

# American Sign Language 506 Credit by Examination Test Specifications Report Measurement and Evaluation Center

## Introduction

### **Purpose**

At the request of the Department of Linguistics, the Measurement and Evaluation Center, a component of the Division of Instructional Innovation and Assessment, conducted a validity study to assist the department in determining decision scores for the award of credit by examination for American Sign Language 506. UT Austin policy requires that score ranges used as the basis for the award of credit by examination shall reflect the same standard of performance for students who receive credit by examination as for students who complete the coursework.

### **Course Description**

American Sign Language (ASL) 506 is a five-credit introductory course in a sequence of four undergraduate courses. ASL 506 covers material that provides extensive exposure to American Sign Language so that students can develop beginning communication skills used with Deaf people. Emphasis is on skills necessary to produce and comprehend ASL. Students are also introduced to topics in Deaf culture and ASL grammar.

## Method

### Participants

A total of 196 undergraduate students enrolled during fall 2002 in ten sections of ASL 506, First-year American Sign Language I, took the UT Austin Test in American Sign Language at the end of the semester. Five students were not included in the study because of incomplete data. These students were primarily upper-division students, at the sophomore and junior year level.

### Materials

### **Test Description**

Faculty members in the Department of Linguistics developed The UT Austin Test for Credit in ASL 506. The test consists of 75 items in two parts: (1) 25 items to assess factual knowledge of ASL and of Deaf culture and (2) 50 items to assess comprehension. The test is administered using the computer-based test (CBT) format with true/false or multiple-choice objective questions. The comprehension items are weighted two points each for a total possible weighted score of 125. The total time allotment for the test is 90 minutes.

The factual knowledge tested includes elements of Deaf culture and the basic linguistics of American Sign Language. The comprehension skills tested includes receptive skills in American Sign Language, that is, the ability to understand ASL signs and sentences. For each comprehension test item, examinees view a video clip of a person presenting an ASL sign or a simple sign sentence. Examinees are allowed to

view the video clip a maximum of two times. Content areas for the comprehension items include: identification of finger-spelled words, identification of facial expressions used in ASL to distinguish sentence types, comprehension of spatial descriptions in ASL, comprehension of ASL number signs, identification of isolated ASL signs, knowledge of linguistic contexts that require particular negative signs, and identification of signs for shapes.

### Procedure

#### **Examination Process**

Students enrolled in ASL 506 in fall 2002 took the UT Austin Test for Credit as their final exam in the course. The test was administered by computer in a secured lab located in the Measurement and Evaluation Center. Students were given a five-day period in which to schedule and complete the exam. The Associate Dean of the College of Liberal Arts granted permission to waive the final exam period time restriction for the course so that students could take the exam over the five-day period instead of in the traditional one-day time-slot format.

#### **Validation Process**

Concurrent validity evidence in the form of preliminary course grades, which included all course grades up to but not including the final exam, was collected prior to the administration of the final exam. Grades were on a scale of 0 to 4, with 0 = *F*, 1 = *D*, 2 = *C*, 3 = *B*, and 4 = *A*. Correlation and regression analyses were conducted to examine the relationship between preliminary course grades and scores on the ASL 506 test.

#### **Reliability and SEM**

An internal consistency measure of reliability was computed using Cronbach alpha, and the standard error of measurement (SEM) was also calculated to yield another measure of reliability. The SEM indicates the amount of error to allow for when interpreting individual test results, and is derived from the reliability coefficient using the formula

$$SEM = s\sqrt{1 - r_n}$$

where  $s$  = standard deviation and  $r_n$  = reliability coefficient.

#### **Item Analysis**

In order to investigate the effectiveness of individual items, an item analysis was performed that consisted of determining (1) the difficulty of each item, expressed as the proportion of test takers who answered the item correctly, (2) the discriminating power of each item, or how well each item discriminated between high and low scorers, and (3) the effectiveness of each incorrect multiple-choice alternative response in enticing low-scoring test takers to choose it. Items that are very easy (0.80 – 1.00) or very difficult (0.00 – 0.20) can be viewed as candidates for modification or removal from the test. Items with low discriminating power (near zero or negative) can also be examined for possible changes. And incorrect alternatives (i.e., distractors) can be analyzed for their effectiveness and modified if necessary.

