriateness of both score interpretations and uses (Carr, 2011).

Test Validation Approaches

Validity evidence for the appropriateness of interpretation and uses of a given test score has been traditionally examined through Messick's validity framework (Messick, 1989). In this framework, Messick argues that evidence supporting different aspects of validity including content, concurrent, construct, and consequential validity should be viewed as a unitary concept that integrates both evidential and consequential basis for validation. However, given that his approach has not been considered useful for practical test development (Bachman, 2005), recent views of test validity and use bear on an argument-based approach in which multiple claims about score-based inferences and test uses are investigated in a logical manner (Kane, 1992, 2011; Toulmin, 2003). This notion of validity as argument allows us to test alternative interpretations and to reduce doubt surrounding the claims we make for the interpretation and use of the test scores (Fulcher & Davidson, 2007). Bachman (2005) further develops

this into assessment use arguments (AUAs) as a framework for building and supporting a case for test score interpretation and use. This approach combines the following two arguments into a single framework: (a) the *validity argument*, which refers to the logical link between test scores and intended score interpretation, and (b) the *utilization argument*, which is related to the coherent connection drawn between score interpretation and intended uses. Thus far, Bachman's AUA has provided language testing researchers with a theoretical tool for explicitly establishing a logical structure for simultaneously justifying the interpretations made about test takers' language abilities and decisions about them based on their given test scores. This framework has been found to be quite useful for language testers attempting to validate a test in that it provides useful guidance on how to collect and organize various information supporting score-based interpretations and uses. AUAs have thus been increasingly used in the field of language testing (e.g., Jia, 2013; Llosa, 2008; Shin, 2008; Wang, Choi, Schmidgall, & Bachman, 2012).

The Present Study

As demonstrated in our review thus far, WBLT has numerous benefits, leading university-level FL programs to adopt the web as a tool for language placement testing development and use. However, it is important for all individuals involved in this process to be aware of the nature of placement tests and the aspects of test quality to address when evaluating a placement test (i.e., reliability and validity). Furthermore, it is imperative that FL placement test designers and evaluators critically examine the claims that are made about the inferences drawn from test scores, as "the single most important consideration in both the development of language tests and the interpretation of their results is the purpose or purposes which the particular tests are intended to serve" (Bachman, 1990, p. 55). To date, studies examining the validity and use of placement tests have not offered models of validation geared toward everyday FL programs. To that end, the present study investigates the functionality and use of a web-based Spanish FL placement test to illustrate how to examine issues of reliability and validity. Specifically, Bachman's (2005) AUA framework is adopted to exemplify the process of linking test scores and scored-based inferences and uses in a logical and accessible manner. The specific research questions guiding the present study are the following:

1. To what extent is the web-based Spanish placement exam a reliable indicator of Spanish language knowledge for incoming university-level learners?
2. To what extent is the web-based Spanish placement exam a valid indicator of Spanish language knowledge within the curricular context of the present study?
3. To what extent can our framework for test validation and use be generalized to other university-level FL contexts?

Research of this kind is essential to illustrate how to gather and link evidence to support claims that are made on the basis of test scores, and such an example would be informative to language test developers in similar, large-scale, university-level FL programs in the United States.

The Setting

Within the institutional context of the current study, Spanish is the most popular choice for FL study among university undergraduate students. The academic department in which Spanish language is organized offers courses at all levels (elementary through graduate) and serves approximately 4,000 students annually. Incoming students often have some knowledge of Spanish, having taken Spanish language or content courses in high school. Therefore (and similar to many postsecondary institutions throughout the United States), a placement test is administered to incoming students to match them (to the greatest extent possible) with the course that is most suitable for their current level of Spanish language knowledge.

Prior to the development of the web-based placement test under examination in the present study, a paper-based version of the Wisconsin Spanish Placement Test (University of Wisconsin Center for Placement

Testing, 2011) was used to place students into lower-division courses of the Spanish language program. The paper-based test contained 60 multiple-choice items that assessed Spanish language grammar, vocabulary, and reading comprehension. It was originally administered to incoming and new transfer students during summer orientation visits on campus. It was scored by a testing center located on the campus of the institution in which the present study took place, and test results (with corresponding placement decisions) were made available to test administrators within one to three days. Placement decisions were based strictly on scores from the placement test, which placed students into one of the following courses: First Year Spanish, Second Year Spanish I, Second Year Spanish II, Spanish Grammar in Context, or Introduction to the Study of Hispanic Cultures. If students received a high enough score, they could be placed out of all of the courses previously mentioned, permitting them to enroll in other upper-division content courses taught in Spanish.

The decision to move to an online format was first initiated by the university administrators, who determined that incoming students' orientation campus visits could be used more effectively if all university placement tests were taken online prior to the campus visit. However, the movement to an online format raised copyright concerns with the paper-based test, which necessitated the creation of an exam without a preexisting copyright. The movement to an online format, however, afforded the opportunity to address certain limitations of the paper-based test. First, the paper-based test did not accurately reflect the breadth of content to which students are exposed in the language courses of the current departmental context. The paper-based test only tested grammar, vocabulary, and reading (with a focus on grammar), but the web-based test was improved to include skills that were assessed in the curriculum, such as listening comprehension. In addition, in the departmental context of the present study, it was important to make quick placement decisions and to inform students, teachers, and administrators of the results promptly, given the large number of students enrolling in Spanish language courses each semester. Taken together, a web-based language test provided greater, more authentic breadth of content; offered instant placement decisions; and permitted diverse, multimodal test inputs (Shin, 2012). As a result, a web-based test that was different from the paper-based test was developed and implemented to replace the previous paper-based test for placement decisions in the Spanish language program.

The web-based placement test was designed to place incoming students into one of the six Spanish language courses outlined in Table 1. Given a high enough score, students could place out of these courses, making them eligible for enrollment in other upper-division courses that focused on topics related to culture, linguistics, and literature. The cutoff score for the web-based test was determined by the performance criteria and levels established using the previous paper-based test that students took on campus. We performed a statistical comparison (Kappa's agreement index) to match as closely as possible the results from the web-based test to the results of the paper-based test, which best represented the placement norms previously established for course placement in the departmental context of the present study.

Table 1. *Spanish Language Courses for Placement Based on Web-Based Placement Test*

Curriculum Division	Course Name
Lower-Division	Elementary Spanish
	Elementary Spanish II
	First Year Spanish ^a
	Second Year Spanish I
	Second Year Spanish II
Upper-Division	Spanish Grammar in Context

^a*This course is an accelerated course, covering all content included in Elementary Spanish and Elementary Spanish II within a single academic semester.*

From Table 1, an observation that is relevant to understanding the context of test use is that the test was designed to assess knowledge of language. Consequently, the majority of the courses in which students were placed fell into the lower-division bracket. Although topics related to culture were included in each of these courses, attention was given to developing and refining knowledge of structural aspects of the language as this knowledge was viewed as the essential preparation for the academic writing required in more advanced, upper-division courses.

