

Prueba de Español

Walter Rudolfo Archuleta

The University of New Mexico

Introduction

Prueba de Español para la Certificación Bilingüe is an instrument designed to measure the Spanish language proficiency of prospective bilingual education teachers in New Mexico. Specifically, it examines the performance on the *Prueba* of a norming sample of 120 subjects, ranging from students in bilingual education teacher training programs to experienced bilingual education teachers.

1. Historical overview of the legislation

In 1969 state legislation was enacted, Senate Bill 270 became the first bilingual law in the nation. In 1971, House bill 270 added \$100,000 funding for K-3 only. In 1973, Senate Bill 421 provided bilingual education for elementary students. The law also contained equal educational opportunity for limited English proficient students. In 1979, the State Board of Education adopted certification requirements for an endorsement in bilingual education. In 1980, the Four Skills Exam was developed to assess teacher's ability in Spanish. In 1987, the expansion of the bilingual law to the twelfth grade. Finally in 1994, in response to concerns regarding teacher preparation for bilingual education, the state legislature appropriated funding to develop a new language proficiency exam.

In 1995, the State Department of Education (SDE) began to work on the implementation of HB 224. The SDE hired a Bilingual Education Assessment Specialist. The revisions of the Four Skills Exam (FSE) turned out to be the development of a new assessment instrument. A task force comprised of experts in the areas of Spanish language, bilingual education, linguistics, test development and assessment and evaluation was assembled.

Initially the task force was engaged in the following activities:

*Reviewed Dr. Guerrero's doctoral dissertation, entitled *A Critical Analysis of the Validity of the Four Skills Exam* (1994)

*Reviewed other states' proficiency tests for certification of bilingual education teachers.

*Preliminary development of a conceptual framework for a new exam that addressed identified inadequacies of the FSE which revolved around the issue of competency requirements. While a minimal level of eighth grade level proficiency was deemed essential for bilingual endorsement, the FSE required only a fourth grade level of proficiency and did not address the state competencies for native language proficiency.

A preliminary draft of the exam contained the following key elements:

*It was tied to actual classroom practices teachers are engaged in on a daily basis.

*It encompassed all the state competencies required by the state for an endorsement in bilingual education.

* It reflected four levels of Spanish language proficiency and literacy.

* It integrated both vernacular Spanish spoken in New Mexico and the standard form of Spanish.

*It was a highly contextualized exam. For example, teachers might correct a letter written to the parents of a student written in the native language or read aloud in the natural sciences.

* It incorporated video and taped voices of New Mexican Speakers of Spanish in natural language contexts.

2. History of activities in the development of Prueba

The task force was instrumental in the development of *Prueba*. A brief overview of development activities follows:

- *Preliminary draft of exam is developed, 1995-1996
- *Pre-piloting of exam-20 examinees, February, 1996
- *Testimony and status report to LESC, Fall, 1996
- *Piloting of exam-120 examinees, April 12, 19, 26, June 7, July 8, 1997

Five evaluators were selected and trained to score the exam. Selections were made based on the individuals' educational background, experience, knowledge of bilingual education and their proficiency in Spanish, 1998

*Development of exam manuals: Manual de Estudios, Scoring & Evaluation Booklet, Technical Manual, Administrator's Manual, 1997-1998

- *Printing of all materials, Summer & Fall, 1998
- *Exam is finalized, 1998
- *Identification of testing center and testing sites, 1998
- *Implementation and Administration begin, 1999

*The examination was administered to 288 examinees. Test dates were May 15, May 29, June 12, October 9, October 16, November 5, November 6, 1999. SDE collected data and reviewed ongoing examination results.

*Recommendation for the examination's passing score based on the evaluation results, must score a 2 (proficient level) on 12 out of 15 subsections, 2000

- *Administration continues, 2003
- *Exam was evaluated by three independent evaluators

3. Congruency between Prueba and the native language competencies for bilingual education

A language matrix for *Prueba* which included school contexts and the four language skills is congruent with the native language competencies for bilingual education. For example, in comunicaciones escolares, given a report card as a prompt, an examinee explains progress to parents. This addresses native language competencies (1) a, b, c,d, and g.