## Score Normalization

When the distribution of raw scores is skewed, as in the present study, any linear transformation of that distribution will also be skewed. To avoid this, a normalization procedure can be performed to achieve the benefits of a normalized distribution, which are that every score has concise statistical meaning. For a given score, the percentage of examinees who achieved scores above and below that score is known exactly. Additionally, differences across individuals or within the individual across exams can be compared when scores have been normalized with common means and units of measurement (Magnusson, 1967).

The first step in normalizing the scores was to tabulate the frequencies for each weighted score. Then the cumulative frequencies for all the weighted scores were calculated using the formula

cumulative frequency =

number of scores *below* a given score value + number of scores *at* a given score value.

In order to normalize the cumulative frequencies, a nonlinear transformation was used in which the cumulative frequency for a given score was computed as the cumulative frequency for the class means in each class interval; that is, the number of scores in the distribution that fell below the class mean in the interval represented by the original score. The normalized cumulative frequencies were computed using the formula

normalized cumulative frequency =

number of scores *below* a given score value +  $\frac{\text{number of scores at a given score value}}{2}$

In this computation the assumption was made that there was a rectangular distribution of scores in the class interval. This is not completely accurate in a normal distribution, but as long as there are sufficiently small intervals, the errors produced will be so small that they will lack practical significance (Magnusson, 1967).

In the next step in the normalization process, the normalized cumulative frequencies were transformed to proportions using the formula

proportion =  $\frac{\text{normalized cumulative frequency at a given score value}}{N}$

where N = total sample size.

The final step was to convert the proportions to normalized standard scores, or z-scores. The z-score is the standard score that corresponds to a given percentage on a normal distribution.

## Scale Transformation

In order to transform the weighted scores into a standard scale that can be used with all UT Austin tests for credit by examination, a scale transformation was used in which the weighted scores were first converted to normalized z-scores as described above. The normalized z-scores were then converted to the UT Austin scale using a mean of 50 and a standard deviation of 21.06. This linear transformation converted the z-scores to a scale of 1 to 100, with all scores rounded to the nearest integer. Any score below one (1) was rounded up to a score of one and any score above 100 was rounded down to a score of 100. The formula for computing the UT Austin scale is

$$\text{UT Austin scale score} = (\text{z-score} * 21.06) + 50$$

## Results

### Validation

The correlation between preliminary course grades and weighted scores on the ASL 506 exam was 0.4488. ASL 506 weighted exam scores were used as the predictor variable and preliminary course grades as the criterion (dependent) variable in a regression analysis to compute the prediction equation for expected grade given a test score. The regression equation was found to be

$$\text{expected grade} = (\text{exam score} * 0.04042) - 1.2584$$

A similar regression analysis was used to predict expected score on the exam given preliminary grades. The resulting regression equation was found to be

$$\text{expected score} = (\text{preliminary grade} * 4.9837) + 93.138$$

### Reliability, SEM, and Item Analysis

For the ASL 506 exam the reliability estimate was 0.76 with standard error of measurement for the weighted scores of 4.65. Five items were selected for analysis: four with a difficulty value of  $p = 1.00$ , meaning that everyone answered them correctly, and one with a negative discrimination value, meaning that it did not contribute to discriminating between high and low ability examinees. Editorial revisions were made to the distractors for four of the items, and one was left unchanged.

### Determining Decision Scores

In February 2003, MEC staff met with three American Sign Language faculty members to discuss possible criterion score ranges using the weighted scores. For the threshold for awarding credit-by-examination with a grade of *C*, MEC staff recommended selecting a weighted score of 103, the score that corresponded to the expected score for students who had a preliminary course grade of *C* in ASL 506. Similarly, the thresholds for awarding grades of *B* and *A* were recommended to be 108 and 113, respectively, the expected scores for students who had preliminary course grades of *B* and *A*, respectively.

