



# **Preliminary Item Response Theory Analysis of the University of Minnesota CLA Language Proficiency Tests in French, German, and Spanish**

BY VASHTI MCCOLLUM LOZIER AND  
MICHELINE CHALHOUB-DEVILLE

CARLA Working Paper #5

**CARLA** | Center for Advanced  
Research on Language  
Acquisition

---

UNIVERSITY OF MINNESOTA

**Preliminary Item Response Theory Analysis of the University of  
Minnesota CLA Language Proficiency Tests  
in French, German, and Spanish**

Prepared by  
Vashti McCollum Lozier  
Graduate Research Assistant  
and  
Micheline Chalhoub-Deville, Ph. D.  
Assessment Manager

Center for Advanced Research on Language Acquisition  
University of Minnesota, Minneapolis

# Preliminary Item Response Theory Analysis of the University of Minnesota CLA Language Proficiency Tests in French, German, and Spanish

First Edition

© 1995, 2013 by the Board of Regents of the University of Minnesota. All rights reserved.

Produced by

Center for Advanced Research on Language Acquisition  
University of Minnesota  
140 University International Center  
331 17<sup>th</sup> Ave SE  
Minneapolis, MN 55414  
USA  
612.626.8600  
carla@umn.edu  
<http://www.carla.umn.edu>

## History of CLA Tests

In January of 1983, the Dean of the College of Liberal Arts (CLA) at the University of Minnesota appointed a task force on language instruction to examine the College's second language requirement. To complete the existing requirement, students had a choice between two options: (1) take five quarters of a language, or (2) take three quarters of language and three quarters of culture courses related to the studied language (Lange, 1988). This "seat-time" requirement mandated the number of classes a student had to take, but did not specify a level of proficiency that must be achieved. Upon completing its study - which revealed a generally low level of functional language proficiency in addition to significant differences in outcome between the two tracks - the task force recommended that new foreign language requirements be established for the three most commonly taught languages (French, German and Spanish).

The main goal of the new requirement was the standardization of language proficiency as a requirement for graduation, rather than seat-time. For the three main languages taught at the University of Minnesota, French, German and Spanish, it mandated attainment of a proficiency level "normally attained after two years of college-level study" (Barnes, Klee, Wakefield, 1991). In addition, the new requirement set an entrance standard for enrollment in a second-year language course. The requirement became fully effective Fall quarter, 1988, after a two year phase-in period that began Fall quarter, 1986. At that time, less commonly taught languages were also asked to develop their own proficiency levels with accompanying tests to be completed by 1990.

The new requirement was implemented in several steps beginning in 1984. In the first step, an oral proficiency workshop familiarized 30 educators from various secondary and post secondary institutions in Minnesota with the ACTFL Provisional Guidelines (1982) and the Oral Proficiency Interview (OPI), as well as their implications for instruction and testing of all modalities. The second step consisted of defining the curriculum, defining the test parameters and constraints, and writing test items to serve as a basis for the listening, reading, speaking, and writing tests that were developed in the 1986 academic year (Lange, 1988).

The minimal levels of proficiency for French, German and Spanish were set as follows (Barnes et al., 1991). At the *entrance* level (equivalent to about one year of university study), intermediate low level competence was required in reading and listening, while novice high level was required for writing and speaking. At the

*graduation* level (after approximately two years of university study), intermediate high level competence was required in reading and listening, while intermediate mid level was required for writing and speaking.

The test developers decided to structure the tests of all four modalities according to the ACTFL OPI format, with items divided into 4 stages: warm-up (items below target level), level check (items at desired level), probe (items above target level), and wind-down (items at or below target level).

Since the language requirement became fully effective in 1988, the University, in four yearly testing sessions, has administered both entrances and exit foreign language proficiency tests to thousands of students.

### Data Collection

The purpose of this document is to provide preliminary item analysis findings based on the data collected to date on the entrance and graduation machine-scored reading and listening tests in the three languages combining the results based on an item response theory (IRT) model with classical test theory (CTT) analyses.

Over the years, CTT item and test analyses have been run on individual administrations of the machine-scored tests of reading and listening. These analyses were performed using the computer program TESTAN. TESTAN is used by the University of Minnesota's Measurement Services Office.

At the time the tests were scanned and scored, TESTAN computed statistics for each individual administration of each test and the **original data was discarded**. Because the data was not saved, the raw item and people scores were not available to the assessment researchers at the Center for Advanced Research on Language Acquisition (CARLA) for in-depth analyses across administrations. Other reports produced at CARLA summarize the information that could be gleaned from the limited data available in the TESTAN printouts and provides a detailed content analysis of each test making up the entrance and graduation batteries in all three languages (Chalhoub-Deville, Mueller, Lozier, Juengling, 1996a; Chalhoub-Deville, Alcaya, Klein, Lozier, Budlong, 1996b, Chalhoub-Deville, Sweet, Schmidt, Lozier, 1996c). In a document dated March 20, 1995, an agreement was signed providing CARLA with access to all raw data from all subsequent test administrations of the entrance and graduation tests in French, German, and Spanish (see Appendix A). Having the raw data allows, for the first time, an analysis of the different tests *across* administrations (providing larger sample sizes) and using the more versatile IRT analyses.

## Item Response Theory Analysis

Developed over the past fifty years through the work of Frederic Lord (1952), Georg Rasch (1960), and others, IRT is an alternative test development and analysis model whose most notable advantage over older classical test theory is that IRT allows for the description of test-items and test-takers on the same scale (Lord & Novick, 1968; Weiss & Yoes, 1988). With IRT, we model a student's probabilistic response to an item with known properties given the ability level,  $\theta$  (theta), of the student (Lord, 1980). Unlike CTT statistics, the item analysis statistics (also referred to as item parameters) in IRT are not dependent on the specific sample of students tested.

The three most frequently used IRT models for dichotomously scored test data (i.e., items scored either right or wrong) are the 1-, 2-, and 3-parameter logistic models. The 3-parameter model (3PLM) describes each test item in terms of its difficulty ( $b$ ), its discrimination ( $a$ ), and a third statistic ( $c$ ) called the pseudo-guessing parameter. The  $c$  parameter describes the probability of a person of infinitely low ability answering an item correctly by randomly guessing and can be important when modeling multiple choice items. In the 2-parameter model (2PLM), each test item is described in terms of its difficulty ( $b$ ) and its discrimination ( $a$ ). The 2PLM is similar to the 3PLM except that the guessing parameter ( $c$ ) is set to 0.0 and is not estimated. The 1-parameter model (1PLM or the Rasch model) describes each item in terms of only its difficulty ( $b$ ). The Rasch model assumes that all the items have equal discrimination of 1.0 and that the  $c$  parameter for all items is 0.0.

There are generally accepted sample size requirements for the use of IRT. Although the 1PLM, Rasch model, can be performed with sample sizes as little as 100 to 200, the 2PLM requires 200 to 400 people, and the 3PLM requires large samples of 1,000 to 2,000 people for the item and ability ( $\theta$ ) parameters to be estimated with confidence (Henning, 1987).

## Method

### Participants and Sample Sizes

All quantitative results are based on samples of college students taking either the reading or listening test as part of the entrance or graduation requirement at the University of Minnesota from May 15, 1995 to September 22, 1995. The data received from the CLA Testing Office is summarized for the entrance tests in Table 1 and for the graduation tests in Table 2.

The sample sizes for the French (reading  $N=274$ , listening  $N=226$ ) and German (reading  $N=193$ , listening  $N=192$ ) entrance tests are clearly too small for a 2PLM or 3PLM IRT analysis to be performed. The sample sizes for the Spanish entrance tests, however, are 690 for the reading test and 666 for the listening test. The French graduation tests also had small sample sizes (reading  $N=219$ , listening  $N=220$ ). However, the graduation tests of Spanish had larger sample sizes of 552 for the reading test and 568 for the listening test. While a sample size of at least 1,000 people is recommended for using the 3PLM, sufficiently large samples for the 2PLM were available for the Spanish tests. The Rasch model was not considered because of its restrictive assumption of equal item discriminations.

### Analysis

The test-takers' responses to the Reading and Listening Tests in French, German, and Spanish were fitted to a 2PLM IRT model. The 2PLM describes each test item in terms of its difficulty ( $b$ ) and its discrimination ( $a$ ).

The IRT analysis program, XCALIBRE, developed by the Assessment Systems Corporation, was used to estimate the discrimination and difficulty parameters for each of the 40 items on each test.

In addition, CTT item analysis statistics provided by the XCALIBRE program were calculated. The CTT analyses reported also provide important information on the test items. As a result of the larger sample sizes available by collecting data across test administrations, the CTT statistics in this report serve to augment the results provided in other reports (Chalhoub-Deville et al., 1996a; Chalhoub-Deville et al., 1996b; Chalhoub-Deville et al., 1996c).

Table 1

Summary of CLA Entrance Test Data Received Based on Joint Agreement Dated March 20, 1995

Language	Reading		Listening	
	Date	N	Date	N
French	5/15/95	29	5/15/95	31
	6/26/95	22	6/26/95	22
	7/8/95	25	7/8/95	27
	7/19/95	44	7/19/95	44
	7/22/95	47	7/22/95	47
	7/29/95	36	7/29/95	36
	8/28/95	52	8/28/95	55
	9/22/95	14	9/22/95	14
German	5/15/95	22	5/15/95	22
	6/26/95	21	6/26/95	21
	7/8/95	25	7/8/95	26
	7/19/95	26	7/19/95	26
	7/22/95	23	7/22/95	23
	7/29/95	35	7/29/95	34
	8/28/95	24	8/28/95	24
	9/22/95	16	9/22/95	16
Spanish	5/15/95	67	5/15/95	66
	6/26/95	78	6/26/95	79
	7/8/95	90	7/8/95	90
	7/19/95	93	7/19/95	77
	7/22/95	87	7/22/95	91
	7/29/95	103	7/29/95	104
	8/28/95	106	8/28/95	112
	9/22/95	49	9/22/95	50

Table 2

Summary of CLA Graduation Test Data Received Based on Joint Agreement Dated  
March 20, 1995

Language	Reading		Listening	
	Date	N	Date	N
French	5/22/95	163	5/22/95	163
	7/14/95	9	7/14/95	10
	8/18/95	34	8/18/95	33
	9/23/95	13	9/23/95	14
Spanish	5/22/95	320	5/22/95	325
	7/14/95	56	7/14/95	62
	8/18/95	113	8/18/95	114
	9/23/95	48	9/23/95	67

## Results

The results of the 2-parameter IRT and CTT analyses of the Spanish reading and listening proficiency tests for the entrance and graduation test batteries will be presented in turn. Because the sample sizes were insufficient for conducting an IRT analysis on the remaining tests, those analyses, although performed, are included only in the appendix (see Appendix B).

The CTT item statistics--proportion correct (item difficulty) and point biserial correlation (item discrimination)--can be found in the first of two tables relating to each test. Unlike the IRT statistics, the CTT item statistics are based on sufficiently large sample sizes to be useful for all tests, not just the Spanish. When combined with the IRT analyses available for the Spanish tests, the item information contained in this report is sizable for the four Spanish tests.