The Web-Based Test

The test was delivered via the web to incoming students. Specifically, incoming students took the test on their own (at home using their computers) before arriving at the institution for their first academic semester of study. The web-based test was not adaptive, meaning that subsequent items were not presented to test takers from a bank of items depending on their response to the previous item. To facilitate comprehension of the demands of the assessment, all instructions were provided in English. Students were permitted to take the test one time only, after which their responses were recorded and scored electronically. The test was constructed using HTML by technology support staff at the institution of the present study. Additionally, all listening files included on the test were recorded directly onto a computer using a USBpre external sound card and a Shure wh10a microphone.

The test contained 100 items across a total of five sections: sound discrimination ($k = 5$, where k is number of items), grammar ($k = 65$), listening comprehension ($k = 13$), reading comprehension ($k = 7$), and vocabulary ($k = 10$). In what follows, a detailed description of each section with sample (not actual) test items is provided.

Sound Discrimination

The first section required test takers to listen to a phrase in Spanish that contained a word targeting a specific sound segment. The word always occurred in phrase-initial position, but targeted segments occurred in either word-initial or word-medial position. Each word was a real word in Spanish, spoken by a native speaker of Spanish reading at a normal, conversational rate of speech. The items of this section took a multiple-choice format with three response options (see Figure 1 for a sample item). Test takers could listen to each phrase up to two times.

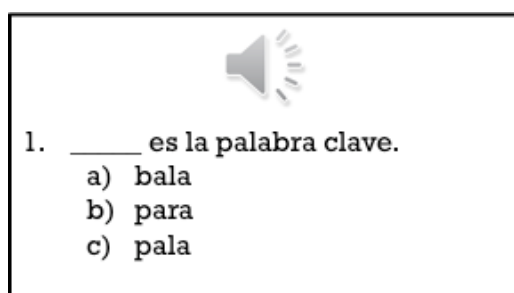


Figure 1. Sample sound discrimination item

Grammar

The second section required test takers to observe a picture, read a contextualized phrase, or read a cloze passage and then select the response that best described the picture or completed the phrase or passage provided (see Figure 2). The items of this section also took a multiple-choice format, with two, three, or four response options. With regard to the content of this section, items covered a wide range of grammatical structures known to present an acquisitional challenge to predominantly English-speaking learners (see Geeslin, 2013). Some of these structures included: the copula contrast (i.e., *ser* vs. *estar*), aspect (i.e., preterit vs. imperfect), mood selection (i.e., subjunctive vs. indicative), verbal morphology across tenses (e.g., present, future, conditional, etc.), grammatical gender and number agreement, clitic

pronouns (e.g., *lo, le, les*), and imperatives (e.g., *dime*), among others.¹ These structures were taught in the curriculum designed for the lower-division courses.


 <p>1. El perro...</p> <ul style="list-style-type: none"> a) estás triste b) es triste c) está triste 	<p>2. Estás en casa y hace mucho calor. Tu amigo está cerca de la ventana y para que él la abra dices:</p> <ul style="list-style-type: none"> a) abre la ventana b) abra la ventana 	<p>Ayer tuve que (3) _____ porque no quería llegar tarde a clase. Cuando llegué la maestra me castigó frente todos los alumnos de la clase...</p> <p>3. Infinitivo: correr</p> <ul style="list-style-type: none"> a) corre b) correr c) corrían d) corrió
---	---	---

Figure 2. Sample grammar items

Listening Comprehension

The third section contained three listening passages of varying lengths: the first, 31 s; the second, 18 s; and the third, 60 s. These passages dealt with a monologic description of people, a telephone message, and a news-related report, respectively. Test takers were instructed to listen to the passage and select the best response to complete each corresponding set of items provided. Each item took either a multiple-choice or true/false response format; multiple-choice items always provided four response options. The first passage contained a set of 5 corresponding items, the second passage 3 items, and the third passage 5 items. No visual input was provided in this section of the test.

Reading Comprehension

The fourth section contained two reading texts of varying lengths: the first was 99 words long and the second was 408 words long. These passages dealt with lifestyle and history, respectively, and were accompanied by a set of items that varied in number to reflect the length of each passage (two and five items, respectively). Test takers were instructed to read the passage and answer each item provided based on the content of its corresponding reading passage. Each item had a multiple-choice format with four response options.

Vocabulary

The final section required test takers to complete phrases that prompted the selection of an appropriate synonym or antonym (see Figure 3). The vocabulary items were selected from Davies' (2006) *A frequency dictionary of Spanish: Core vocabulary for learners*. Specifically, five words were selected from the last 10 words of each block of 100 words (there are 5,000 words total in the dictionary). With regard to the types of words selected, this section targeted mostly adjectives, followed by nouns and nonfinite verbs. These items took a multiple-choice format with four response options.

<p>1. La palabra bonito es similar a:</p> <ul style="list-style-type: none"> a) feo b) sucio c) limpio d) guapo 	<p>2. La palabra difícil es opuesta a:</p> <ul style="list-style-type: none"> a) vago b) alto c) fácil d) duro
--	---

Figure 3. Sample vocabulary items

Method for Test Evaluation

Participants and Procedure

For the present study, the web-based placement exam was evaluated based on the performance of 2,201 incoming first-year and transfer students. All students reported having 0 to 4 years of classroom Spanish instruction at the time of the test. Although the majority of students reported English as their native language ($n = 2,111$), there were some students who reported a non-English native language (Spanish, $n = 28$; Other, $n = 39$).²

As part of the validation framework developed for the present study, 1,622 of the 2,201 incoming first-year and transfer students who took the web-based test also took the paper-based test that was being used in the department to place students into lower-division courses of the Spanish language program. This subset of the participants completed the paper-based test during the summer orientation session before their enrollment into courses for the following fall semester.

AUA Model for the Validity of the Spanish Placement Exam

Bachman's (2005) AUA framework was adopted to articulate the uses of and inferences drawn from the web-based Spanish language placement test described. Using Bachman's model, the validity and utilization arguments pertaining to each score-based inference and use of the web-based Spanish placement exam were articulated, as shown in [Figure 4](#) and [Figure 5](#), respectively.³ These arguments consisted of warrants to support each claim made, as well as rebuttals that could weaken the warrants stated for the Spanish language placement exam. Additionally, each argument outlined the type of evidence collected to serve as backing or rebuttal data.