The ASL faculty approved the recommendations and these three criterion score ranges were adopted for awarding credit by examination:

Weighted Test Score Range	Credit and Letter Grade
113-125	Credit with Grade of <i>A</i>
108-112	Credit with Grade of <i>B</i>
103-107	Credit with Grade of <i>C</i>

### **Transformation of Weighted Scores to the UT Austin Scale**

After the decision scores ranges were selected, the weighted scores were transformed to the normalized UT Austin score scale to make the interpretation of scores more meaningful. On the UT Austin scale, score points are equidistant with a zero point, so that scores may be compared mathematically. For example, a score of 80 may be interpreted as twice as good as a score of 40. Additionally, the distribution of expected grades based on the weighted scores holds true for the UT Austin scale scores as well. UT Austin scale scores yield the same distribution of credit-by-examination grade assignments as do the weighted scores.

Table 1 shows the relationships among the weighted scores, the normalization steps, converted z-scores, and UT Austin scale scores. The rows in bold indicate the threshold criterion scores for credit-by-examination grade assignments. Table 2 summarizes the definitions, formulas, and meanings of variables for the column headings of Table 1.

Table 1  
 Obtained Weighted Scores, Frequencies, Cumulative Frequencies, Score Interval Cumulative  
 Frequencies, Transformed UT Austin Scale Scores, and Percentiles for ASL 506 Test

Weighted Score	Frequency <i>f</i>	Cumulative Frequency <i>cf1</i>	Score Interval Cumulative Frequency <i>cf2</i>	<i>p</i> ( $p=cf2/N$ )	<i>z</i> Score	UT Austin Scale Score	UT Austin Scale Score Rounded	Percentile
81	1	1	0.5	.0026	-2.794	-8.84	1	1
82	1	2	1.5	.0076	-2.428	-1.13	1	1
83	1	3	2.5	.0128	-2.232	2.99	3	1
84	0	3	3	.0153	-2.162	4.47	4	2
85	2	5	4	.0204	-2.046	6.91	7	2
86	1	6	5.5	.0281	-1.909	9.80	10	3
87	1	7	6.5	.0332	-1.836	11.33	11	3
88	1	8	7.5	.0383	-1.771	12.70	13	4
89	1	9	8.5	.0434	-1.713	13.92	14	4
90	3	12	10.5	.0536	-1.611	16.07	16	5
91	0	12	12	.0612	-1.545	17.46	17	6
92	1	13	12.5	.0638	-1.524	17.90	18	6
93	2	15	14	.0714	-1.465	19.15	19	7
94	3	18	16.5	.0842	-1.377	21.00	21	8
95	4	22	20	.1020	-1.270	23.25	23	10
96	4	26	24	.1224	-1.163	25.51	26	12
97	2	28	27	.1378	-1.090	27.04	27	14
98	3	31	29.5	.1505	-1.034	28.22	28	15
99	2	33	32	.1633	-0.9810	29.34	29	16
100	2	35	34	.1735	-0.9404	30.20	30	17
101	3	38	36.5	.1862	-0.8920	31.21	31	19
102	3	41	39.5	.2015	-0.8363	32.39	32	20
<b>103</b>	<b>2</b>	<b>43</b>	<b>42</b>	<b>.2143</b>	<b>-0.7916</b>	<b>33.33</b>	<b>33</b>	<b>21</b>
104	6	49	46	.2347	-0.7235	34.76	35	23
105	13	62	55.5	.2832	-0.5734	37.92	38	28
106	9	71	66.5	.3393	-0.4144	41.34	41	34
107	7	78	74.5	.3801	-0.3052	43.57	44	38
<b>108</b>	<b>8</b>	<b>86</b>	<b>82</b>	<b>.4184</b>	<b>-0.2060</b>	<b>45.66</b>	<b>46</b>	<b>42</b>
109	6	92	89	.4541	-0.1153	47.57	48	45
110	10	102	97	.4949	-0.0128	49.73	50	49
111	8	110	106	.5408	0.1024	52.16	52	54
112	12	122	116	.5918	0.2322	54.89	55	59
<b>113</b>	<b>5</b>	<b>127</b>	<b>124.5</b>	<b>.6352</b>	<b>0.3457</b>	<b>57.28</b>	<b>57</b>	<b>64</b>
114	6	133	130	.6633	0.4215	58.88	59	66
115	9	142	137.5	.7015	0.5287	61.13	61	70
116	8	150	146	.7449	0.6585	63.86	64	74
117	9	159	154.5	.7883	0.8005	66.85	67	79
118	9	168	163.5	.8342	0.9709	70.45	70	83
119	5	173	170.5	.8699	1.126	73.71	74	87
120	9	182	177.5	.9056	1.314	77.67	78	91
121	4	186	184	.9388	1.545	82.54	83	94
122	3	189	187.5	.9566	1.713	86.08	86	96
123	5	194	191.5	.9770	1.995	92.01	92	98
124	0	194	194	.9898	2.319	98.84	99	99
125	2	196	195	.9949	2.569	104.10	100	99