4. Main features

The main features of *Prueba* are the following: 1) Language tasks are linked to school contexts: Comunicaciones escolares, Ciencias naturales, Lengua y cultura, and Estudios sociales; 2) Language tasks are built around authentic situations and prompts: Report card, telephone message, Principal's memo, Reading from actual textbooks, and Video; 3) Language tasks integrate four language skills: Reading and speaking, Listening and writing, Reading and writing, and Listening and speaking; 4) Language tasks are linked to NM mandated competencies. Enhances the validity of the test, Focus on content areas and Focus on eighth grade level of proficiency.

5. Content

The content of *Prueba* is divided into four parts. The first part is School Communication. Under this part there is correction of correspondence, conference with the parents, and response to a telephone message. Part two is Language and Culture. There is a video on the rural life of New Mexico. In part two there is an essay on the video and meaning of sentences. Part three is Social Studies. There is an article on the New Mexicans. This part addresses summary of a text and constructing questions about the text. Part four deals with the Natural Sciences. The main topic is the respiratory system. Under part four there is reading aloud, a dictation and a "cloze" test.

6. Description of Prueba

In describing *Prueba* there are four parts. Part one, School Communication, is comprised of three subtests, each built around typical events that occur in communication between parents and teachers. The first subtest consists of a letter sent home from school to parents which assesses editing

skills. The second subtest is configured around a parent/teacher conference which assesses oral language. The third subtest elicits a writing sample by requesting a written note to a parent in response to a telephone message.

Part two, Language and Culture, consists of a video about rural life in New Mexico. Two subtests are built around a video. Each subtest assesses understanding of life in rural New Mexico. The first subtest requires a written essay based on the video. The second uses a multiple choice format to assess listening comprehension.

Part three, Social Studies, is about the New Mexicans. Two subtests assess oral and written production within the social studies content area. The first subtest requires the examinee to produce an oral language sample. The second requires a written response.

Part four, the Natural Sciences, the Respiratory System. Three subtests collectively assess reading comprehension, oral reading, listening comprehension and writing within the science content area. Each subtest is based on a separate text. The first subtest requires the examinee to produce a sample of oral reading, the second requires listening and writing, and the third assesses reading comprehension within a "cloze" format.

7. Methodology

The exam was administered to a sizable norming sample of 120 persons whose backgrounds are appropriate for the purposes of the exam. Of the norming sample, 93 subjects were female and 27 subjects were male. The main measure used for this study was *Prueba* which has been described above. In addition to *Prueba*, one open-ended questionnaire was also used addressing the question, To what do you attribute your Spanish language skills? This is the qualitative portion of the exam

8. Sociodemographic variables

With regards to *Prueba*, the sociodemographic variables are gender, age ethnicity, geographical location where raised, self-evaluation or self-rating, experience in a Spanish-speaking country and major area of study. Gender addressed male or female. Age was broken down into 3 categories, 20-39, 40-49 and 50-59. Ethnicity addresses Hispanic, Spanish-American, Chicano, Mexican-American, Other and Anglo. Geographical location where raised deals with New Mexico, Other parts of the U.S. and Outside the U.S. Self-Evaluation of Spanish Proficiency addresses beginner's, adequate (or proficient), advanced and superior. Experience in a Spanish-speaking country is broken down into three categories, 0 months, 1-11 months and 12 months or longer. In this study, the sociodemographic variables were treated as independent variables tied to seven hypotheses.

To briefly summarize the sociodemographic characteristics, the testing population largely female. The norming sample is comprised of mature, experienced adults (72%) over the age of 40. More than half of the examinees were raised speaking Spanish and continue to speak it. Nearly 70% were raised in New Mexico and nearly 60% consider themselves Hispanic or Spanish-American. Nine out of ten rated their Spanish as at least adequate. Nearly half had never been to a Spanish speaking country and approximately one-third majored in Spanish or Spanish/Bilingual Education combination.