Each test has two tables of results associated with it, which communicate distinct types of information. The first table reports information on the 2PLM item-model fit and the statistics on each item's difficulty and discrimination in both the CTT and IRT metric. The second table contains specific information about the alternative answer options (distracters) associated with each test item. Below is a more detailed depiction of the information provided by each of these two tables. For ease of communication, these two tables are referred to as item and option tables, respectively.

### Information Reported in Item Table

The item table format contains a list of the item numbers and each item's intended function (i.e., warm-up, level check, probe, or wind-down). Also presented are the item parameter estimates of item difficulty ( $b$ ) and discrimination ( $a$ ). Following the item parameters, is the standardized residual statistics which indicate the degree to which the item conforms or fails to conform to the 2-parameter model. The standardized residual is distributed normally and values above 2.0 indicate items that do not fit the model well. Next, the proportion correct, a measure of item difficulty (percentage of students answering an item correctly) and point biserial, a measure of item discrimination (correlation between each item response and number correct total score) are listed. Additionally, the point biserial is re-calculated as the correlation between each item response and the IRT estimate of ability,  $\theta$ . The point biserial and the point biserial  $\theta$  differ only in that the former uses number correct score as its performance criterion while the latter uses the IRT ability estimate  $\theta$ . The point biserial

correlation which is the correlation between number correct score and item score is included in these analyses because it is one of the most common measures of item discrimination. The point biserial  $\theta$  may be a more accurate measure of item discrimination because  $\theta$  is a better estimator of ability than is number correct score.

The XCALIBRE program automatically flags items that are potentially problematic. Flagged items are indicated in the column labeled “flags.” The Flags and their meanings are listed below (Assessment Systems Corporation,1995).

### Potential Item Flags

<b>Flag</b>	<b>Meaning</b>
P	<p>Potentially Problematic item: can mean one (or more) of the following:</p> <p style="padding-left: 40px;"><i>a</i> value &lt; 0.30</p> <p style="padding-left: 40px;"><i>b</i> value &gt; 2.95 or <i>b</i> value &lt; -2.95</p> <p style="padding-left: 40px;"><i>c</i> value &gt; 0.40</p>
K	<p>Indicates a possible keying error. This indicates that one of the incorrect item response options has a higher correlation with total score than does the correct response.</p>
R	<p>Indicates that the standardized residual (item model fit) statistic exceeded a value of 2.0.</p> <p>This item does not meet the assumptions of the 2PLM.</p>

### Information Reported in the Option Table

The option table provides detailed information regarding the effectiveness of the various options (the correct answer and the distracters). First, the endorsement rate provides the percentage of students choosing each of the option answers. In the case of the correct answer, indicated with an asterisk, this is the same as the percentage of students answering the item correctly. Secondly, the point biserial correlation between each alternative item response and the person's ability,  $\theta$ , (multiplied by 100 to eliminate decimal points) is an indication of the discriminating power of the correct answer and each distracter.

In the case of the correct answer, good discrimination is indicated by positive correlations between the response and overall ability (theta). For the distracters, good discrimination is indicated by negative correlations. It is important to note that when an item is very easy (i.e. the percentage of people answering the item correctly is very

high), it curtails the correlation between the option and total score so that neither the correct answer nor the distracters appear to be doing a good job discriminating between test-takers. In the case of warm-up items, for example, the purpose of these items is to put students at ease, not to discriminate between them. Poor item-theta correlations are troublesome only when they occur in level-check and probe items that are supposed to serve the purpose of discriminating between the more able students and the less able students. The item option-theta correlation can also be curtailed by extremely difficult items. However, items in these tests are rarely so difficult that 95% or more of students are answering them incorrectly. When items are that difficult, it is often because they are miskeyed.

When using the second table to evaluate the effectiveness of each distracter, the first set of statistics to examine are the endorsement rates for each option. In general, an effective distracter is one that entices a certain number of test-takers to select it. The point-biserial  $\theta$  correlations for each option provide further information. A strong negative correlation between a distracter and overall ability,  $\theta$ , is a good indication of a distracter that is functioning properly. A positive or very small negative correlation between a distracter and  $\theta$  indicates a distracter that attracts many of the better foreign language students. To be effective, a distracter must attract students of lesser ability, but not be very distracting to those of the highest ability levels.

Many examples of the sorts of problems described above occur on each of the four Spanish tests. For explication, certain instances of problems and trends will be pointed out for each of the tests.

## Spanish Entrance Reading Test

Tables 3 and 4 present the results of the Spanish entrance reading test. Table 3, information about the items, contains the item parameters. The difficulties of the items, as reflected in both the  $b$  parameter and the proportion correct, should correspond with the pattern of item functions (i.e., warm up, level check, probe, or wind down). It is clear that, in this test, the expected pattern of difficulty is violated in as many as 17 items out of 40. For example, items 38 to 40 are supposed to function as wind down items. Item 38 with a  $b$  of -2.89 and proportion correct of .91 and item 40 with a  $b$  of -2.87 and proportion correct of .94 are, in fact, easy items that allow the student to end the test on a positive note. Item 39, however, has a  $b$  of only -0.15 and proportion correct of .54. This is a very difficult item that appears to be functioning more as a probe than as a wind down item.

In addition, XCALIBRE flags certain items as problematic. In the cases of items 4, 11, 14, 17, and 19, these items were flagged with a “P” because they are extremely easy with  $b$ 's of -3.00 and proportion corrects of .93 and higher. In the instance of item 4, this is not a problem because it is intended as a warm-up item. The remaining items with low difficulty (easy items), however, are level check items and would be expected to be somewhat more difficult.

Item 20 is flagged with a “P” because of its low discrimination ( $a$ ) value of .29 and its low point biserial correlation of .15. Looking at the option table 4, the correlation between the correct answer and overall ability (multiplied by 100 to eliminate decimal points) is just 9. It is very similar to the correlation between option number 1 and overall ability which is 7. This indicates that, statistically speaking, option 2 (i.e., choice “b”) is not clearly more correct than option 1 (choice “a”). Therefore, it would be reasonable to revise this item so that there is one clear answer.

Item 29 is also problematic. This item was flagged by XCALIBRE because of its low discrimination,  $a=.26$ , and high standardized residual of 2.10 (above the 2.0 cutoff). This indicates that the more proficient students are no more likely to answer this item correctly than are the less proficient students. Not only is this item not discriminating well among students, it is not fitting the 2-parameter IRT model as indicated by the “R” flag and may be miskeyed as indicated by the “K” flag. The content analysis reported in Chalhoub-Deville et al. (1996c) indicates that item 29 has two potentially correct answers (both the keyed response, “b,” and the option “c”). This could certainly account for the statistics found in the current analysis.

Table 3

Spanish Entrance Reading Item Table: 2 Parameter Logistic Model Analysis, N=690

Item	Function	a	b	Flags	Residua I	Pr. Correct	Pt. Biserial	Pt. Biserial $\theta$
1	WU	0.41	-1.96		1.28	0.76	0.22	0.21
2	WU	0.77	-2.80		0.62	0.94	0.36	0.33
3	WU	0.63	-2.12		0.59	0.85	0.38	0.37
4	WU	0.67	-3.00	P	0.83	0.96	0.23	0.22
5	WU	0.46	-2.18		0.83	0.80	0.27	0.25
6	LC	0.81	-2.51		0.65	0.92	0.39	0.37
7	WU	0.67	-1.87		0.67	0.83	0.42	0.41
8	WU	0.60	-2.44		0.81	0.87	0.35	0.32
9	LC	0.87	-2.36		0.41	0.92	0.44	0.41
10	LC	0.76	-2.50		0.45	0.91	0.40	0.36
11	LC	0.52	-3.00	P	1.59	0.93	0.15	0.14
12	LC	0.64	-1.99		0.64	0.84	0.39	0.38
13	LC	0.46	-1.73		0.77	0.76	0.32	0.29
14	LC	0.50	-3.00	P	0.74	0.92	0.18	0.17
15	LC	0.65	-1.73		0.59	0.81	0.41	0.42
16	LC	0.62	-2.78		0.96	0.91	0.31	0.29
17	LC	0.84	-3.00	P	0.88	0.98	0.28	0.23
18	LC	0.56	0.94		0.69	0.34	0.40	0.44
19	LC	0.84	-3.00	P	0.68	0.98	0.28	0.25
20	LC	0.29	0.07	P	1.72	0.50	0.15	0.09
21	LC	0.84	-2.35		0.55	0.91	0.43	0.40
22	LC	0.56	0.38		0.63	0.44	0.42	0.46
23	LC	0.47	-1.75		1.23	0.76	0.33	0.30
24	LC	0.46	-1.77		0.76	0.76	0.29	0.29
25	LC	0.48	-2.11		0.89	0.81	0.30	0.28
26	LC	0.54	0.88		0.70	0.35	0.39	0.43
27	LC	0.49	-1.14		0.75	0.68	0.36	0.35
28	LC	0.56	-2.36		0.87	0.85	0.33	0.30
29	LC	0.26	-2.08	RPK	2.10	0.70	0.00	-0.03
30	P	0.42	0.36		0.74	0.45	0.31	0.32
31	P	0.63	-2.42		0.85	0.88	0.36	0.33
32	P	0.52	-0.29		0.79	0.56	0.38	0.41
33	P	0.71	-1.23		0.87	0.74	0.47	0.48
34	LC	0.63	-2.44		0.73	0.88	0.35	0.33
35	LC	0.51	-2.30		1.04	0.83	0.29	0.28
36	P	0.49	-0.48		0.65	0.59	0.38	0.37
37	P	0.43	-1.43		0.93	0.70	0.28	0.27
38	WD	0.60	-2.89		0.73	0.91	0.29	0.28
39	WD	0.56	-0.15		0.56	0.54	0.43	0.44
40	WD	0.71	-2.87		0.94	0.94	0.35	0.31

Table 4

## Spanish Entrance Reading Test Option Effectiveness (2PLM), N=690

Item	Function	Endorsement Rate				Item-Theta Correlation (Pt. Biserial $\theta$ )			
		1	2	3	4	1	2	3	4
1	WU	19	4	76 *	1	-17	-9	-21 *	-5
2	WU	1	4	1	94 *	-17	-23	-15	33 *
3	WU	8	5	85 *	2	-27	-18	37 *	-13
4	WU	96 *	4	0	0	22 *	-22	0	0
5	WU	13	80 *	3	4	-13	25 *	-18	-14
6	LC	4	92 *	3	1	-27	37 *	-23	-10
7	WU	3	5	83 *	9	-19	-22	41 *	-26
8	WU	4	87 *	1	8	-30	32 *	-17	-11
9	LC	4	2	92 *	1	-27	-24	41 *	-16
10	LC	1	91 *	4	4	-16	36 *	-23	-22
11	LC	93 *	0	5	1	14 *	4	-13	-9
12	LC	1	13	84 *	2	-12	-29	38 *	-22
13	LC	76 *	7	9	8	29 *	-11	-19	-15
14	LC	2	0	6	92 *	-14	1	-12	17 *
15	LC	5	11	3	81 *	-22	-27	-18	42 *
16	LC	2	2	91 *	5	-15	-16	29 *	-19
17	LC	1	98 *	0	1	-15	23 *	-9	-16
18	LC	16	34 *	45	5	-22	44 *	-20	-14
19	LC	1	1	1	98 *	-7	-15	-20	25 *
20	LC	37	50 *	7	6	7	9 *	-22	-10
21	LC	4	3	2	91 *	-23	-20	-25	40 *
22	LC	10	10	36	44 *	-16	-28	-20	46 *
23	LC	12	76 *	6	6	-15	30 *	-15	-18
24	LC	15	7	76 *	2	-16	-25	29 *	-3
25	LC	81 *	6	11	3	28 *	-23	-13	-11
26	LC	9	35 *	34	22	-11	43 *	-26	-13
27	LC	8	68 *	18	6	-8	35 *	-31	-9
28	LC	9	2	4	85 *	-23	-13	-12	30 *
29	LC	3	25	70 *	1	-16	11	-3 *	-5
30	P	14	33	45 *	7	-28	-12	32 *	-1
31	P	3	4	88 *	5	-7	-25	33 *	-22
32	P	23	56 *	5	16	-27	41 *	-12	-18
33	P	8	11	7	74 *	-30	-22	-24	48 *
34	LC	6	88 *	3	3	-17	33 *	-20	-19
35	LC	2	12	83 *	3	-11	-19	28 *	-16
36	P	26	1	59 *	14	-15	-13	37 *	-29
37	P	4	70 *	24	1	-4	27 *	-24	-9
38	WD	5	91 *	2	1	-21	28 *	-13	-10
39	WD	11	54 *	32	3	-15	44 *	-33	-13
40	WD	4	1	2	94 *	-26	-8	-14	31 *

Notes. An asterisk indicates the item that is keyed as correct.  
Endorsement rates are given as percentages.