Table 2  
Definitions, Formulas, and Variables Used in Table 1

Table 1 Column Heading	Meaning
Weighted Score	The obtained raw score (0-125) from the ASL examination
Frequency ( $f$ )	The frequency of each weighted score
Cumulative Frequency ( $cf1$ )	The cumulative frequency at each weighted score value
Score Interval Cumulative Frequency ( $cf2$ )	The cumulative frequency at the midpoint of each weighted score value interval $cf2 = (\# \text{ of scores below a score}) + (\# \text{ of scores at a score interval}) / 2$
Proportion ( $p$ )	The proportion of the score interval cumulative frequency to the total number of examinee scores; $p = cf2 / N$
z Score	The standard score corresponding to a given $p$ in a normal distribution
UT Austin Scale Score	A linear transformation of a normalized z score to a score with a mean of 50 and standard deviation of 21.06; $scale \text{ score} = (z * 21.06) + 50$
UT Austin Scale Score Rounded	Closest positive integer values (1-100) for UT Austin scale scores
Percentile	The percent of scores falling at or below this value on the UT Austin scale score distribution

### Correlation and Regression Analyses with UT Austin Scale Scores

The correlation between preliminary course grades and UT Austin scale scores was 0.4402. UT Austin scale scores were used as the predictor variable and preliminary course grades as the criterion (dependent) variable in a regression analysis to compute the prediction equation for expected grade given a test score. The regression equation was found to be

$$\text{expected grade} = (\text{UT Austin scale score} * 0.01821) + 2.2266$$

A similar regression analysis used to predict expected score on the exam given preliminary grades resulted in the prediction equation

$$\text{expected score} = (\text{preliminary grade} * 10.6418) + 16.6546$$

### Comparison of Weighted and Normalized UT Austin Scale Scores Used as Criterion Score Ranges

A comparison was made between the grade distributions that would result from using credit-by-examination criterion score ranges based on the weighted and the UT Austin scale scores. Table 3 shows a comparison of the parameters for the two score distributions, and Table 4 shows a comparison of the criterion score ranges and of the resulting credit eligibility for the two score scales.

Table 3

Comparison of the Frequency Distributions for Weighted Scores and for UT Austin Scale Scores for the UT Austin Test for Credit in American Sign Language 506

Distribution Characteristic	Weighted Scores	UT Austin Scale Scores
Mean	108.78	50.05
Standard Deviation	9.49	20.67
SEM	4.65	10.13
Correlation with Preliminary Grades	0.4488	0.4402

Table 4

Comparison of the Criterion Score Ranges and of Credit Eligibility for ASL 506 Based on Weighted Scores and on UT Austin Scale Scores

Credit Eligibility Status	Weighted Scores			UT Austin Scale Scores		
	Criterion Score Range	N	Expected Grade	Criterion Score Range	N	Expected Grade
No credit	81-102	41	2.02-2.86	1-32	41	2.24-2.81
C	103-107	37	2.90-3.07	33-45	37	2.83-3.05
B	108-112	44	3.11-3.27	46-56	44	3.06-3.25
A	113-125	74	3.31-3.79	57-100	74	3.26-4.00

As the comparison of criterion score ranges and credit eligibility status under the two score scales shows, using UT Austin score scale yields the same results as using the weighted scores. The frequency distributions for both scales are identical, and the expected grade ranges vary only slightly, due to measurement error in the prediction equations for expected grade. Therefore, the UT Austin scale can be used to present results with the added benefits of ease of interpretation, clarity, and consistency.

#### Reference

Magnusson, D. (1967). Test theory. Reading, MA: Addison-Wesley Publishing Company.