9. Test variables

There are 15 test variables. Five address writing skills, four address reading comprehension, two address oral skills or oral production, and one addresses reading aloud. The test variables are Correction of Exam, Response to a Telephone Message, Written Essay, Formulation of Questions, and Dictation which all address **Writing**. Conference with Parents and Summary of Text both address **Oral Production**. Conference with Parents, Formulation of Questions, Summary of Text and Cloze Test address **Written Comprehension**. Response to Message, Written Essay, and Meaning of Sentences addresses **Listening Comprehension**. Reading Aloud addresses **Oral Reading**. The two most difficult parts of *Prueba* are Correction of Exam (72.9%) and Cloze Test (43.3%) scored at the novice level. The most difficult skill areas appear to be Writing/Editing and Reading Comprehension. The two easiest parts of *Prueba* are Response to Message (9.2%) and Dictation (10%). The easiest

skill areas appear to be Listening Comprehension and Writing based on listening comprehension. Just as would be expected Listening Comprehension is superior to Oral Production, whereas Oral Production is easier than Writing (the written language) for native Spanish-speakers due to lack of none or minimal Spanish language instruction.

10. Reliability

Split half-reliability was calculated to determine the extent to which items are consistent to the test sections. While most of the coefficients exceeded the general accepted criterion of .80, one section, *Segunda Parte B: Significado de Oraciones*, was .53. This results indicated a section behaves differently than the other sections. Alderson, Clapham, and Wall (1995) note that "... it is more common to give the test only once, and to measure what is called *inter-item consistency*. One way of doing this is to stimulate the parallel forms method by calculating the *split half reliability index*"(p. 88). Thus, regarding test reliability, the test development team has analyzed the consistency of the test appropriately, and to the extent possible, given the constraints of time and money.

11. Validity

Efforts to validate *Prueba* draw on evidence regarding the construct measures, the content selected for measuring the construct, and input from the team of experts who developed the test.

Construct validity lies in the view of language proficiency that the test team adopted and used to create an exam aligned with the state competencies. Because language is not a set of isolated skills the test development team devised integrated test tasks which reflect, to the extent possible in a test, how language and literacy are integrated in a classroom context. For example in one section of the test, the teacher must *listen* to a phone message about a student's report card and write the parent a note to answer his or her questions. Teachers must also *read* a social studies text excerpt and write appropriate text questions for the students. Such tasks, while challenging from a measurement perspective because of their integrated nature (Lewkowicz, 1997), are really state-of-the-art measurement in language testing. In terms of construct validity, *Prueba* rates highly, and might even be considered a *model* for teacher language testing in bilingual education. The test evaluators of *Prueba* found that New Mexico is far ahead of other states in regards to bilingual certification/endorsement with respect to teacher language assessment. According to Grant (1997), some states have no testing at all, some leave testing up to the teacher preparation institutions, some use a single general measure but only a few test proficiency in the context of teaching. The major benefit of testing language proficiency in the teaching context is the *match* between the test and the real-life setting or target language use (TLU) domain described by language testing experts Bachman and Palmer (1996).

Content validity addresses how well the content reflects a sampling of the kinds of tasks required in the domain of bilingual education. The test team, made up of bilingual education experts, determined the task according to what they knew about bilingual education and what they determined to be feasible to test. The test evaluators review of the content indicated a close alignment between the tasks required of bilingual teachers and the test tasks themselves; such evidence is important in providing evidence for test validity and overall test usefulness (Bachman and Palmer, 1996).

Regarding criterion-related evidence for validity, lack of time and evidence prevented them from gathering evidence of this nature. However concurrent validity has been gathered in my doctoral dissertation by examining the relationships between the self assessment done by examinees on the questionnaires given prior to the test and the test performance of those same examinees. I found that those who rated themselves higher in terms of language ability did in fact do better on the test. Those who rated themselves lower performed worse on the test.

12. Seven null hypotheses and results

The results of the seven null hypotheses are: 1) There is no difference in Spanish proficiency based on Gender. On almost all tests males performed somewhat better than the females. However,