## Spanish Entrance Listening Test

Tables 5 and 6 present the results of the Spanish entrance listening test. Once again, the difficulties of these items do *not* reflect their intended functions for at least 17 of the 40 test items. For example, as table 5 shows, items 14 and 15 have *b*'s of .36 and .32, respectively, which when compared to the values of the other items, indicate that they are extremely difficult items. Only about 45 % of examinees were able to answer each of these items correctly. These items are supposed to be level check items, but may be more appropriate as probes. In table 6, items 14 and 15 appear to be functioning well in terms of the correlations between each option and overall ability indicating that these may be good items that are just misclassified in terms of their difficulty function.

Additionally, a number of items are flagged with a "P" because of their low difficulty (i.e., items 2, 18, and 25). Item number 2 is not a concern because it is a warm-up item. Items 18 and 25, however, seem to be much easier than other level check items on the same test.

Item 29 on this test is particularly troublesome. It is extremely difficult ( $b=+3.00$ ), has a low point biserial correlation (-.11) and has a standardized residual of 2.80 which is well above the 2.0 misfit cutoff. Turning to option table 6, the nature of item 29's problem becomes more clear. For example, only 5% of examinees choose number 2 as the correct answer. The largest percent of students choose option 3 (83%). In addition, only option three has a positive correlation with overall ability. The correlation between option 2 and overall ability is quite poor (-.11). From this analysis, it appears that item 29 is miskeyed with option 3, not 2, being the correct answer. In fact, the CLA Language Testing Office gives students credit for choosing either option 2 or 3. Based on this statistical analysis alone, it appears that credit should only be given to students choosing option 3 instead of credit being given for both options 2 and 3 as is currently done.

Table 5

Spanish Entrance Listening Item Table: 2 Parameter Logistic Model Analysis, N=666

Item	Function	a	b	Flags	Residua l	Pr. Correct	Pt. Biserial	Pt. Biserial $\theta$
1	WU	0.74	-2.17		0.87	0.88	0.37	0.36
2	WU	0.84	-3.00	P	1.13	0.98	0.25	0.22
3	WU	0.72	-1.36		0.41	0.77	0.45	0.45
4	WU	0.78	-0.33		0.70	0.59	0.54	0.56
5	WU	0.46	-1.24		0.69	0.69	0.30	0.30
6	WU	0.61	-1.85		0.52	0.81	0.36	0.36
7	LC	0.38	-0.61		1.02	0.59	0.25	0.24
8	LC	0.80	-1.81		0.86	0.85	0.43	0.43
9	P	0.53	-0.70		1.07	0.63	0.39	0.39
10	LC	0.53	-1.00		0.44	0.67	0.38	0.37
11	LC	0.46	-2.77		1.15	0.86	0.18	0.17
12	LC	0.50	-2.56		1.25	0.85	0.23	0.23
13	P	0.63	-1.91		0.92	0.82	0.36	0.36
14	LC	0.68	0.36		0.96	0.44	0.49	0.53
15	LC	0.53	0.32		1.04	0.46	0.40	0.42
16	P	0.40	-0.79		1.25	0.62	0.27	0.25
17	LC	0.41	-1.24		1.39	0.67	0.25	0.23
18	LC	0.67	-3.00	P	1.41	0.97	0.11	0.11
19	LC	0.56	-1.86		0.74	0.80	0.35	0.33
20	P	0.47	-2.00		0.79	0.79	0.28	0.26
21	LC	0.40	-1.93		1.45	0.75	0.20	0.18
22	LC	0.78	-2.91		0.42	0.95	0.31	0.28
23	P	0.53	-1.67		0.87	0.77	0.33	0.33
24	LC	0.63	-1.96		0.73	0.83	0.36	0.36
25	LC	0.69	-3.00	P	0.77	0.96	0.25	0.22
26	P	0.35	-2.53		1.82	0.79	0.12	0.08
27	LC	0.89	-0.41		0.66	0.61	0.57	0.60
28	P	0.34	-0.01		1.25	0.51	0.22	0.19
29	P	0.47	3.00	RPK	2.80	0.05	-0.11	-0.11
30	LC	0.50	-1.84		0.79	0.78	0.30	0.29
31	LC	0.75	-2.36		0.79	0.90	0.36	0.34
32	LC	0.33	-2.15		1.91	0.74	0.10	0.07
33	LC	0.60	-1.37		0.59	0.74	0.40	0.39
34	LC	0.54	-2.55		1.09	0.86	0.27	0.25
35	LC	0.66	-1.84		0.56	0.82	0.39	0.38
36	LC	0.52	-0.78		0.87	0.64	0.38	0.38
37	P	0.54	-1.22		0.44	0.71	0.38	0.37
38	WD	0.54	-1.79		0.66	0.78	0.34	0.32
39	WD	0.41	0.01		1.01	0.51	0.31	0.29
40	WD	0.44	0.43		0.97	0.44	0.34	0.33

Table 6

Spanish Entrance Listening Test Option Effectiveness (2PLM), N=666

Item	Function	Endorsement Rate					Item-Theta Correlation (Pt. Biserial $\theta$ )				
		1	2	3	4	5	1	2	3	4	5
1	WU	1	6	5	1	88 *	-5	-28	-18	-11	36 *
2	WU	98 *	0	1	1	0	22 *	-10	-13	-13	-9
3	WU	1	77 *	11	6	6	-15	45 *	-24	-17	-28
4	WU	31	6	4	59*	-	-48	-18	-9	56*	-
5	WU	1	17	69 *	12	-	-11	-17	30 *	-19	-
6	WU	6	3	81 *	10	-	-18	-16	36 *	-23	-
7	LC	34	59 *	7	0	-	-13	24 *	-20	-5	-
8	LC	6	85 *	3	7	-	-30	43 *	-18	-22	-
9	P	63 *	1	20	16	-	39 *	-5	-28	-19	-
10	LC	1	6	26	67 *	-	-10	-2	-37	37 *	-
11	LC	0	13	1	86 *	-	-5	-15	-8	17 *	-
12	LC	2	10	85 *	3	-	-11	-12	23	-18	-
13	P	82 *	9	6	3	-	36 *	-31	-16	-7	-
14	LC	44 *	15	34	7	-	53 *	-11	-41	-12	-
15	LC	18	46 *	12	24	-	-26	42 *	-2	-24	-
16	P	25	62 *	8	5	-	0	25 *	-26	-22	-
17	LC	20	3	9	67 *	-	-12	-15	-12	23 *	-
18	LC	0	97 *	3	0	-	0	11 *	-11	-1	-
19	LC	9	6	80 *	6	-	-17	-14	33 *	-22	-
20	P	2	79 *	8	11	-	-9	26 *	-24	-9	-
21	LC	7	8	75 *	10	-	-9	-9	18 *	-11	-
22	LC	1	2	2	95*	-	-14	-14	-18	28 *	-
23	P	1	8	83 *	8	-	-9	-27	36 *	-19	-
24	LC	1	8	83 *	8	-	-9	-27	36 *	-19	-
25	LC	2	96 *	1	1	-	-8	22 *	-13	-18	-
26	P	2	16	3	79 *	-	-9	-2	-7	8 *	-
27	LC	22	9	61 *	9	-	-46	-23	60 *	-12	-
28	P	12	25	12	51 *	-	-11	1	-19	19 *	-
29	P	8	5 *	83	4	-	-9	-11 *	22	-18	-
30	LC	78 *	2	4	16	-	29 *	-6	-13	-24	-
31	LC	9	90 *	1	0	-	-34	34 *	-7	2	-
32	LC	8	16	74 *	2	-	-1	-4	7 *	-9	-
33	LC	19	4	74 *	2	-	-35	-11	39 *	-8	-
34	LC	3	1	9	86 *	-	-20	-12	-14	25 *	-
35	LC	82 *	3	9	6	-	38 *	-15	-27	-18	-
36	LC	64 *	7	20	9	-	38 *	-22	-17	-18	-
37	P	8	9	71 *	11	-	-18	-18	37 *	-20	-
38	WD	12	78 *	4	6	-	-28	32 *	-12	-8	-
39	WD	22	51 *	20	7	-	-12	29 *	-17	-11	-
40	WD	11	44 *	37	8	-	-24	33 *	-14	-10	-

Notes. An asterisk indicates the item that is keyed as correct.  
Endorsement rates are given as percentages.

## Spanish Graduation Reading Test

The results of the Spanish Reading Graduation test are summarized in tables 7 and 8. Similar to the other tests, table 7 indicates that the items in the Spanish graduation reading test often, at least 16 of 40, deviate in difficulty from what would be expected based on their intended function. Item 40, for instance, is very difficult for a wind-down while item 20 is extremely easy for a probe item.

A huge proportion of this test's items, 23 of 40, are extremely easy for students (items 1-9, 11, 13, 15, 18, 20, 22, 23, 25, 26, 33, 35, 36, and 39). Of these items, 16 were not designed as either warm-up or wind-down items. There are two possible explanations. First, the students of Spanish tested in this sample may be extremely proficient at reading and are all well above the criterion cutoff. Alternatively, these items may, in fact, be too easy for the proficiency level targeted. If the latter is the case, these items should be considered for revision so that they are more appropriate for their intended function. For example, table 8 shows that the endorsement rates for the various options associated with item 13 are very low (only 3% of students chose option 2 and no student chose options 1 or 3). In order to revise this item, the test developer might wish to focus on making these options more desirable or increasing the complexity of the text or the question asked.

## Spanish Graduation Listening Test

The results of the Spanish Graduation Listening Test are reported in tables 9 and 10. Similar to the earlier results, table 9 reveals that at least 15 of the 40 items in this test deviate in difficulty from what would be expected based on their intended function. For example, four of the five wind-down items are quite difficult (items 36, 37, 38, and 40). These items had *b* values ranging from 1.24 to -1.37 with 34% to 72% of students able to answer these questions correctly. These items would not allow students to complete the test "feeling good," a function of wind-down items.