The desired claim based on our interpretations of students' scores ([Figure 4](#)) on the placement test was that the exam is a reliable and valid indicator of Spanish language ability for university-level classroom learners. To support this claim, we addressed whether or not students' performance on the test was consistent (Warrant 1) by examining statistical estimates of internal consistency (Backing 1). A potential weakness of this warrant was if there were several items that individually influenced the statistical estimates of internal consistency to a great extent (Rebuttal 1). Therefore, we identified any such items and examined the extent to which they influenced statistical estimates of reliability of the test (Rebuttal data 1).

To provide further support for the score-based interpretation claim, we addressed the extent to which the content of the placement test reflected the content present in the courses in which students could be placed (Warrant 2) by matching test items with the corresponding course content (Backing). However, a potential rebuttal was that the content of the placement test was limited in comparison to the content of the courses in which students were being placed (Rebuttal 2). As a result, we identified additional evidence to demonstrate that the content of the exam matched the content of the courses in which students could enroll. In this case, we compared the content of the placement test to the content reflected in course syllabi and textbooks (Rebuttal data 2).

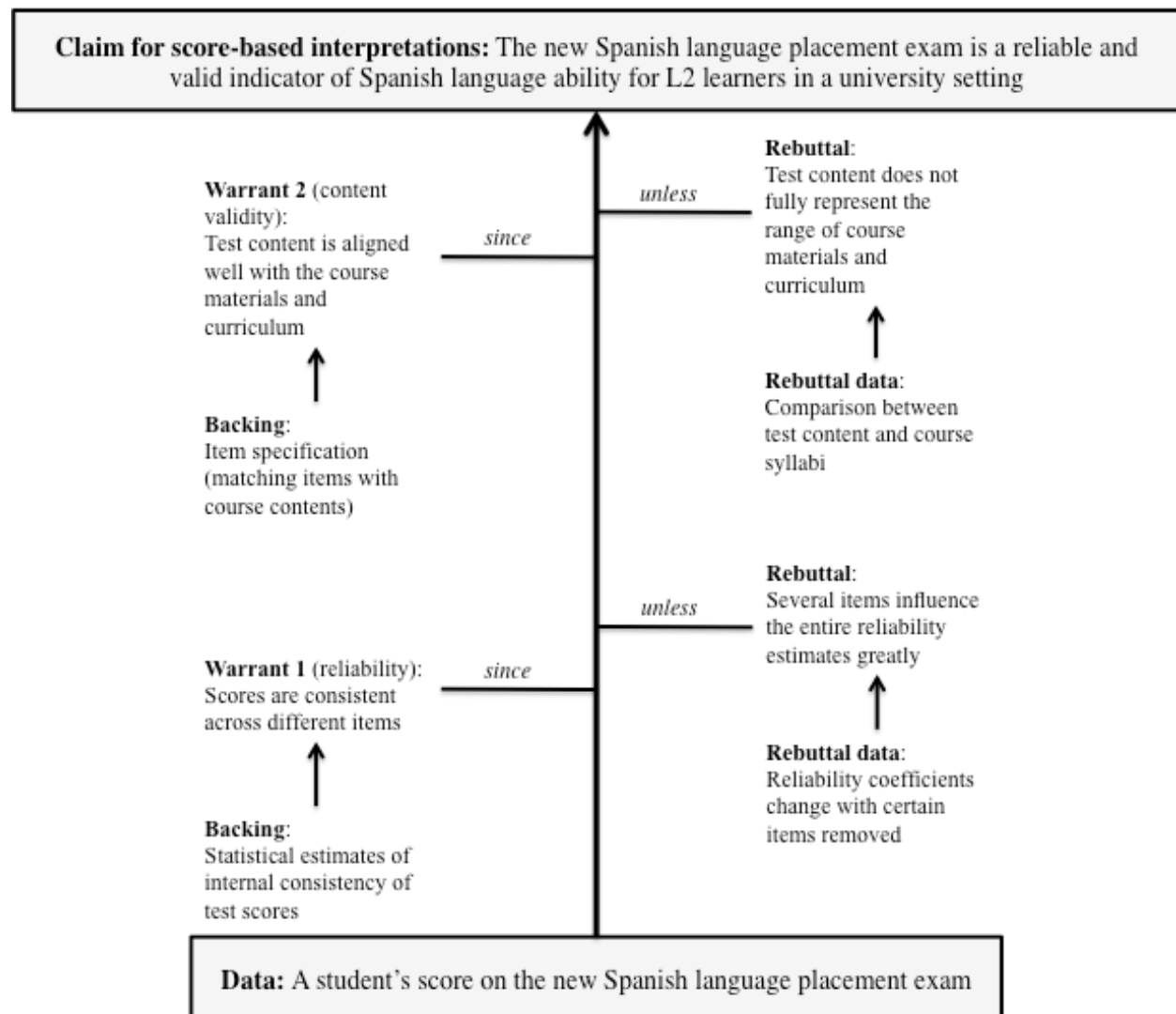


Figure 4. Test validity argument for the Spanish language placement exam

Next is the utilization argument, which refers to the link that is made between the score-based interpretation claim and use of the test (see Figure 5). In making this link, the desired claim was that the placement exam placed students into Spanish courses that were appropriate for their knowledge of Spanish as a FL. Although there were several approaches to validating this claim, we chose to compare the placement decisions of scores based on the new, web-based placement test of the present study with the placement decisions of scores based on the previous, paper-based test used in the present testing context (Warrant 3). The rationale for this method of validation was that placement decisions made based on the scores of the previous, paper-based test yielded a low rate of misplacement in the current departmental context each academic year (Way & McClarty, 2012). Thus, agreement between the placement decisions made based on the scores gained by students completing both tests provided validity for the use of the web-based Spanish language placement exam as a tool for determining which Spanish course would be appropriate for enrollment by test takers. Our backings for this warrant included statistical estimates of the extent to which scores on the new test correlated with scores on the previous test (Backing 1) as well as the extent of agreement between placement decisions made based on the scores of the new test versus those based on scores of the previous test (Backing 2). However, a potential source of weakening of this argument would be a relatively high rate of course withdrawals or replacements based on the scores generated by the new placement test (Rebuttal 3). Therefore, we examined the

number and nature of reported withdrawals and replacements resulting from placement decisions based on the web-based test scores (Rebuttal data 3).