the differences are not considered significant; 2) There is no difference in Spanish language proficiency based on Age. Age group 2 (40-49) scored significantly higher than age group 1 (21-39). Age group 3 (50-59) scored significantly higher than age group 1 (21-39). Proficiency in Spanish diminished dramatically from one generation to the next from older to younger based on age; 3) There is no difference in Spanish language proficiency based on Ethnicity. Even though the Chicano group scored higher than the other two groups on most tests, there were no significant differences in *Prueba* scores based on ethnicity; 4) There is no difference in Spanish language proficiency based on where the Spanish speaker is born and raised. Geographical location, the examinees who are born and raised in a Spanish-speaking country scored significantly higher than examinees from New Mexico and other parts of the U.S.; 5) There is no difference in Spanish language proficiency based on self evaluation or self rating. Self-evaluation, the examinees at levels “4” and “3” scored consistently higher than level “1” examinees; 6) There is no difference in Spanish language proficiency based on time spent in a Spanish speaking country. The ones that spend more time in a Spanish speaking country scored higher than those who have spent no time outside the country; 7) There is no difference in Spanish language proficiency based on major area of study. Spanish majors scored consistently higher than bilingual education majors and higher than other/elementary education majors.

13. Recommendations

My recommendations are the following: 1) Spanish Heritage language classes or Spanish for Native Speaker classes in addition to language revitalization type of programs for those who are on the verge or have lost their ancestral language; 2) Require spending a summer session, a semester or possibly a year in a Spanish-speaking country acquiring or strengthening their Spanish language and literacy skills; 3) Bilingual education programs need to be strengthened at the universities; 4) Spanish language immersion programs need to be expanded to other universities and community colleges and offered more often; 5) There must be collaboration among the different academic departments promoting bilingualism; 6) Curriculum Development courses for bilingual educators with a scope and sequence or in which there is articulation from one level to the next need to be offered; 7) Everyone needs to be *Educated* as to what comprises a strong and effective bilingual education program.

14. Conclusions

Prueba directly correlates to the New Mexico State Competencies. *Prueba* was developed by a group of educators from New Mexico who worked to develop a test to respond specifically to the needs of their state. *Prueba* sets high standards for ensuring that the teacher has the entry-level competencies deemed necessary to teach bilingual education in New Mexico. In addition, the state has invested time and funds in developing the infrastructure to support its candidates.

References

- Alderson, J., Clapham, C. & Wall D. (1995). *Language test construction and evaluation*. NY: Cambridge University Press.
- Archuleta, W. (2002). *Prueba de español para la certificación bilingüe: Testing Spanish language proficiency of bilingual teachers in New Mexico*. Doctoral dissertation, Albuquerque, NM: University of New Mexico.
- Bachman, L. & Palmer, A. (1996) *Language Testing in Practice*. Oxford: Oxford University Press.
- Garcia, E., Grant L. & Izquierdo, E. (2002). *Evaluation of New Mexico Spanish Language Instruments for Bilingual Teacher Endorsement*. Santa Fe, New Mexico: New Mexico State Department of Education.
- Grant, L. (1997). Testing the language proficiency of bilingual teachers: Arizona's Spanish proficiency test. *Language Testing*, 14, 1, 23-46.
- Guerrero, Michael D. (1994). *A Critical Analysis of the Validity of the Four Skills Exam*. Doctoral dissertation, Albuquerque, NM: University of New Mexico.

Lewkowitz, J.A. (1997). Integrated Testing. In C. Clapham and D. Corson (Eds) *Encyclopedia of Language and Education, Volume 7, Language Testing and Assessment* (pp. 121-130). Norwell, MA: Kluwer Academic Publishers.

Nelson C. & Guerrero, M. (1999). *Prueba de español para la certificación bilingüe: Technical Manual*. Santa Fe, New Mexico: New Mexico State Department of Education.

ISB4: Proceedings of the 4th International Symposium on Bilingualism

edited by James Cohen, Kara T. McAlister,
Kellie Rolstad, and Jeff MacSwan

Cascadilla Press Somerville, MA 2005

Copyright information

ISB4: Proceedings of the 4th International Symposium on Bilingualism
© 2005 Cascadilla Press, Somerville, MA. All rights reserved

ISBN 978-1-57473-210-8 CD-ROM
ISBN 978-1-57473-107-1 library binding (5-volume set)

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Press.

Ordering information

To order a copy of the proceedings, contact:

Cascadilla Press
P.O. Box 440355
Somerville, MA 02144, USA

phone: 1-617-776-2370
fax: 1-617-776-2271
sales@cascadilla.com
www.cascadilla.com

Web access and citation information

This paper is available from www.cascadilla.com/isb4.html and is identical
to the version published by Cascadilla Press on CD-ROM and in library binding.