As before, a number of items were flagged as being too easy (items 3, 5, 7, 9, 18, 19, 31, and 39). These items, except for 3 and 39, are neither warm-up or wind-down items.

Table 7

Spanish Graduation Reading Item Table: 2 Parameter Logistic Model Analysis, N=552

Item	Function	a	b	Flags	Residual	Pr. Correct	Pt. Biserial	Pt. Biserial $\theta$
1	WU	0.66	-3.00	P	1.33	0.97	0.12	0.11
2	WU	0.52	-3.00	P	1.07	0.93	0.12	0.12
3	WU	0.85	-3.00	P	1.25	0.98	0.29	0.22
4	WU	0.75	-3.00	P	1.38	0.99	0.07	0.08
5	WU	0.68	-3.00	P	1.27	0.97	0.13	0.13
6	WU	0.78	-3.00	P	1.51	0.99	0.11	0.10
7	LC	0.82	-3.00	P	1.72	0.99	0.10	0.07
8	LC	0.74	-3.00	P	0.52	0.96	0.31	0.26
9	LC	0.52	-2.79		0.98	0.88	0.26	0.24
10	LC	0.44	-0.77		0.82	0.61	0.33	0.32
11	LC	0.57	-3.00	P	0.28	0.93	0.24	0.22
12	LC	0.61	-2.94		0.65	0.92	0.28	0.26
13	LC	0.70	-3.00	P	0.77	0.97	0.24	0.21
14	LC	0.60	-2.27		0.87	0.85	0.37	0.33
15	LC	0.57	-2.91		0.93	0.90	0.26	0.25
16	LC	0.57	-2.17		0.88	0.83	0.35	0.33
17	LC	0.50	-2.53		1.03	0.85	0.27	0.25
18	LC	0.62	-2.59		1.09	0.89	0.32	0.31
19	P	0.59	-0.56		0.61	0.59	0.44	0.48
20	P	0.77	-2.67		0.51	0.93	0.39	0.34
21	LC	0.58	-2.20		0.94	0.84	0.34	0.34
22	LC	0.58	-3.00	P	0.35	0.92	0.26	0.24
23	LC	0.72	-3.00	P	0.29	0.95	0.34	0.28
24	LC	0.59	-1.66		0.89	0.77	0.40	0.40
25	LC	0.55	-3.00	P	0.58	0.93	0.20	0.19
26	LC	0.72	-3.00	P	1.45	0.98	0.11	0.11
27	LC	0.57	-2.35		0.88	0.85	0.35	0.31
28	P	0.49	-0.15		0.65	0.52	0.37	0.40
29	LC	0.64	-3.00	P	0.64	0.95	0.22	0.21
30	LC	0.41	-1.32		1.45	0.67	0.26	0.25
31	LC	0.58	-1.20		0.96	0.70	0.42	0.43
32	LC	0.53	-2.11		0.91	0.81	0.34	0.32
33	LC	0.73	-3.00	P	0.51	0.96	0.31	0.26
34	P	0.39	0.13		0.94	0.48	0.29	0.27
35	LC	0.69	-3.00	P	0.38	0.95	0.28	0.25
36	LC	0.63	-2.77		0.89	0.91	0.32	0.29
37	LC	0.39	-2.41		1.60	0.79	0.19	0.15
38	WD	0.59	-2.18		0.86	0.84	0.35	0.34
39	WD	0.56	-3.00	P	0.48	0.93	0.21	0.20
40	WD	0.52	-0.96		0.73	0.65	0.39	0.40

Table 8

## Spanish Graduation Reading Test Option Effectiveness (2PLM), N=552

Item	Function	Endorsement Rate				Item-Theta Correlation (Pt. Biserial $\theta$ )			
		1	2	3	4	1	2	3	4
1	WU	2	97 *	0	0	-11	11 *	2	-7
2	WU	3	2	93 *	2	-7	-9	12 *	-5
3	WU	1	1	98 *	1	-9	-13	22 *	-16
4	WU	99 *	1	0	0	8 *	-9	-1	-2
5	WU	1	97 *	1	0	-6	13 *	-10	-8
6	WU	99 *	0	0	1	10 *	-3	0	-10
7	LC	1	99 *	0	0	-7	7 *	0	0
8	LC	96 *	0	4	0	26 *	-10	-25	0
9	LC	9	1	88 *	3	-18	-8	24 *	-13
10	LC	1	34	61 *	5	-7	-31	32 *	-2
11	LC	0	93 *	5	2	-9	22 *	-15	-14
12	LC	92 *	1	7	0	26 *	-7	-24	-9
13	LC	0	3	0	97 *	-5	-20	-5	21 *
14	LC	6	3	7	85 *	-13	-20	-23	33 *
15	LC	2	90 *	4	4	-7	25 *	-15	-19
16	LC	4	8	83 *	6	-15	-22	33 *	-16
17	LC	7	8	85 *	0	-12	-21	25 *	-4
18	LC	1	9	89 *	1	-14	-28	31 *	-3
19	P	22	7	59 *	12	-28	-20	48 *	-22
20	P	4	1	2	93 *	-25	-15	-15	34 *
21	LC	10	3	4	84 *	-20	-11	-25	34 *
22	LC	3	92 *	4	1	-15	24 *	-17	-6
23	LC	95 *	2	0	3	28 *	-18	-12	-19
24	LC	3	77 *	17	4	-9	40 *	-33	-16
25	LC	2	3	93 *	2	-14	-13	19 *	-5
26	LC	1	98 *	1	0	-6	11 *	-8	-4
27	LC	0	85 *	9	5	-1	31 *	-32	-8
28	P	1	45	52 *	3	-2	-38	40 *	-4
29	LC	1	95 *	1	3	-9	21 *	-11	-15
30	LC	2	1	30	67 *	-3	-5	-24	25 *
31	LC	10	70 *	18	3	-16	43 *	-33	-14
32	LC	10	81 *	6	2	-19	32 *	-19	-14
33	LC	0	3	1	96 *	-11	-18	-15	26 *
34	P	15	48 *	29	8	-17	27 *	-12	-8
35	LC	1	3	1	95 *	-10	-20	-11	25 *
36	LC	4	3	2	91 *	-16	-19	-14	29 *
37	LC	10	6	79 *	5	-10	-8	15 *	-5
38	WD	3	84 *	1	12	-9	34 *	-9	-30
39	WD	93 *	2	1	4	20 *	-17	-14	-7
40	WD	19	10	7	65 *	-19	-19	-23	40 *

Notes. An asterisk indicates the item that is keyed as correct.  
Endorsement rates are given as percentages.

Table 9

Spanish Graduation Listening Item Table: 2 Parameter Logistic Model Analysis, N=568

Item	Function	a	b	Flags	Residual	Pr. Correct	Pt. Biserial	Pt. Biserial $\theta$
1	WU	0.51	-2.42		0.91	0.85	0.27	0.25
2	WU	0.39	-2.09		1.74	0.76	0.18	0.16
3	WU	0.79	-3.00	P	1.57	0.99	0.04	0.05
4	WU	0.56	-2.51		0.56	0.87	0.28	0.27
5	LC	0.71	-3.00	P	0.57	0.95	0.29	0.26
6	LC	0.42	-2.19		1.25	0.78	0.21	0.18
7	LC	0.40	-3.00	P	1.47	0.87	0.09	0.08
8	LC	0.51	-1.51		0.78	0.73	0.35	0.33
9	LC	0.64	-3.00	P	1.14	0.96	0.15	0.15
10	LC	0.49	-2.67		1.13	0.86	0.22	0.21
11	LC	0.40	-1.94		1.38	0.74	0.19	0.18
12	LC	0.56	-0.69		0.42	0.62	0.42	0.41
13	LC	0.57	-2.64		0.48	0.89	0.28	0.26
14	LC	0.57	-3.00	P	0.46	0.92	0.24	0.22
15	LC	0.48	-2.25		0.84	0.82	0.27	0.24
16	LC	0.59	-1.46		0.75	0.75	0.40	0.38
17	LC	0.36	-0.74		1.29	0.60	0.20	0.19
18	LC	0.76	-3.00	P	1.52	0.99	0.13	0.12
19	LC	0.41	-3.00	P	1.46	0.88	0.11	0.09
20	LC	0.50	0.05		0.56	0.49	0.39	0.39
21	LC	0.37	0.23		1.63	0.47	0.24	0.23
22	LC	0.44	0.64		0.69	0.41	0.34	0.34
23	LC	0.51	-1.99		0.70	0.80	0.30	0.28
24	LC	0.51	-2.68		0.76	0.87	0.25	0.23
25	LC	0.48	1.52		0.70	0.26	0.35	0.37
26	LC	0.56	-1.68		0.86	0.77	0.36	0.35
27	LC	0.56	-1.34		0.87	0.72	0.38	0.37
28	P	0.69	-0.73		0.42	0.65	0.48	0.49
29	P	0.60	-0.68		0.34	0.63	0.43	0.44
30	LC	0.46	-2.08		1.09	0.78	0.25	0.24
31	LC	0.70	-3.00	P	0.39	0.95	0.30	0.26
32	LC	0.38	-1.53		1.39	0.69	0.20	0.19
33	P	0.64	-0.69		0.36	0.63	0.46	0.47
34	P	0.69	-0.48		0.30	0.60	0.48	0.51
35	P	0.65	-0.69		0.52	0.63	0.46	0.47
36	WD	0.58	-0.62		0.30	0.62	0.43	0.43
37	WD	0.52	-1.37		0.56	0.72	0.35	0.35
38	WD	0.55	-0.93		0.61	0.66	0.40	0.40
39	WD	0.63	-3.00	P	1.54	0.96	0.15	0.14
40	WD	0.34	1.24		1.45	0.34	0.20	0.17

Table 10

## Spanish Graduation Listening Test Option Effectiveness (2PLM), N=568

Item	Function	Endorsement Rate				Item-Theta Correlation (Pt. Biserial $\theta$ )			
		1	2	3	4	1	2	3	4
1	WU	85 *	0	2	13	25 *	0	-12	-22
2	WU	1	21	76 *	2	-6	-13	16 *	-8
3	WU	0	1	99 *	0	0	-5	5 *	0
4	WU	2	1	87 *	11	-8	-6	27 *	-24
5	LC	2	1	2	95 *	-18	-11	-15	26 *
6	LC	1	12	78 *	9	-12	-8	18 *	-14
7	LC	87 *	3	10	0	8 *	-5	-7	4
8	LC	13	6	73 *	8	-26	-15	33 *	-8
9	LC	0	96 *	3	0	0	15 *	-14	-6
10	LC	5	0	8	86 *	-11	-2	-16	21 *
11	LC	1	74 *	18	7	-7	18 *	-4	-21
12	LC	0	3	35	62 *	-9	-15	-35	41 *
13	LC	0	89 *	1	9	-5	26 *	-11	-23
14	LC	6	1	1	92 *	-20	-4	-8	22 *
15	LC	82 *	4	5	10	24 *	1	-24	-15
16	LC	14	75 *	10	1	-29	38 *	-17	-13
17	LC	19	60 *	5	16	-11	19 *	-16	-4
18	LC	1	0	0	99 *	-9	-6	-6	12 *
19	LC	3	88 *	2	7	-8	9 *	-10	1
20	LC	49 *	2	26	22	39 *	-10	-16	-27
21	LC	47 *	5	43	4	23 *	1	-17	-18
22	LC	45	41 *	11	4	-23	34 *	-10	-11
23	LC	9	80 *	8	3	-14	28 *	-21	-11
24	LC	1	8	87 *	3	-11	-16	23 *	-11
25	LC	2	69	3	26 *	-9	-27	-13	37 *
26	LC	77 *	2	3	18	35 *	-6	-16	-29
27	LC	11	11	6	72 *	-20	-19	-22	37 *
28	P	18	15	65 *	3	-29	-29	49 *	-13
29	P	9	20	9	63 *	-21	-31	-11	44 *
30	LC	0	78 *	14	7	-10	24 *	-22	-6
31	LC	3	95 *	1	0	-20	26 *	-14	-8
32	LC	3	3	25	69 *	-10	-14	-10	19 *
33	P	63 *	12	2	22	47 *	-27	-13	-28
34	P	34	60 *	3	3	-43	51 *	-15	-13
35	P	12	17	63 *	7	-24	-28	47 *	-18
36	WD	62 *	17	19	2	43 *	-18	-32	-11
37	WD	9	19	0	72 *	-16	-29	2	35 *
38	WD	2	66 *	32	0	-9	40 *	-38	0
39	WD	96 *	1	1	1	14 *	-7	-7	-12
40	WD	34 *	2	63	1	17 *	-9	-14	-1

Note. An asterisk indicates the item that is keyed as correct.  
Endorsement rates are given as percentages.