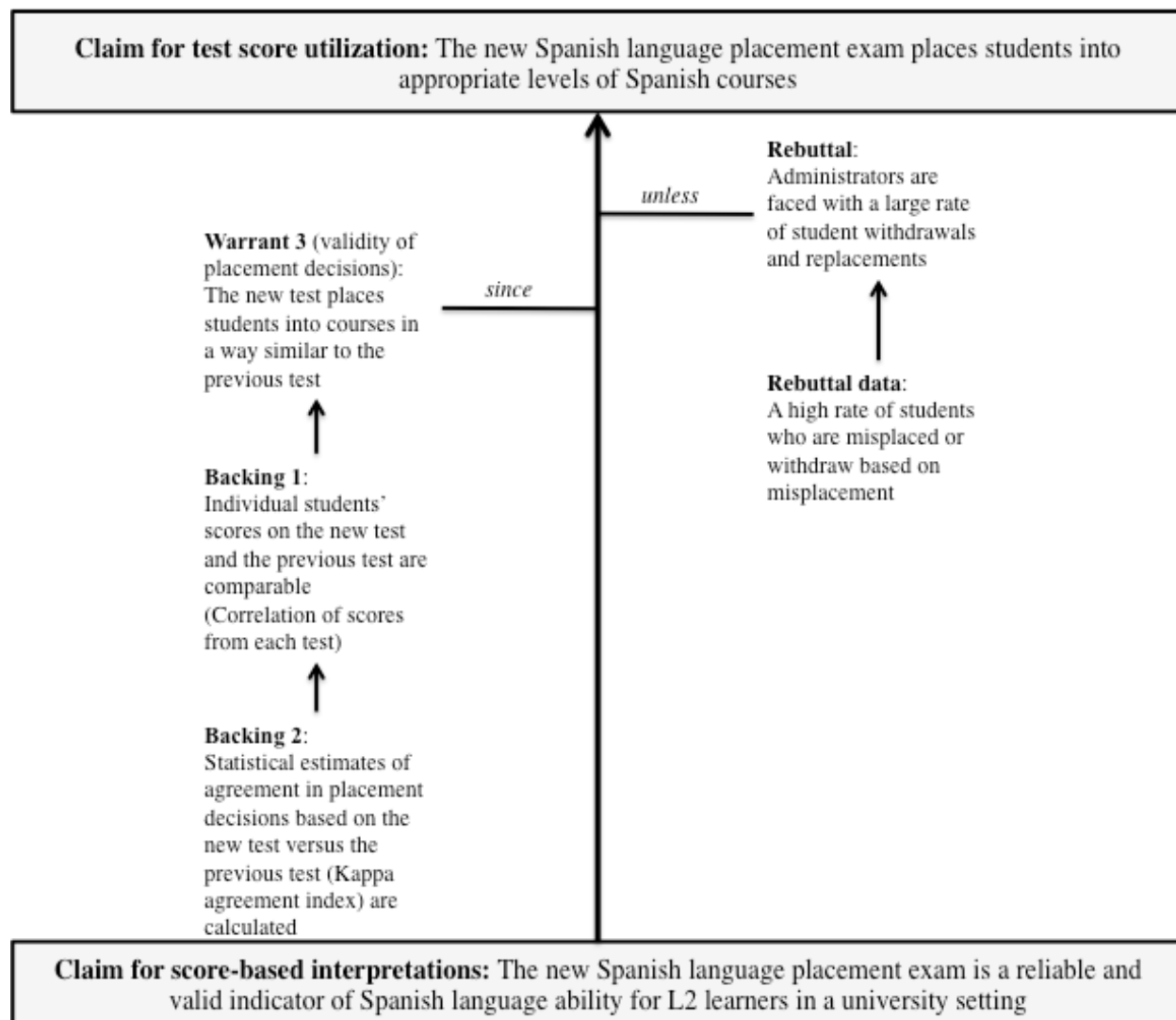


Figure 5. Test utilization argument for Spanish language placement exam

Test Scoring and Data Analysis

Students' responses to each item of the web-based Spanish placement test were scored electronically. A correct response received 1 point, whereas an incorrect response received 0 points, for a maximum possible score of 100. No partial credit was given in the scoring of students' responses to each test item. Students' test scores were computed as the sum of all correct responses to each item of the test.

Cronbach's alpha was examined as the statistical measure of internal consistency (Cronbach, 1951) for the test scores of the 2,201 test takers. Reliability coefficients were also examined to identify items that may have influenced the internal consistency of the test to a greater degree. Specifically, we examined Cronbach's alpha for each item of the test if it were to be removed from the test.

Following each analysis of internal consistency just described, two analyses of validity were conducted. The first examined test content in relation to curriculum content, and the second examined the nature of placement decisions generated by the scores of the new, web-based test. Regarding the assessment of content validity, we manually examined the content of the textbooks for each course students could be

placed into. The assigned texts for each course (at the time of the present study) that we examined were *Vistazos: Un curso breve* (VanPatten, Lee, Ballman, & Farley, 2010) for First Year Spanish; *¡Anda! Curso Intermedio* (Heining-Boynton, Leloup, & Cowell, 2010a) for Second Year Spanish I; *¡Anda! Curso Intermedio Volume II* (Heining-Boynton, Leloup, & Cowell, 2010b) for Second Year Spanish II; and *Spanish grammar in review* (Holton, Hadlich, & Gómez-Estrada, 2001), *Lazos: Gramática y vocabulario a través de la literatura* (Frantzen, 2009), and *Spanish grammar: A quick reference* (Wren, 2005) for Spanish Grammar in Context. With regard to the second method of test score validation, we compared performance and placement decisions based on the new, web-based Spanish placement test with those of the previous, paper-based test used in the institutional context of the current study. Both the scores and placement decisions of a subset of 1,622 students who took the paper-based test and the web-based test were examined by means of correlation analysis (Pearson r) and Kappa's agreement index (Cohen, 1968), respectively. Kappa's agreement index provided a statistical measure of the degree of dependability of the two testing procedures, which differed from reliability in that the consistency of absolute decisions (i.e., placing a student in one course and not another) was under examination (Carr, 2011; Kane & Brennan, 1980).

Results

This section is organized to address each warrant outlined in the AUA model developed for the web-based Spanish placement test. Thus, we report the findings first for reliability (Warrant 1), followed by content validity (Warrant 2) and validity of placement decisions (Warrant 3).

Warrant 1: Reliability

The analysis of internal consistency revealed the entire test ($k = 100$) to be very reliable, $\alpha = .92$ ($N = 2,201$). As indicated in our validation framework, there may have been items that greatly influenced estimates of reliability. To address this rebuttal, we examined reliability coefficients to identify those items that may have had a greater influence on the overall internal consistency of the test relative to other test items. With regard to this measure, we found that no single item greatly influenced estimates of the test's internal consistency. Specifically, hypothetically removing any item of the test did not result in change of the test's overall reliability by more than .002. This finding suggests that the test was internally consistent.

