

**STANDARD-SETTING
STUDY REPORT**

**Standard-Setting Study of the
University of Texas at Austin
Credit by Exam
for Credit in Biology 311D
Spring 2006**

May 31, 2006

Mi-Suk Shim
Misuk.shim@austin.texas.edu



Division of Instructional Innovation and Assessment

research, evaluation and assessment section

THE UNIVERSITY OF TEXAS AT AUSTIN

<http://www.utexas.edu/academic/diia>

P.O. Box 7246
2616 Wichita Street
Austin, TX 78713-7246

Contents

I. Introduction	3
1. Purpose	3
2. Course Descriptions	3
II. Method	4
1. Participants	4
2. Test Description	4
3. Procedure	4
Examination Process	4
Validation Process	5
III. Results	6
1. Summary Statistics	6
2. Cut-score Analyses	6
3. Item Analyses	11
IV. Recommendation	14
1. Decision Score Determination	14
2. Other Considerations	15
V. References	16

I. Introduction

Purpose

At the request of the Department of Biological Sciences, the Instructional Assessment and Evaluation (IAE) group, a component of the Division of Instructional Innovation and Assessment (DIIA), conducted a validity study for the Credit by Exam (CBE) in Biology to assist the department in examining the decision scores for student placement and the award of credit by examination for Biology 311D. The Biological Sciences department has recently changed its sequence of courses from BIO 213 and BIO 214 to BIO 311D. 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 Descriptions

Biology (BIO) 311D, Introductory Biology II, is a three-credit introductory course that follows BIO 311C, Introductory Biology I. The completion of BIO 311C with a grade of C or better is a prerequisite for BIO 311D. BIO 311D provides an introduction to the topics of diversity, ecology, structure, and function (*Biological Sciences Course Descriptions*, 2006).

II. Method

Participants

A total of 1,071 undergraduate students enrolled during spring 2006 in BIO 311D (encompasses 63 sections) took the credit by examination test at the end of the semester. Among them, 59 students were not included in the final data set due to incomplete information or dropping the course. Final data resulted in a sample size of 1,012.

Test Description

The Credit by Exam (CBE) in Biology 311D was developed by faculty members in the Biological Sciences department at the University of Texas at Austin. This is a 90 minute test with 50 multiple-choice questions covering diversity, ecology, structure, and function. Each item has only one correct answer. The test is scored as number of correct answers, thus results in a possible score range of 0-50.

Procedure

Examination process. Students enrolled in BIO 311D during spring 2006 took the test that was designed for Credit by Exam (CBE) in Biology 311D during the last week of classes. Students were informed about the purpose of the test, which was to establish the criteria for credit-qualifying scores from the test for future use. Students would not receive credit by examination for this administration. However, students were required to take the test for 5% of their final grade. Students were directed to register for the test online in the DIIA test registration system. The test was administered on the last class day and the test was scheduled in large groups in classroom settings using UT testing personnel (*Testing Period and Test Descriptions, 2005-2006*).

Validation process. Concurrent validity evidence in the form of preliminary course grades, which included all course grades except for the UT CBE in BIO 311D and final exam, was collected prior to the administration of the UT CBE in BIO 311D. Grades were on a scale of 0 to 4, with 0 = F, 1 = D, 2 = C, 3 = B, and 4 = A. Using the data consisting of preliminary course grade and test score, IAE staff calculated the coefficient of correlation between test scores and preliminary course grades and tabulated a two-way frequency table of test scores by preliminary course grades. A linear regression equation was used to calculate expected test scores and expected grades. Finally, using six possible guidelines for making cut-score decisions, a candidate for the cut-off score was determined for each of the six guidelines.

III. Results

Summary Statistics

Descriptive statistics for BIO 311D course are shown in Table 1. Correlation between the preliminary course grades and the test scores on UT CBE in BIO 311D was .55, which is statistically significant at $p < .001$ level.