## Discussion

The purpose of this document was to provide detailed statistical information on the entrance and graduation reading and listening test items currently used by CLA at the University of Minnesota to test French, German, and Spanish proficiency based on the data collected from May 15, 1995 to September 22, 1995. In this study, the items making up the reading and listening tests were analyzed using CTT and IRT. Each of the tests was fitted to a 2-parameter logistic IRT model. With sufficiently large sample sizes for performing IRT analyses, special attention was given in this report to the interpretation of the Spanish entrance and graduation reading and listening test results. This information can be used for the revision of these tests or to guide future test development.

The items in these four Spanish tests fit the 2-parameter IRT model well. Evidence for this can be found in the fact that the vast majority of standardized residual statistics reported in tables 3, 5, 7, and 9 are below the cutoff of 2.0. The 1-parameter, Rasch, model appears inappropriate for this data because of the great range in discrimination ( $a$  parameter) values reported across items. Recall, the Rasch model assumes that all item discriminations are equal and have a value of 1.0. When larger sample sizes become available, it might be interesting to explore the data using a 3-parameter analysis.

The preliminary results based on the IRT 2-parameter logistic model for the Spanish entrance and graduation reading and listening tests provide an important resource for future exam revision especially when combined with the information gained from the CTT item statistics. From these analyses, the test developer can identify and rework items that are too easy or too difficult based on their intended functions, items that discriminate poorly, and even distracters that discriminate poorly. This analysis, coupled with the earlier quantitative and content analysis provided by Chalhoub-Deville et al., (1996a), Chalhoub-Deville et al. (1996b), and Chalhoub-Deville et al. (1996c) provide a wealth of information for detailed test evaluation and revision.

The current study had modest sample sizes for the CLA French and German tests. In the future, analyses similar to these performed here must be conducted as testing continues and more data becomes available. Furthermore, as the sample sizes increase, there will be increasing confidence in the accuracy of the results for all the test analyses.

Finally, the data contained in this analysis, as with any statistical analysis of test items, should always be interpreted in conjunction with item analysis performed by

expert item writers and content experts. Item statistics can point to particular item problems, but only the content and item experts can determine the root of the problem and make the appropriate revisions.

## References

- American Council on the Teaching of Foreign Languages. (1982). *ACTFL provisional proficiency guidelines*. Hastings-on-Hudson, New York: Author.
- Assessment Systems Corporation. (1995). *User's manual for the XCALIBRE Marginal Maximum-Likelihood Estimation Program*. Minneapolis: Author.
- Barnes, B., Klee, C., & Wakefield, R. (1991). Reconsidering the foreign language requirement: From seat-time to proficiency in the Minnesota experience. In S. Magnan (Ed.), *Challenges in the 1990s for college foreign language programs* (pp. 55-69). Boston: Heinle and Heinle.
- Chalhoub-Deville, M., McCollum Lozier, V., Sweet, G., Juengling, F., Klein, F., Mueller, I., Alcaya, C., & Kim, S. (1995). *Technical report #1: Qualitative and quantitative critique of the entrance and graduation University of Minnesota CLA language proficiency tests in French, German, and Spanish*. Minneapolis: The Center for Advanced Research on Language Acquisition.
- Chalhoub-Deville, M., Mueller, I., McCollum Lozier, V., & Juengling, F. (1996). *Qualitative and quantitative review of the University of Minnesota CLA language proficiency tests in German for entrance and graduation*. Minneapolis: The Center for Advanced Research on Language Acquisition.
- Henning, G. (1987). *A guide to language testing*. Cambridge, MA: Newbury House Publishers.
- Lange, D. (1988). Developing and implementing proficiency oriented tests for a new language requirement at the University of Minnesota: Issues and problems for implementing the ACTFL/ETS/ILR proficiency guidelines. pp. 275-290.
- Lord, F. M. (1952). A theory of test scores. *Psychometric Monograph*, 7.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielsen and Lydiche.
- Weiss, D. J., & Yoes, M. (1988). Item response theory. In R. K. Hambleton and J. Zaal (eds.), *Handbook of Testing*. Amsterdam: North-Holland.

## Appendix A

# UNIVERSITY OF MINNESOTA

Twin Cities Campus

National Language Resource Center  
Center for Advanced Research on  
Language Acquisition

Suite 105  
1313 Fifth Street S.E.  
Minneapolis, MN 55414  
612-627-1870  
Fax: 612-627-1875

March 20, 1995

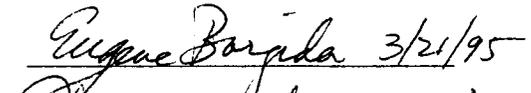
This letter of agreement is being drawn, based on Gene Borgida's request of February 27, 1995.

This letter is to confirm that the Center for Advanced Research on Language Acquisition (CARLA) is authorized to access the College of Liberal Arts' (CLA) entrance and graduation proficiency test data in French, German, and Spanish. It is hereby guaranteed that CARLA will ensure the anonymity of all student records and will cover the cost of: (a) storing the scannable test data on diskette; (b) programming to substitute students' IDs; (c) hiring a person to enter the non-scannable test data on the computer; and (d) other research-related costs that arise during the process. CARLA will coordinate with the Language Center for the data entry of these non-scannable tests.

Name

Signature

Eugene Borgida,  
Associate Dean, CLA



Micheline Chalhoub-Deville,  
Assessment Manager, CARLA



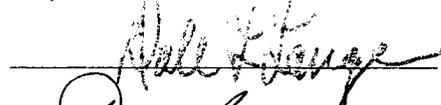
Andrew Cohen,  
Director, NLRC



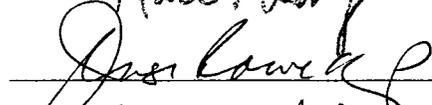
Monica Eden,  
Director, Testing Program



Dale Lange,  
Director, CARLA



Jenise Rowekamp,  
Director, Language Center



Ray Wakefield,  
Coordinator, Assessment Project



## Appendix B

Table 1

French Entrance Reading Item Table: 2 Parameter Logistic Model Analysis, N=274

Item	Function	a	b	Flags	Residual	Pr. Correct	Pt. Biserial	Pt. Biserial $\theta$
1	WU	0.68	-1.92		0.78	0.83	0.40	0.39
2	WU	0.62	-1.63		0.87	0.78	0.38	0.38
3	WU	0.67	-3.00	P	1.07	0.97	0.13	0.10
4	WU	0.77	-2.75		0.30	0.93	0.38	0.33
5	WU	0.75	-3.00	P	0.76	0.98	0.20	0.16
6	WU	0.63	-2.38		0.76	0.88	0.34	0.30
7	LC	0.69	-2.98	P	0.60	0.94	0.27	0.25
8	LC	0.82	-3.00	P	0.66	0.97	0.38	0.29
9	LC	0.57	-0.79		0.49	0.64	0.38	0.39
10	LC	0.82	-2.73		0.40	0.94	0.42	0.36
11	LC	0.59	-2.56		1.11	0.88	0.27	0.24
12	LC	0.65	-1.67		0.59	0.79	0.41	0.39
13	LC	0.78	-3.00	P	0.31	0.96	0.33	0.27
14	LC	0.63	-1.83		0.78	0.81	0.38	0.36
15	LC	0.90	-3.00	P	1.05	0.99	0.39	0.27
16	LC	0.81	-3.00	P	0.88	0.98	0.23	0.20
17	LC	0.52	-0.50		0.66	0.59	0.33	0.34
18	LC	0.84	-2.07		0.61	0.88	0.49	0.45
19	LC	0.77	-3.00	P	0.61	0.97	0.23	0.23
20	LC	0.88	-3.00	P	0.95	0.99	0.33	0.24
21	LC	0.82	-3.00	P	0.70	0.98	0.30	0.24
22	LC	0.80	-3.00	P	1.03	0.98	0.26	0.17
23	LC	0.81	-2.43		0.82	0.91	0.41	0.39
24	LC	0.84	-2.50		0.41	0.92	0.45	0.40
25	LC	0.85	-2.33		0.42	0.91	0.48	0.42
26	LC	0.75	-3.00	P	0.52	0.96	0.26	0.25
27	P	0.56	-3.00	P	1.25	0.93	0.12	0.09
28	P	0.81	-2.57		0.40	0.92	0.44	0.37
29	P	0.63	-1.81		0.76	0.81	0.38	0.36
30	P	0.85	-3.00	P	1.12	0.99	0.31	0.22
31	P	0.71	-1.03		0.29	0.70	0.48	0.49
32	P	0.52	-1.54		1.23	0.74	0.28	0.28
33	P	0.62	-1.12		0.50	0.70	0.42	0.42
34	P	0.49	-0.13		0.95	0.53	0.28	0.30
35	P	0.65	-1.43		0.49	0.76	0.43	0.42
36	WD	0.38	0.55		1.80	0.42	0.09	0.06
37	WD	0.53	-1.34		1.05	0.72	0.32	0.32
38	WD	0.53	-0.47		0.49	0.58	0.33	0.36
39	WD	0.55	-2.13		0.86	0.82	0.28	0.25
40	WD	0.68	-1.87		1.02	0.82	0.41	0.39