Examining the findings for reliability by test section, in contrast, it was revealed that only the grammar section ($k = 65$) was reliable, $\alpha = .88$. The other sections were not found to be as internally consistent as the grammar subsection, possibly due to the low number of test items in each of these subsections: sound discrimination, $\alpha = .29$ ($k = 5$); listening comprehension, $\alpha = .57$ ($k = 13$); reading comprehension, $\alpha = .50$ ($k = 7$); and vocabulary, $\alpha = .62$ ($k = 10$).

Warrant 2: Content Validity

The analysis of content validity revealed that the content of the test reflected the content incorporated in the course materials and curriculum within the departmental context of the present study. Specifically, the test sections (i.e., grammar, listening comprehension, vocabulary, etc.) were very representative of the skill sets on which students received practice and instruction. For example, the specific sounds targeted in the sound discrimination section were the same sounds that received attention in students' electronic workbook that was to be completed outside of class. With regard to grammar, the target structures included in this section of the test received some focus in terms of instruction and assessment in at least two of the six courses into which students could be placed (see [Table 1](#)). For example, the copula contrast was a structure in the grammar section of the test that received instruction and assessment in Elementary Spanish, First Year Spanish, and Spanish Grammar in Context. Aspect received instruction and assessment in Elementary Spanish II, First Year Spanish, and Spanish Grammar in Context.

For listening comprehension, the types of listening passages found in the test were generally

representative of passage types found in course materials. Although a combination of monologic and dialogic tasks could be found in the course materials, a greater proportion of monologic tasks was evident in the classroom materials and assessment, suggesting that the placement test closely reflected classroom materials in terms of listening passage types. Further, the content of the listening passages accurately reflected the content of listening passages encountered in course materials, which included (but was not limited to) descriptions of people, places, and things; narrations of past events; telephone messages; radio announcements; and dialogues between coworkers and friends.

With regard to reading comprehension, the test similarly reflected content present in course materials. For reading comprehension, an important observation was that the variable length of reading passages was representative of the variable lengths of reading passages that could be encountered in course materials. Reading passages in course textbooks ranged from as few as five lines to well over 100 lines in upper-level courses (e.g., Second Year Spanish II and Spanish Grammar in Context). Reading passages in assessments, in contrast, tended to range between 15 and 25 lines, which was more similar to passage lengths in the placement test. With regard to reading passage content, students were exposed to a variety of topics in courses: daily life, health, Spanish-speaking cultures, history of Spanish-speaking countries, notable people and places in the Spanish-speaking world, and several pieces of fiction, to name just a few. Two of the preceding topics (i.e., lifestyle and history) were used in the placement test.

Lastly, for vocabulary, the words included on the test were also present in the materials of several courses into which students could be placed. For example, vocabulary related to describing people and places (very generally) was present in both course materials and the placement test. In addition, the way in which vocabulary was tested in the test was very similar to the method for testing vocabulary in courses—namely, contextualized sentences that students had to complete using a list of vocabulary words (provided in a word bank or in multiple choice format).

A potential rebuttal against the warrant that the test was valid in terms of content was that the content of the test was not fully representative of course materials or curriculum input. To that end, we examined the syllabus and textbooks for each course in terms of stated objectives and content covered for each class meeting (for a detailed summary of the objectives and content presented in each course syllabus and textbook, see [Appendix](#)). Upon examination of the course objectives and content for each course, we found that the placement test was comprehensive in its inclusion of linguistic structures and topics, but not topics related to culture. Specifically, the range of topics related to culture in test items and the reading and listening passage was not as extensive as that observed upon in-depth review of course content. Nevertheless, we do not believe that this lack of range of culture topics negatively impacts the validity of the test in terms of content given that, albeit minimally, a consideration of culture is reflected in the placement exam. Thus, it appears that, overall, test content matches content provided in course materials and, in many cases, curriculum-based assessment.

Warrant 3: Validity of Placement Decisions

A paper-based assessment was used previously to place incoming and transfer students into the courses offered by the department. To address the claim that the web-based placement test placed students into Spanish courses that were appropriate for their knowledge of Spanish as a FL, we compared the placement decisions of scores based on the new, web-based test with those of scores based on the previous, paper-based test. To investigate this comparison, we asked (a) whether or not the scores for the web-based test correlated with those of the paper-based test for each student and (b) how consistent placement into the levels determined by the web-based test were with placement by the paper-based test. With regard to the first investigation, we calculated correlations between the scores of the web-based test and the paper-based test completed by the subset of 1,622 students who took both tests. The Pearson product-moment correlation coefficient (r) was .73, indicating that the scores calculated for each student on the web-based test and the scores on the paper-based test were highly correlated. For the second investigation, we examined Kappa measures of agreement between placement decisions made based on the scores of the new, web-based test as compared to placement decisions made based on the previous,

paper-based test (again by the same subset of 1,622 students who took both tests). Depending on the score, students could be placed into one of the six courses outlined in Table 1. For the previous, paper-based test, students could be placed into Introduction to Hispanic Cultures, a different upper-division course, in addition to the six courses listed in Table 1. For the purposes of this analysis, we examined agreement measures between placement decisions for equivalent courses (i.e., courses with the same objectives and content: First Year Spanish, Second Year Spanish I, Second Year Spanish II, and Spanish Grammar in Context). The kappa coefficient was .85 (a value of 1.0 indicates perfect agreement), indicating considerably strong agreement between placement decisions made based on the scores of each test (Fleiss, 1981).

A potential rebuttal for the warrant that the web-based test placed students into courses in a manner similar to the previous test was that we encountered a high rate of course withdrawals or replacements based on the scores generated by the new placement test. In response to this rebuttal, we examined the distribution and rates of withdrawal and replacement that occurred during the fall semester following our study (see Table 2).

Table 2. *Distribution and Rates of Withdrawal and Replacement*

Course	N Withdrawals and Replacements	%
First Year Spanish	57	2.6
Second Year Spanish I	78	3.5
Second Year Spanish II	89	4.0

Note. Percentages were based on the number of withdrawals and replacements divided by the number of students enrolled in each class level.