Table 2

## French Entrance Reading Test Option Effectiveness (2PLM), N=274

Item	Function	Endorsement Rate				Item-Theta Correlation (Pt. Biserial $\theta$ )			
		1	2	3	4	1	2	3	4
1	WU	83 *	0	16	1	39 *	0	-37	-10
2	WU	11	1	78 *	10	-25	0	38 *	-27
3	WU	2	97 *	1	0	-7	10 *	-12	4
4	WU	4	1	2	93 *	-29	-11	-12	33 *
5	WU	0	1	0	98 *	0	-14	-8	16 *
6	WU	88 *	5	0	7	30 *	-21	-6	-19
7	LC	2	94 *	1	3	-9	25 *	-18	-18
8	LC	2	1	97 *	0	-18	-24	29 *	-9
9	LC	64 *	21	4	11	39 *	-14	-26	-27
10	LC	0	3	4	94 *	0	-25	-25	36 *
11	LC	88 *	8	1	3	24 *	-15	-8	-19
12	LC	14	79 *	4	3	-26	39 *	-26	-12
13	LC	1	1	96 *	1	-19	-18	27 *	-11
14	LC	1	81 *	16	2	-7	36 *	-29	-22
15	LC	99 *	0	0	1	27 *	-13	0	-24
16	LC	1	0	98 *	1	-12	-13	20 *	-10
17	LC	13	4	59 *	21	-21	-22	34 *	-12
18	LC	8	2	88 *	2	-33	-20	45 *	-22
19	LC	1	1	0	97 *	-17	-14	0	23 *
20	LC	1	0	99 *	0	-14	-21	24 *	0
21	LC	98 *	0	0	1	24 *	-6	-7	-23
22	LC	0	98 *	1	1	0	17 *	3	-25
23	LC	3	4	91 *	1	-19	-29	39 *	-17
24	LC	92 *	3	4	1	40 *	-24	-29	-9
25	LC	3	91 *	5	2	-24	42 *	-32	-12
26	LC	1	0	2	96 *	-16	-3	-19	25 *
27	P	0	1	93 *	6	7	-20	9 *	-5
28	P	92 *	1	1	5	37 *	-23	-12	-27
29	P	4	12	81 *	3	-21	-20	36 *	-21
30	P	99 *	1	0	0	22 *	-24	-13	4
31	P	4	70 *	8	18	-26	49 *	-28	-25
32	P	10	3	74 *	14	-20	-15	28 *	-11
33	P	19	5	70 *	7	-19	-19	42 *	-30
34	P	9	53 *	26	12	-4	30 *	-22	-12
35	P	7	8	8	76 *	-21	-23	-21	42 *
36	WD	5	53	42 *	0	-17	1	-6	0
37	WD	5	20	72 *	3	-22	-18	32 *	-13
38	WD	2	58 *	2	38	-19	36 *	-15	-28
39	WD	82 *	1	7	9	25 *	-20	-13	-14
40	WD	82 *	10	7	1	39 *	-23	-26	-17

Notes. An asterisk indicates the item that is keyed as correct.  
Endorsement rates are given as percentages.

Table 3

French Entrance Listening Item Table: 2 Parameter Logistic Model Analysis, N=226

Item	Function	a	b	Flags	Residual	Pr. Correct	Pt. Biserial	Pt. Biserial $\theta$
1	WU	.74	-3.00	P	0.40	0.97	0.18	0.18
2	WU	.83	-1.17		0.50	0.75	0.52	0.51
3	WU	.76	-2.71		0.57	0.93	0.29	0.28
4	WU	.77	-3.00	P	0.38	0.97	0.22	0.21
5	WU	.68	-1.96		0.39	0.85	0.36	0.34
6	WU	.75	-2.01		0.35	0.86	0.40	0.38
7	WU	.65	-1.45		0.71	0.77	0.40	0.38
8	WU	.64	-1.88		0.56	0.83	0.34	0.32
9	LC	.52	0.97		1.00	0.34	0.29	0.29
10	LC	.78	-0.27		0.35	0.57	0.56	0.56
11	LC	.73	-1.75		0.53	0.83	0.41	0.40
12	LC	.81	0.99		0.44	0.29	0.51	0.59
13	LC	.79	-1.91		0.42	0.86	0.44	0.41
14	LC	.67	-2.59		0.53	0.91	0.27	0.24
15	LC	.76	-1.78		0.49	0.84	0.44	0.41
16	LC	.69	-0.25		0.62	0.56	0.48	0.49
17	LC	.67	-1.53		0.67	0.79	0.39	0.38
18	LC	.88	-1.92		0.69	0.88	0.48	0.44
19	LC	.86	-0.48		0.53	0.62	0.58	0.59
20	LC	.74	-2.38		0.40	0.90	0.34	0.32
21	LC	.65	-2.19		0.62	0.87	0.30	0.29
22	LC	-	-		-	-	-	-
23	LC	.59	-1.64		0.61	0.78	0.32	0.30
24	LC	.79	-2.11		0.40	0.88	0.40	0.38
25	LC	.84	-2.44		0.35	0.92	0.39	0.35
26	LC	.46	-1.29		1.81	0.69	0.12	0.11
27	LC	.77	-2.95		0.65	0.96	0.28	0.25
28	LC	.45	-0.77		1.70	0.62	0.13	0.11
29	P	.63	-0.18		1.02	0.55	0.45	0.44
30	P	.57	-2.01		0.87	0.83	0.23	0.23
31	P	.62	-0.10		0.57	0.53	0.42	0.44
32	P	.65	-1.84		0.39	0.82	0.36	0.33
33	P	.48	-0.07		0.90	0.52	0.24	0.22
34	P	.71	-1.27		0.41	0.75	0.45	0.44
35	P	.51	-0.90		0.89	0.65	0.30	0.27
36	P	.44	0.63		1.56	0.41	0.13	0.14
37	WD	.67	-1.75		0.51	0.81	0.37	0.36
38	WD	.66	-2.03		0.99	0.85	0.33	0.32
39	WD	.54	-0.87		0.81	0.65	0.32	0.32
40	WD	.69	-1.54		0.36	0.79	0.42	0.40

Note. All examinees answered item 22 correctly. Statistics cannot be calculated.

Table 4

French Entrance Listening Test Option Effectiveness (2PLM), N=226

Item	Function	Endorsement Rate				Item-Theta Correlation (Pt. Biserial $\theta$ )			
		1	2	3	4	1	2	3	4
1	WU	0	2	1	97 *	-8	-12	-9	-18 *
2	WU	75 *	5	8	12	51 *	-20	-34	-25
3	WU	0	93 *	0	6	0	28 *	-12	-26
4	WU	1	1	97 *	1	-18	-8	21 *	-9
5	WU	5	0	10	85 *	-9	0	-34	34 *
6	WU	86 *	12	1	1	38 *	-29	-17	-18
7	WU	5	15	77 *	4	-17	-26	38 *	-17
8	WU	6	3	83	9	-21	-16	32	-17
9	LC	4	34 *	8	54	-15	29 *	-30	-4
10	LC	57 *	2	36	4	56 *	-12	-51	-7
11	LC	11	83 *	4	2	-30	40 *	-17	-15
12	LC	44	26	1	29 *	-52	0	-6	59 *
13	LC	12	1	86 *	1	-36	-10	41 *	-16
14	LC	91 *	1	1	7	24 *	-5	-2	-24
15	LC	3	84 *	10	3	-13	41 *	-29	-23
16	LC	3	24	56 *	16	-6	-19	49 *	-40
17	LC	4	8	79 *	9	-19	-24	38 *	-19
18	LC	4	88 *	5	4	-14	44 *	-31	-26
19	LC	62 *	19	15	4	59 *	-29	-39	-19
20	LC	1	3	5	90 *	-14	-21	-19	32 *
21	LC	2	87 *	11	-	-12	29 *	-24	-
22	LC	-	-	-	-	-	-	-	-
23	LC	78 *	6	5	11	30 *	-7	-27	-16
24	LC	0	88 *	10	1	-9	38 *	-33	-16
25	LC	1	5	2	92 *	-17	-26	-17	35 *
26	LC	2	23	69 *	6	7	-2	11 *	-23
27	LC	96 *	2	3	0	25 *	-14	-20	0
28	LC	23	14	62 *	1		-21	11 *	-13
29	P	4	55 *	24	16	-18	44 *	-32	-12
30	P	5	83 *	9	4	-12	23 *	-18	-7
31	P	17	17	53 *	12	-20	-14	44 *	-28
32	P	82 *	2	13	2	33 *	-4	-32	-9
33	P	52 *	2	4	42	22 *	-5	-13	-17
34	P	4	13	8	75 *	-10	-34	-20	44 *
35	P	5	65 *	16	14	1	27 *	-21	-15
36	P	10	32	41 *	18	-4	-22	14 *	11
37	WD	3	7	9	81 *	-17	-16	-25	36 *
38	WD	5	85 *	2	8	-18	32 *	-14	-20
39	WD	27	5	65 *	3	-24	-8	32 *	-17
40	WD	4	10	8	79 *	-12	-23	-27	40 *

Notes. Item 21 has only 3 alternatives.

All examinees answered item 22 correctly. Statistics cannot be calculated

An asterisk indicates the item that is keyed as correct.

Endorsement rates are given as percentages.

Table 5

German Entrance Reading Item Table: 2 Parameter Logistic Model Analysis, N=193

Item	Function	a	b	Flags	Residual	Pr. Correct	Pt. Biserial	Pt. Biserial $\theta$
1	WU	0.78	-3.00	P	1.02	0.99	0.16	0.15
2	WU	0.68	-2.75		0.52	0.93	0.28	0.27
3	WU	0.68	-3.00	P	0.80	0.97	0.06	0.07
4	LC	0.70	-3.00	P	0.36	0.95	0.26	0.23
5	LC	0.65	-2.41		0.80	0.89	0.31	0.29
6	LC	0.61	-2.08		0.58	0.84	0.31	0.30
7	LC	0.59	-1.83		0.47	0.80	0.33	0.31
8	LC	0.63	-3.00	P	0.92	0.94	0.15	0.16
9	LC	0.47	-0.39		1.04	0.56	0.23	0.22
10	LC	0.76	-2.99	P	0.68	0.95	0.33	0.30
11	LC	0.66	-1.40		0.44	0.75	0.43	0.41
12	LC	0.70	-2.10		0.51	0.86	0.39	0.37
13	LC	0.65	-2.82		0.79	0.93	0.27	0.22
14	P	0.54	-2.62		1.16	0.88	0.14	0.12
15	P	0.57	-0.18		0.54	0.54	0.40	0.40
16	P	0.53	-1.96		0.97	0.80	0.23	0.20
17	LC	0.56	0.05		0.82	0.49	0.39	0.39
18	LC	0.69	-2.07		0.49	0.85	0.39	0.37
19	LC	0.61	-1.42		0.87	0.74	0.38	0.37
20	LC	0.58	-1.12		0.89	0.69	0.36	0.36
21	LC	0.57	-1.24		0.92	0.71	0.34	0.34
22	LC	0.55	-0.89		0.70	0.65	0.34	0.34
23	P	0.63	-3.00	P	0.89	0.94	0.13	0.14
24	LC	0.62	1.45		0.77	0.23	0.40	0.42
25	LC	0.72	-3.00	P	0.90	0.96	0.26	0.22
26	P	0.64	-1.60		0.77	0.78	0.40	0.38
27	P	0.53	1.16		0.86	0.30	0.30	0.30
28	LC	0.71	-3.00	P	0.74	0.96	0.25	0.21
29	LC	0.70	-1.89		0.43	0.83	0.42	0.40
30	LC	0.71	-0.31		0.18	0.57	0.53	0.54
31	P	0.52	-0.66		0.81	0.61	0.30	0.31
32	P	0.61	-2.04		0.55	0.83	0.32	0.30
33	P	0.62	-1.81		0.51	0.81	0.36	0.34
34	LC	0.62	-2.86		0.75	0.92	0.17	0.17
35	LC	0.67	0.09		0.36	0.48	0.50	0.52
36	LC	0.69	-1.46		0.53	0.77	0.46	0.43
37	LC	0.55	-1.79		0.99	0.78	0.28	0.26
38	LC	0.67	0.50		0.47	0.40	0.50	0.52
39	WD	0.60	-1.89		0.63	0.81	0.34	0.31
40	WD	0.72	-3.00	P	0.94	0.98	0.07	0.07

Table 6

German Entrance Reading Test Option Effectiveness (2PLM), N=193

Item	Function	Endorsement Rate				Item-Theta Correlation (Pt. Biserial $\theta$ )			
		1	2	3	4	1	2	3	4
1	WU	99 *	1	0	1	15 *	-7	0	-14
2	WU	3	4	1	93 *	-15	-22	-2	27 *
3	WU	1	1	97 *	1	1	-11	7 *	-2
4	LC	0	95 *	1	4	0	23 *	-10	-21
5	LC	1	5	6	89 *	-9	-22	-13	29 *
6	LC	15	84 *	1	0	-28	30 *	-10	0
7	LC	14	1	5	80 *	-14	-11	-30	31 *
8	LC	1	94 *	3	3	-13	16 *	-10	-7
9	LC	12	22	9	56 *	-19	-1	-16	22 *
10	LC	2	2	95 *	1	-16	-22	30 *	-12
11	LC	75 *	15	0	10	41 *	-25	0	-29
12	LC	5	86 *	5	4	-24	37 *	-22	-15
13	LC	1	5	93 *	2	-10	-13	22 *	-16
14	P	4	88 *	6	3	-21	12 *	-2	3
15	P	54 *	3	18	25	40 *	-13	-25	-17
16	P	4	80 *	5	11	-13	20 *	-12	-9
17	LC	49 *	16	32	3	39 *	1	-39	-10
18	LC	9	85 *	2	4	-24	37 *	-16	-21
19	LC	74 *	6	12	8	37 *	-14	-32	-10
20	LC	0	69 *	30	1	0	36 *	-35	-11
21	LC	3	8	71 *	19	-8	-14	34 *	-27
22	LC	2	11	65 *	22	-13	-14	34 *	-24
23	P	1	94 *	1	4	-19	14 *	-3	-6
24	LC	23 *	28	26	23	42 *	-4	-35	1
25	LC	96 *	3	1	1	22 *	-16	-12	-11
26	P	78 *	9	4	9	38 *	-8	-26	-28
27	P	34	30 *	19	16	-20	30 *	-6	-6
28	LC	96 *	1	3	0	21 *	5	-25	0
29	LC	2	2	14	83 *	-20	-4	-35	40 *
30	LC	26	12	57 *	6	-34	-25	54 *	-15
31	P	34	2	61 *	3	-24	-18	31 *	-7
32	P	2	83 *	14	1	-12	30 *	-26	-7
33	P	6	4	10	81 *	-16	-7	-27	34 *
34	LC	92 *	8	0	0	17 *	-17	0	0
35	LC	9	48 *	39	4	-17	52 *	-37	-17
36	LC	77 *	6	16	2	43 *	-21	-33	-9
37	LC	11	78 *	4	7	-8	26 *	-12	-23
38	LC	40 *	36	9	15	52 *	-32	-12	-19
39	WD	2	9	8	81 *	-10	-29	-7	31 *
40	WD	1	98 *	1	1	2	7 *	-4	-11

Notes. An asterisk indicates the item that is keyed as correct.  
Endorsement rates are given as percentages.

Table 7

German Entrance Listening Item Table 2 Parameter Logistic Model Analysis, N=192

Item	Function	a	b	Flags	Residual	Pr. Correct	Pt. Biserial	Pt. Biserial $\theta$
1	WU	0.70	-2.76		0.32	0.93	0.33	0.29
2	LC	0.60	-0.69		0.93	0.61	0.41	0.41
3	WU	0.51	-2.36		1.41	0.84	0.13	0.12
4	WU	0.67	-3.00	P	0.63	0.95	0.21	0.19
5	WU	0.70	-3.00	P	0.83	0.96	0.26	0.20
6	WU	0.78	-3.00	P	1.11	0.99	0.02	0.04
7	WU	0.65	-1.79		0.56	0.81	0.41	0.38
8	WU	0.82	-3.00	P	1.21	0.99	0.38	0.24
9	LC	0.74	-2.78		0.63	0.94	0.37	0.32
10	LC	0.62	-0.10		0.47	0.51	0.44	0.46
11	LC	0.60	-2.10		0.69	0.84	0.27	0.29
12	LC	0.79	-2.82		0.74	0.95	0.48	0.37
13	P	0.64	-0.20		0.24	0.53	0.46	0.48
14	LC	0.52	-0.43		1.10	0.56	0.32	0.32
15	LC	0.46	2.74	R K	2.90	0.11	-0.18	-0.18
16	LC	0.70	-2.12		0.34	0.86	0.42	0.38
17	P	0.53	-1.26		0.96	0.69	0.30	0.28
18	LC	0.72	0.01		0.43	0.49	0.54	0.56
19	LC	0.49	-0.45		1.09	0.56	0.25	0.25
20	LC	0.68	-1.85		0.97	0.82	0.39	0.39
21	LC	0.65	-2.43		0.78	0.89	0.31	0.29
22	LC	0.75	-3.00	P	0.61	0.97	0.33	0.26
23	P	0.49	-0.31		1.31	0.54	0.28	0.26
24	LC	0.57	0.05		0.86	0.48	0.37	0.41
25	P	0.75	-1.55		0.92	0.79	0.49	0.48
26	LC	0.72	-2.91		0.15	0.94	0.34	0.28
27	LC	0.81	-3.00	P	1.16	0.99	0.32	0.23
28	LC	0.63	-2.98	P	0.56	0.94	0.19	0.17
29	LC	0.70	-3.00	P	0.59	0.96	0.24	0.19
30	LC	0.66	-2.78		0.68	0.92	0.28	0.25
31	LC	0.78	-3.00	P	1.03	0.98	0.31	0.21
32	LC	0.53	-0.85		0.92	0.63	0.31	0.30
33	P	0.59	-2.33		0.85	0.86	0.23	0.24
34	P	0.61	-1.20		0.71	0.70	0.41	0.39
35	LC	0.68	-3.00	P	0.78	0.97	0.04	0.06
36	LC	0.61	-3.00	P	0.95	0.94	0.11	0.12
37	WD	0.76	-3.00	P	0.97	0.97	0.33	0.24
38	WD	0.70	-1.93		0.76	0.76	0.46	0.47
39	WD	0.74	-2.94		0.34	0.95	0.35	0.30
40	WD	0.66	-2.50		0.65	0.90	0.30	0.29

Table 8

German Entrance Listening Test Option Effectiveness (2PLM), N=192

Item	Function	Endorsement Rate				Item-Theta Correlation (Pt. Biserial $\theta$ )			
		1	2	3	4	1	2	3	4
1	WU	6	1	93 *	1	-31	2	29 *	-4
2	LC	8	61 *	1	30	-6	2	29 *	-37
3	WU	15	84 *	1	1	-9	12 *	-12	-4
4	WU	4	1	95 *	0	-17	-10	19 *	0
5	WU	96 *	1	3	0	20 *	-21	-12	0
6	WU	1	99 *	0	0	-4	4 *	0	0
7	WU	17	81 *	2	1	-31	38 *	-11	-19
8	WU	0	1	0	99 *	0	-24	0	24 *
9	LC	2	94 *	2	2	-13	32 *	-22	-19
10	LC	3	51 *	38	8	-16	46 *	-32	-17
11	LC	84 *	4	5	8	29 *	-7	-11	-26
12	LC	95 *	2	3	1	37 *	-23	-25	-13
13	P	17	28	2	53 *	-25	-30	-10	48 *
14	LC	56 *	27	15	2	32 *	-19	-20	-3
15	LC	56	31	11 *	2	-11	29	-18 *	-12
16	LC	2	2	86 *	10	-9	-8	38 *	-36
17	P	18	69 *	1	12	-29	28 *	-6	-4
18	LC	43	49 *	5	4	-44	56 *	-15	-15
19	LC	56 *	10	13	21	25 *	-12	-8	-14
20	LC	3	82 *	11	3	-10	39 *	-32	-18
21	LC	4	6	89 *	1	-18	-21	29 *	-6
22	LC	97 *	3	1	0	26 *	-23	-13	0
23	P	6	28	11	54 *	-17	-14	-8	26 *
24	LC	48 *	32	7	13	41 *	-29	-7	-15
25	P	79 *	11	3	6	48 *	-33	-19	-23
26	LC	94 *	2	3	2	28 *	-13	-19	-16
27	LC	1	1	99 *	0	-11	-21	23 *	0
28	LC	1	94 *	1	5	-6	17 *	-5	-15
29	LC	1	96 *	3	1	-6	19 *	-13	-13
30	LC	7	1	92 *	1	-20	-13	25 *	-8
31	LC	1	1	98 *	0	-13	-17	21 *	0
32	LC	4	63 *	24	9	-21	30 *	-5	-29
33	P	86 *	3	6	6	24 *	-5	-11	-22
34	P	4	70 *	8	18	-15	39 *	-17	-28
35	LC	97 *	1	1	2	6 *	-6	-5	-2
36	LC	2	1	94 *	3	0	-7	12 *	-12
37	WD	2	0	0	97 *	-28	0	0	24 *
38	WD	18	76 *	5	2	-35	47 *	-24	-11
39	WD	95 *	5	0	0	30 *	-30	0	0
40	WD	3	90 *	8	0	-11	29 *	-26	0

Notes. An asterisk indicates the item that is keyed as correct.  
Endorsement rates are given as percentages.