Rates of withdrawal and replacement for each course listed in Table 2 are relatively low. On a more practical level, however, the raw number of withdrawals and replacements may suggest that the web-based test is in need of further examination. It is important to point out here that *withdrawal rates* do not necessarily reflect misplacement. However, we cannot ignore the possibility that a student was not placed into the most appropriate course for his or her level of Spanish language knowledge, but did not choose to withdraw from the course until later in the semester.⁴

To provide further insight into the withdrawal and replacement rates observed in Table 2, we consulted with administrators of the web-based placement test and the Spanish language program in which the test was used. Administrators reported that, in general, many students requested to be moved from Second Year Spanish II to Second Year Spanish I in particular because they found the course to be overwhelming. Within the departmental context of the present study, incoming students often come from high schools in which Spanish language instruction involves a strong focus on grammar or vocabulary and may not necessarily contain a strong communicative component or actual use of Spanish in the classroom. It appears that students' familiarity with grammar, which constitutes a noticeable bulk of the web-based test, results in test scores that place them into courses (such as Second Year Spanish II or even Spanish Grammar in Context) that require not only extensive knowledge of grammar and vocabulary, but also the ability to use such knowledge for communicative purposes in the classroom.

Discussion

Every year, language instructors and language program administrators face the task of developing, administering, and assessing web-based placement tests for the purpose of enrolling incoming and transfer students into courses that are suitable for their knowledge of the language being tested. However, it is also likely that these individuals lack a systematic method or procedure for undertaking a task such as the one outlined in the present study. In response to the need for a concise, accessible example of test validation from which everyday FL programs can benefit, this paper reported the findings of a study that

used Bachman's (2005) AUA framework to examine the reliability and validity of a web-based Spanish placement test.

To address the first research question of our study, we found that the web-based test was reliable, as demonstrated by means of statistical estimates of internal consistency (Cronbach's alpha and examination of reliability coefficients for each test item). In response to our second research question, we found that the web-based test was valid in terms of content and the placement decisions being made based on test scores. Nevertheless, there are areas in which the test validity could be strengthened. For example, test content could be modified to reflect the range of cultural topics observed in the course materials and syllabi of the target enrollment classes. Additionally, test developers could consider how performance demonstrated on the placement exam translates to use of the assessed language knowledge in Spanish language classrooms. Related to this point, it may be that the cut-offs for placement into the two Second Year Spanish courses need to be adjusted to address the needs of high-scoring students who have extensive knowledge of Spanish and grammar but may not actually be able to use such knowledge for communicative ends.

In keeping with the central contribution of our study and in response to the final research question guiding this investigation, we believe that the nuts and bolts of our method for test validation can be used in other university-level FL contexts. Particularly for large-scale FL programs in which administrators or coordinators need to make fairly quick decisions about which course to place students into, this report illustrates how to outline, substantiate, and provide evidence for the reliability, validity, and use of a placement test. Specifically, using our study as an example, first it must be demonstrated that the test consistently produces similar results, which can be shown by examining simple statistical measures such as Cronbach's alpha. Once the test has been determined to be reliable, it must be demonstrated that, as a measure, it is valid for its intended uses. Validity can be determined in a number of ways (for other methods not employed in the present study, see Wall et al., 1994; Fulcher, 1997), of which we examined content validity—a qualitative technique. Finally, it is important to demonstrate that actual use of the test scores is appropriate. In this study, we examined the validity of placement decisions, a technique that may not be an option for other FL contexts.⁵ In FL programs in which a previous placement test is not available (or is not in place), an alternative technique could be the yes/no Angoff method (Hsieh, 2013) that allows instructors—who oftentimes have a good sense of whether or not a given student belongs in a course—to establish several cut-off points in a single test by indicating at what point (in a test with items arranged from easiest to hardest) a borderline student would be able to answer a test item correctly.⁶

Another point of discussion that is important to raise here is the pedagogical impact of designing and implementing a web-based test. Perhaps the most notable impact in the departmental context of the present study is the ability to keep test and course content more closely matched over time. The web-based test offers the flexibility to update test content as minor adjustments to curriculum are made over time. Additionally, in a society that has been witness to an ever-increasing use of computer- and web-based technologies, using an online test has the potential to deliver input and content in a manner that not only is familiar to language learners, but also mirrors the use of technology that learners are exposed to in the classroom. For instance, within the institutional context of the present study, instructors are increasingly moving to a web-based format for the design and delivery of course examinations. This pedagogical practice is one that has likely been influenced by current placement testing practices, which has proven to be an accessible and flexible means for test delivery.

Conclusion

This study has provided an examination of the evidence to support claims made about the intended interpretations and uses of scores based on a new web-based Spanish language placement test. Specifically, we employed an argument-based approach (Bachman, 2005) to objectively and logically link students' scores on the test to actual placement in Spanish FL courses based on those scores. In doing so, we offer an accessible, replicable example for test validation to the many researchers and educators

involved in the process of creating and assessing FL placement tests in similar institutions across the United States. The following three important limitations of our study should be considered. First, although a more thorough discussion of test development fell outside the scope of this study, this aspect of language placement testing should not be ignored when assessing test reliability and validity. Administrators and instructors who adapt our method for placement testing assessment should examine test development carefully—specifically, the procedure for setting cut-off scores, the relationship between any previous tests and the current test under development, and analyses of test scores (overall and by subsections, if applicable)—as a necessary prerequisite to investigating test reliability and validity (for an overview, see Bachman & Palmer, 1996). Second, issues of security are equally crucial to consider in the design and assessment of an online language test, particularly in medium- to high-stakes testing settings.⁷ Administrators and instructors must consider this issue in the logistics of testing administration and address all potential concerns in the design, development, and use of online placement tests.⁸ Third, it is important to note that Bachman's (2005) AUA model is quite exhaustive, covering a wide range of aspects of validity claims. Thus, there are other types of warrants and required backing and rebuttals not addressed in this study. For example, warrants and claims related to construct validity and relevance, utility, intended consequences, and sufficiency in the assessment utilization argument are not made or investigated. Such validity evidence should be further collected to support the claims about the intended test score interpretations and uses particularly for a high-stakes testing context in future studies.

Acknowledgements

We would like to acknowledge the student participants without whom this study would not have been possible. We also wish to thank Stephanie Dickinson for her assistance with statistics in an earlier version of this manuscript. Finally, we are grateful to the members of the SLRF 2013 audience and the Language Assessment Lab at Indiana University for their insightful comments and suggestions. All remaining errors are our own.

Notes

1. The items in the grammar section of the placement test were not evenly distributed by structure type (i.e., there was not an even number of copula, mood, aspect, etc. items).
2. 23 students did not report their native language.
3. As pointed out by an anonymous reviewer, Bachman and Palmer (2010) broke down AUA components into four different claims (assessment records, interpretation, decisions, and consequences), but our study is based on Bachman's original AUA framework (Bachman, 2005). Note that there are no substantial differences between the two models, and his original one (Bachman, 2005) has been more widely used, which makes this study more comparable with other studies using the AUA model for test validation.
4. As pointed out by an anonymous reviewer, comparisons drawn between withdrawal or replacement rates for the paper- and web-based tests can be stronger with additional statistical support. Future studies should strive to incorporate statistical comparisons where possible.
5. Additionally, as pointed out by an anonymous reviewer, it is important to address qualitative-based placement decisions (e.g., in the case of students who indicate that they have been misplaced and wish to be placed into a different course) in a reliable manner. Although we would like to provide a systematic decision procedure for such cases, this procedure may not be feasible across institutional settings, and it is likely that cases of replacement will need to be addressed on a case-by-case basis. However, a high number of misplacements may suggest the need for further evaluation (or reevaluation) of placement decisions based on tests scores (i.e., the quantitative-based method).
6. The yes/no Angoff method is not without its limitations (for a review, see Cizek & Bunch, 2007);

however, it is logical and practical, and it has been shown to be used with confidence (Hsieh, 2013).