Table 9

French Graduation Reading Item Table: 2 Parameter Logistic Model Analysis, N=219

Item	Function	a	b	Flags	Residual	Pr. Correct	Pt. Biserial	Pt. Biserial $\theta$
1	WU	0.58	-2.46		0.84	0.87	0.28	0.28
2	WU	0.66	-3.00	P	0.73	0.97	0.12	0.14
3	WU	0.54	-2.78		1.06	0.89	0.18	0.18
4	WU	0.68	-3.00	P	0.56	0.96	0.24	0.21
5	WU	0.50	-1.63		1.07	0.73	0.30	0.28
6	WU	0.72	-3.00	P	1.11	0.98	0.15	0.13
7	LC	0.65	-3.00	P	0.67	0.95	0.24	0.19
8	LC	0.39	-1.36		1.96	0.66	0.08	0.10
9	LC	-	-	-	-	-	-	-
10	LC	0.61	-3.00	P	0.63	0.93	0.22	0.21
11	LC	0.46	-0.37		1.16	0.53	0.28	0.29
12	LC	0.64	-3.00	P	1.13	0.97	0.07	0.07
13	LC	0.67	-2.98	P	0.20	0.94	0.32	0.28
14	LC	0.65	-2.82		0.53	0.92	0.32	0.29
15	LC	0.78	-3.00	P	1.27	0.99	0.19	0.15
16	LC	0.74	-3.00	P	1.08	0.99	0.14	0.13
17	LC	-	-	-	-	-	-	-
18	LC	0.68	-3.00	P	0.78	0.97	0.14	0.14
19	LC	0.57	-2.27		0.74	0.84	0.30	0.30
20	LC	0.73	-3.00	P	0.91	0.98	0.23	0.19
21	LC	0.63	-3.00	P	0.75	0.96	0.12	0.13
22	LC	0.60	-2.99	P	0.48	0.93	0.24	0.21
23	LC	0.53	-0.74		0.99	0.59	0.40	0.40
24	P	0.49	-2.09		1.07	0.79	0.24	0.22
25	P	0.61	-2.83		0.54	0.91	0.26	0.25
26	P	0.60	-3.00	P	0.89	0.95	0.06	0.11
27	P	0.63	-3.00	P	0.40	0.95	0.22	0.20
28	LC	0.59	-0.93		0.86	0.63	0.44	0.46
29	LC	0.73	-3.00	P	0.90	0.98	0.25	0.20
30	LC	0.70	-3.00	P	0.61	0.96	0.29	0.24
31	LC	0.64	-2.96	P	0.47	0.93	0.29	0.26
32	P	0.54	-2.15		0.97	0.82	0.29	0.28
33	LC	0.62	-2.62		0.47	0.89	0.34	0.30
34	LC	0.58	-3.00	P	0.64	0.94	0.14	0.13
35	LC	0.68	-2.99	P	0.16	0.94	0.34	0.28
36	WD	0.73	-3.00	P	1.27	0.99	0.12	0.10
37	WD	-	-	-	-	-	-	-
38	LC	0.73	-2.52		0.69	0.91	0.46	0.39
39	WD	0.66	-3.00	P	0.70	0.95	0.26	0.21
40	WD	0.66	-2.99	P	0.31	0.94	0.30	0.25

- All examinees answered items 9, 17, 37 correctly. Statistics cannot be calculated.

Table 10

## French Graduation Reading Test Option Effectiveness (2PLM), N=219

Item	Function	Endorsement Rate				Item-Theta Correlation (Pt. Biserial $\theta$ )			
		1	2	3	4	1	2	3	4
1	WU	5	0	87 *	9	-21	0	28 *	-18
2	WU	2	1	97 *	0	-10	-9	14 *	0
3	WU	2	89 *	1	8	-8	18 *	-12	-12
4	WU	2	1	96 *	1	-10	-10	21 *	-17
5	WU	73 *	0	26	1	28 *	5	-27	-12
6	WU	0	1	1	98 *	0	-16	-3	13 *
7	LC	95 *	0	4	0	19 *	0	-18	-8
8	LC	66 *	33	0	0	10 *	-9	2	-7
9	LC	-	-	-	-	-	-	-	-
10	LC	1	93 *	2	3	-12	21 *	-9	-15
11	LC	4	1	42	53 *	-12	-4	-24	29 *
12	LC	0	0	3	97 *	-8	0	-5	7 *
13	LC	4	94 *	2	0	-20	28 *	-15	-11
14	LC	6	2	92 *	0	-23	-17	29 *	-2
15	LC	0	0	99 *	0	-6	-15	15 *	0
16	LC	0	99 *	1	0	0	13 *	-13	0
17	LC	-	-	-	-	-	-	-	-
18	LC	3	97 *	0	0	-14	14 *	0	0
19	LC	6	6	84 *	3	-14	-20	30 *	-15
20	LC	0	2	0	98 *	-9	-17	0	19 *
21	LC	96 *	2	2	0	13 *	-7	-8	-8
22	LC	3	1	3	93 *	-7	-3	-24	21 *
23	LC	4	22	15	59 *	-15	-21	-23	40 *
24	P	21	0	0	79 *	-22	0	0	22 *
25	P	2	3	5	91 *	-6	-23	-12	25 *
26	P	3	95 *	1	0	-8	11 *	-8	-2
27	P	2	2	95 *	1	-10	-15	20 *	-9
28	LC	1	35	63 *	1	-6	-44	46 *	-6
29	LC	0	98 *	2	0	0	20 *	-20	0
30	LC	96 *	3	0	0	24 *	-19	-15	-6
31	LC	6	1	93 *	0	-25	-7	26 *	0
32	P	14	1	82 *	3	-22	-16	28 *	-7
33	LC	89 *	6	3	1	30 *	-20	-17	-14
34	LC	0	94 *	1	5	-8	13 *	-8	-9
35	LC	4	1	1	94 *	-20	-19	-8	28 *
36	WD	1	99 *	0	0	-2	10 *	-15	0
37	WD	-	-	-	-	-	-	-	-
38	LC	4	3	91 *	2	-24	-19	39 *	-22
39	WD	95 *	1	2	2	21 *	-11	-13	-12
40	WD	0	6	0	94 *	-10	-24	0	25 *

Notes. An asterisk indicates the item that is keyed as correct.  
Endorsement rates are given as percentages.

Table 11

French Graduation Listening Item Table: 2 Parameter Logistic Model Analysis, N=220

Item	Function	a	b	Flags	Residual	Pr. Correct	Pt. Biserial	Pt. Biserial $\theta$
1	WU	0.59	-2.85		0.53	0.91	0.24	0.23
2	WU	0.68	<b>-3.00</b>	PK	1.51	0.99	<b>-0.00</b>	<b>0.01</b>
3	WU	0.65	<b>-3.00</b>	P	0.93	0.97	0.10	0.12
4	WU	0.60	<b>-3.00</b>	P	0.90	0.96	0.09	0.09
5	LC	0.53	-2.75		0.80	0.89	0.20	0.19
6	LC	0.49	-0.92		1.02	0.65	0.31	0.30
7	LC	0.56	-0.47		0.62	0.59	0.42	0.43
8	LC	0.43	-0.29		1.13	0.54	0.22	0.22
9	LC	0.52	-2.27		1.48	0.83	0.20	0.23
10	LC	0.70	-1.75		0.57	0.82	0.50	0.46
11	LC	0.43	-1.81		1.44	0.75	0.13	0.12
12	LC	0.68	<b>-3.00</b>	P	1.00	0.98	0.15	0.15
13	LC	0.59	-2.92		0.56	0.92	0.23	0.22
14	LC	0.61	<b>-3.00</b>	P	0.44	0.95	0.23	0.20
15	LC	0.63	<b>-3.00</b>	P	0.35	0.95	0.26	0.22
16	LC	0.58	-1.48		1.05	0.75	0.39	0.38
17	LC	0.51	-1.52		0.98	0.73	0.30	0.29
18	LC	0.63	<b>-3.00</b>	P	0.67	0.96	0.19	0.17
19	LC	0.48	-1.09		0.92	0.67	0.31	0.29
20	LC	0.44	0.32		0.99	0.45	0.26	0.27
21	P	0.53	0.09		0.96	0.48	0.43	0.41
22	P	0.60	-1.38		0.87	0.74	0.41	0.41
23	P	0.57	-0.71		0.45	0.63	0.42	0.42
24	P	0.57	-1.17		0.78	0.70	0.42	0.40
25	LC	0.52	-1.49		0.62	0.73	0.33	0.31
26	LC	0.45	-2.52		1.81	0.84	0.10	0.10
27	LC	0.59	-1.86		0.54	0.80	0.38	0.35
28	LC	0.54	-1.85		0.95	0.79	0.31	0.31
29	P	0.57	-0.62		0.35	0.61	0.44	0.43
30	P	0.41	-1.23		1.73	0.66	0.12	0.11
31	P	0.59	-0.40		0.27	0.58	0.44	0.46
32	LC	0.61	-2.85		0.40	0.92	0.27	0.25
33	LC	0.57	<b>-3.00</b>	P	0.91	0.94	0.17	0.16
34	LC	0.47	-1.05		0.86	0.66	0.30	0.27
35	P	0.54	-0.56		0.41	0.60	0.40	0.39
36	P	0.57	-1.12		0.59	0.69	0.41	0.40
37	P	0.53	-0.89		0.51	0.65	0.38	0.37
38	LC	0.61	-2.66		0.17	0.90	0.32	0.28
39	WD	0.53	-2.57		0.91	0.87	0.22	0.21
40	WD	0.52	-2.53		0.83	0.86	0.22	0.20

Table 12

## French Graduation Listening Test Option Effectiveness (2PLM), N=220

Item	Function	Endorsement Rate				Item-Theta Correlation (Pt. Biserial $\theta$ )			
		1	2	3	4	1	2	3	4
1	WU	91 *	0	0	8	23 *	-4	-12	-20
2	WU	0	1	0	99 *	4	-4	0	1 *
3	WU	0	97 *	0	2	0	12 *	-1	-13
4	WU	0	0	96 *	4	0	4	9 *	-11
5	LC	6	1	3	89 *	-14	-10	-6	19 *
6	LC	65 *	15	15	5	30 *	-18	-17	-9
7	LC	59 *	5	33	3	43 *	-2	-39	-12
8	LC	7	9	54 *	30	-11	-2	22 *	-16
9	LC	1	83 *	13	3	-5	23 *	-21	-6
10	LC	3	1	14	82 *	-13	-10	-41	46 *
11	LC	1	75 *	22	2	-14	12 *	-3	-17
12	LC	0	2	0	98 *	0	-15	0	15 *
13	LC	1	92 *	0	6	-7	22 *	-13	-18
14	LC	5	0	0	95 *	-15	-9	-13	20 *
15	LC	1	2	95 *	2	-5	-14	22 *	-17
16	LC	75 *	18	6	1	38 *	-25	-24	-12
17	LC	20	73 *	0	7	-32	29 *	0	-1
18	LC	2	1	1	96 *	-11	-7	-10	17 *
19	LC	7	6	20	67 *	0	-4	-32	29 *
20	LC	45 *	4	46	5	27 *	-24	-23	12
21	P	48 *	2	20	30	41 *	-7	-31	-16
22	P	4	13	74 *	10	-12	-28	41 *	-22
23	P	17	10	63 *	10	-25	-17	42 *	-18
24	P	22	70 *	4	4	-34	40 *	-16	-5
25	LC	4	19	4	73 *	-28	-18	-9	31 *
26	LC	2	11	3	84 *	-6	-3	-9	10 *
27	LC	80 *	3	3	13	35 *	-5	-16	-31
28	LC	11	5	5	79 *	-17	-11	-22	31 *
29	P	16	21	61 *	2	-30	-22	43 *	-7
30	P	66 *	18	15	0	11 *	-10	-3	0
31	P	27	5	10	58 *	-31	-19	-16	46 *
32	LC	1	92 *	4	3	-11	25 *	-19	-11
33	LC	1	94 *	3	2	-4	16 *	-9	-13
34	LC	10	4	20	66 *	-14	-6	-18	27 *
35	P	60 *	2	3	35	39 *	-13	-10	-33
36	P	3	24	69 *	5	-6	-37	40 *	-7
37	P	6	23	6	65 *	-5	-30	-16	37 *
38	LC	90 *	7	2	1	28 *	-19	-14	-17
39	WD	1	11	87 *	1	-8	-17	21 *	-11
40	WD	5	86 *	2	6	-9	20 *	-13	-13

Notes. An asterisk indicates the item that is keyed as correct.  
Endorsement rates are given as percentages.