7. Concerns about security are related to test reliability and validity, as pointed out by an anonymous reviewer. In our particular institutional context, students do not automatically receive credit for past Spanish courses. Thus, it is not in students' best interest to cheat, as doing so would place them in a level beyond their ability. This is one method for dealing with the difficulty and cost of creating a secure exam, but security nonetheless should factor into the evaluation of a web-based placement test.
8. An additional limitation of this study pointed out by an anonymous reviewer is the lack of interviews with administrators or students regarding withdrawals from courses in which students were placed. This is an excellent suggestion that we recommend testing administrators incorporate into their testing evaluation practices.

References

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, UK: Oxford University Press.
- Bachman, L. F. (2004). *Statistical analysis for language assessment*. Cambridge, UK: Cambridge University Press.
- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2(1), 1–34.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, UK: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford, UK: Oxford University Press.
- Bardovi-Harlig, K., & Shin, S.-Y. (2014). Expanding traditional testing measures with tasks from L2 pragmatics research. *Iranian Journal of Language Testing*, 4, 26–49.
- Bernhardt, E. B., Rivera, R. J., & Kamil, M. L. (2004). The practicality and efficiency of web-based placement testing for college-level language programs. *Foreign Language Annals*, 37(3), 356–365.
- Bernstein, J., Van Moere, A., & Cheng, J. (2010). Validating automated speaking tests. *Language Testing*, 27, 355–377.
- Brown, J. D. (1989). Improving ESL placement tests using two perspectives. *TESOL Quarterly*, 23(1), 65–83.
- Brown, J. D., & Hudson, T. (2002). *Criterion-referenced language testing*. Cambridge, UK: Cambridge University Press.
- Carr, N. T. (2011). *Designing and analyzing language tests*. Oxford, UK: Oxford University Press.
- Carr, N. T., & Xi, X. (2010). Automated scoring of short-answer reading items: Implications for constructs. *Language Assessment Quarterly*, 7, 205–218.
- Chapelle, C. A., & Douglas, D. (2006). *Assessing language through computer technology*. Cambridge, UK: Cambridge University Press.
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220.

- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- Davies, M. (2006). *A frequency dictionary of Spanish: Core vocabulary for learners*. New York, NY: Routledge.
- Eda, S., Itomitsu, M., & Noda, M. (2008). The Japanese skills test as an on-demand placement test: Validity comparisons and reliability. *Foreign Language Annals*, 41(2), 218–236.
- Elder, C., & Randow, J. V. (2008). Exploring the utility of a web-based English language screening tool. *Language Assessment Quarterly*, 5, 173–194.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York, NY: John Wiley.
- Frantzen, D. (2009). *Lazos: Gramática y vocabulario a través de la literatura*. Upper Saddle River, NJ: Prentice Hall.
- Fulcher, G. (1997). An English language placement test: Issues in reliability and validity. *Language Testing*, 14(2), 113–138.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. New York, NY: Routledge.
- Geeslin, K. L. (Ed.). (2013). *The handbook of Spanish second language acquisition*. Malden, MA: Wiley.
- Green, A. (2012). Placement testing. In C. Coombe, B. O’Sullivan, P. Davidson, & S. Stoyanoff (Eds.), *The Cambridge guide to language assessment* (pp. 164–170). Cambridge, UK: Cambridge University Press.
- Heining-Boynton, A. L., Leloup, J. W., & Cowell, G. S. (2010a). *¡Anda! Curso intermedio*. Upper Saddle River, NJ: Prentice Hall.
- Heining-Boynton, A. L., Leloup, J. W., & Cowell, G. S. (2010b). *¡Anda! Curso intermedio Volume II*. Upper Saddle River, NJ: Prentice Hall.
- Holton, J. S., Hadlich, R. L., & Gómez-Estrada, N. (2001). *Spanish grammar in review* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Hsieh, M. (2013). Comparing yes/no Angoff and Bookmark standard setting methods in the context of English assessment. *Language Assessment Quarterly*, 10, 331–350.
- Huff, K. L., & Sireci, S. G. (2001). Validity issues in computer-based testing. *Educational Measurement: Issues and Practices*, 20, 16–25.
- Jia, Y. (2013). *Justifying the use of a second language oral test as an exit test in Hong Kong: An application of assessment use argument framework* (Unpublished doctoral dissertation). University of California, Los Angeles, Los Angeles, CA.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527–535.
- Kane, M. T. (2011). Validating score interpretations and uses: Messick Lecture, Language Testing Research Colloquium, Cambridge, April 2010. *Language Testing*, 29(1), 3–17.
- Kane, M. T., & Brennan, R. L. (1980). Agreement coefficients as indices of dependability for domain-referenced tests. *Applied Psychological Measurement*, 4, 105–126.
- Kline, P. (2000). *The handbook of psychological testing* (2nd ed.). London, UK: Routledge.
- Llosa, L. (2008). Building and supporting a validity argument for a standards-based classroom assessment of English proficiency based on teacher judgments. *Educational Measurement: Issues and Practice*, 27(3), 32–42.

- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York, NY: American Council on Education and Macmillan Publishing Company.
- Ockey, G. J. (2007). Construct implications of including still image or video in computer-based listening tests. *Language Testing*, 24, 517–537.
- Ockey, G. J. (2009). Developments and challenges in the use of computer-based testing for assessing second language ability. *Modern Language Journal*, 93, 836–847.
- Pardo-Ballester, C. (2010). The validity argument of a web-based Spanish listening exam: Test usefulness evaluation. *Language Assessment Quarterly*, 7, 137–159.
- Roever, C. (2001). Web-based language testing. *Language Learning & Technology*, 5(2), 84–94. <https://dx.doi.org/10125/25129>
- Shin, S.-Y. (2008). Examining the construct validity of a web-based academic listening test: An investigation of the effects of constructed response formats in a listening test. *The Spaan Fellowship Working Papers in Second or Foreign Language*, 6, 95–129.
- Shin, S.-Y. (2012). Web-based language testing. In C. Coombe, B. O'Sullivan, P. Davidson, & S. Stoyonoff (Eds.), *The Cambridge guide to language assessment* (pp. 274–279). Cambridge, UK: Cambridge University Press.
- Toulmin, S. E. (2003). *The uses of argument* (2nd ed.). Cambridge, UK: Cambridge University Press.
- University of Wisconsin Center for Placement Testing. (2011). *Spanish placement test*. Retrieved from <http://testing.wisc.edu/centerpages/spanishtest.html>
- VanPatten, B., Lee, J. F., Ballman, T. L., & Farley, A. P. (2010). *Vistazos: Un curso breve* (3rd ed.). Boston, MA: McGraw-Hill.
- Wagner, E. (2010). The effect of the use of video texts on ESL listening test-taker performance. *Language Testing*, 27, 493–513.
- Wall, D., Clapham, C., & Alderson, J. C. (1994). Evaluating a placement test. *Language Testing*, 11(3), 321–344.
- Wang, H., Choi, I., Schmidgall, J., & Bachman, L. F. (2012). Review of Pearson Test of English Academic: Building an assessment use argument. *Language Testing*, 29(4), 603–619.
- Way, W. D., & McClarty, K. L. (2012). Standard setting for computer-based assessments. In G. J. Cizek (Ed.), *Setting performance standards* (2nd ed., pp. 451–466). New York, NY: Taylor & Francis.
- Williamson, D. M., Bejar, I. I., & Sax, A. (2004). Automated tools for subject matter expert evaluation of automated scoring. *Applied Measurement in Education*, 17, 323–357.
- Wren, D. (2005). *Spanish grammar: A quick reference* (2nd ed.). New York, NY: Pearson Education.

Appendix. Summary of Course Objectives and Content for First Year Spanish, Second Year Spanish I, Second Year Spanish II, and Spanish Grammar in Context

Course	Summary of Course Objectives	Summary of Content for All Class Meetings
First Year Spanish ^a	Describe, narrate, ask, and answer questions related to everyday topics in Spanish; comprehend main idea and some supporting details in spoken Spanish related to authentic situations in daily life; write short texts on familiar topics; recognize cross-cultural differences and appreciate perspectives and peoples of the Spanish-speaking world	<p>Topics: self-introduction; school and studying; daily routines; weather; family; age; describing and comparing people; food and eating habits; health; culture-specific examples of topics previously listed</p> <p>Linguistic structures: <i>ser</i> (to be); <i>hay</i> (there is, there are); present tense; hobbies and activities; preterit; <i>estar</i> (to be) + adjectives; direct and indirect object pronouns; comparatives and superlatives; reflexives; passive constructions; <i>gustar</i>-type verbs (to be pleasing)</p>
Second Year Spanish I	Improve basic communicative competence in Spanish, including listening, speaking, reading, and writing; provide practical vocabulary and communicative situations; move students to higher proficiency levels; focus on contemporary Spanish-speaking culture; increase awareness of cross-cultural differences of perspectives and peoples of the Spanish-speaking world	<p>Topics: descriptions of people and personality; pastimes; objects and activities in and around the home; celebrations and events; traveling; culture-specific examples of topics previously listed</p> <p>Linguistic structures: <i>gustar</i>-type verbs; preterit; present perfect; present subjunctive; formal and informal imperatives; definite and indefinite articles; <i>estar</i> + past participles; preterit vs. imperfect; past perfect; relative pronouns</p>
Second Year Spanish II	Improve basic communicative competence in Spanish, including listening, speaking, reading, and writing; provide practical vocabulary and communicative situations; move students to higher proficiency levels; focus on contemporary Spanish-speaking culture; increase awareness of cross-cultural differences of perspectives and peoples of the Spanish-speaking world	<p>Topics: environment; professional settings and the workplace; art; objects and living things of the world; health; culture-specific examples of topics previously listed</p> <p>Linguistic structures: <i>ser</i> and <i>estar</i>; subjunctive; progressive tenses; future tense; conditional tense; demonstrative adjectives; <i>if</i> clauses; prepositions; imperfect subjunctive; past perfect subjunctive; reflexive verbs; reciprocal and passive constructions</p>

Spanish Grammar in Context	Strengthen comprehension and application of challenging grammatical structures within a meaningful language context	<p>Topics: expressing likes and preferences; describing; narrating in the past; plans and dreams for the future; talking about probability in the past and present; expressing desires, expectations, and emotional reactions; talking about the passing of time; argumentative discourse</p> <p>Linguistic structures: descriptive adjectives and their placement; <i>ser</i> vs. <i>estar</i> vs. <i>haber</i> (to exist); preterit vs. imperfect of <i>saber</i> (to know), <i>conocer</i> (to know), and other verbs; subjunctive vs. indicative in all clause types; uses of the future and the conditional tenses; subjunctive and sequence of tenses; expressions of time with <i>hacer</i> (to make, to do); <i>por</i> vs. <i>para</i>; passive constructions; <i>if</i> clauses</p>
----------------------------	---	---

^aBecause this course covers all objectives and content of Elementary Spanish and Elementary Spanish II, the objectives and content of those two courses are not provided here.

About the Authors

Avizia Y. Long (PhD, Indiana University) is Visiting Assistant Professor of Spanish at Texas Tech University. Her research interests include (L2) variation, L2 phonetics and phonology, and pronunciation in task-based language teaching. She is co-author of *Sociolinguistics and Second Language Acquisition: Learning to Use Language in Context* (Routledge, 2014).

E-mail: avizia.long@ttu.edu

Sun-Young Shin is Associate Professor in the Department of Second Language Studies at Indiana University. His research interests include test bias, standard setting, and rater behaviors. His work has been published in *Language Testing*, *Language Teaching Research*, *Assessing Writing*, and *ReCALL*.

E-mail: shin36@indiana.edu

Kimberly L. Geeslin (PhD, University of Arizona) is Professor at Indiana University. She co-authored *Sociolinguistics and Second Language Acquisition* (Routledge, 2014) and edited *The Handbook of Spanish Second Language Acquisition* (Wiley-Blackwell, 2013). Her research appears in *Studies in Second Language Acquisition*, *Language Learning*, *Hispania*, *Spanish in Context*, *Linguistics*, and *SHLL*.

E-mail: kgeeslin@indiana.edu

Erik W. Willis is Associate Professor of Hispanic Linguistics at Indiana University. His research investigates Spanish phonetics and phonology from a laboratory approach and uses acoustic measures to characterize the sounds and systems in question. His research primarily investigates variation in the Spanish system, but also includes learner populations.

E-mail: ewwillis@indiana.edu