Table 1

Descriptive Statistics on the UT CBE in BIO 311D in Relation to Student Performance in BIO 311D

Preliminary Grade	N	Mean	SD	Minimum	Maximum
Total (All grades)	1012	31.6	6.65	12	48
Grade A	143	37.73	5.49	21	48
Grade B	403	33.35	5.94	12	46
Grade C	336	29.2	5.07	15	42
Grade D	93	25.75	5.01	15	41
Grade F	37	24.05	6.16	15	36

Cut-score Analyses

The results of the standard-setting study of BIO 311D are presented in Tables 2-5. Table 2 shows a frequency distribution for test scores by preliminary course grade. In order to shorten Table 2, grade frequencies for test scores of 12-18 have been summarized in one row, and the same has been done for test scores of 45-48. The second column shows the expected grade corresponding to each test score, calculated from the expected grade regression equation. The bottom row shows the expected test score corresponding to each preliminary grade, calculated from the expected score regression equation. Those regression equations are:

$$\text{Expected Grade} = (\text{Test Score} \times 0.08016) - 0.01346$$

$$\text{Expected Score} = (\text{Preliminary Grade} \times 3.7863) + 22.0268$$

From Table 2, the criterion for awarding credit based on the same standard of performance for students who take an examination as for students who complete course work can be determined. According to Table 2, the minimally satisfactory grade of C (2.00) may be expected for students with a test score of 26, or for students receiving the minimally satisfactory grade of C the expected test score is 30.

For convenience in interpreting the data for awarding credit by examination, Table 3 collapses the preliminary course grades into two categories – Unsatisfactory (preliminary course grades *F* and *D*) and Satisfactory (preliminary course grades *C*, *B* and *A*) – and shows the frequencies in these two categories by test score.

Table 4 shows the accuracy of student placement (supposing that the students who participated in the study would have been placed based on the UT CBE in BIO 311D score) for each test score as a hypothetical cut-off score. The table shows the numbers and percentages of students, for each test score as a cut-off score, whose placement would have been Correct, Too High, and Too Low, as determined by the students' preliminary course grade. Table 4 indicates the accuracy with which students would be placed by possible scores ranging from 19 to 30 (the range shown in Table 4 was truncated to the most relevant portion from the whole range 12 to 48).

Table 2

*Scores on the UT CBE in BIO 311D in Relation to Student Performance in BIO 311D:
Frequency distributions, Expected Grades, and Expected Scores*

Test Score	Expected Grade	Preliminary Grade in BIO 311D					Total N
		0 F	1 D	2 C	3 B	4 A	
45-48	3.59-3.83	0	0	0	3	14	17
44	3.51	0	0	0	4	8	12
43	3.43	0	0	0	15	9	24
42	3.35	0	0	1	7	7	15
41	3.27	0	1	1	10	10	22
40	3.19	0	0	4	18	12	34
39	3.11	0	1	2	25	9	37
38	3.03	0	0	5	24	5	34
37	2.95	0	0	7	27	11	45
36	2.87	2	0	12	26	7	47
35	2.79	1	1	24	30	6	62
34	2.71	0	3	21	20	15	59
33	2.63	0	2	19	29	8	58
32	2.55	1	3	17	18	7	46
31	2.47	2	5	30	24	3	64
30	2.39	2	4	23	26	5	60
29	2.31	4	8	23	19	1	55
28	2.23	0	4	24	10	2	40
27	2.15	0	8	18	14	0	40
26	2.07	0	6	22	12	0	40
25	1.99	6	8	15	12	1	42
24	1.91	2	8	18	8	0	36
23	1.83	3	9	16	6	0	34
22	1.75	0	5	10	2	0	17
21	1.67	3	3	8	4	3	21
20	1.59	1	2	10	3	0	16
19	1.51	1	4	3	1	0	9
12-18	0.95-1.43	9	8	3	6	0	26
Total		37	93	336	403	143	1012
Percent		4%	9%	33%	40%	14%	100%
Expected Score		22	26	30	33	37	

Table 3

Scores on the UT CBE in BIO 311D in Relation to Student Performance in BIO 311D: Combined Frequency Distributions

Test Score	Preliminary Grade in BIO 311D		Total N
	Unsatisfactory (0,1)	Satisfactory (2, 3, 4)	
45-48	0	17	17
44	0	12	12
43	0	24	24
42	0	15	15
41	1	21	22
40	0	34	34
39	1	36	37
38	0	34	34
37	0	45	45
36	2	45	47
35	2	60	62
34	3	56	59
33	2	56	58
32	4	42	46
31	7	57	64
30	6	54	60
29	12	43	55
28	4	36	40
27	8	32	40
26	6	34	40
25	14	28	42
24	10	26	36
23	12	22	34
22	5	12	17
21	6	15	21
20	3	13	16
19	5	4	9
12-18	17	9	26
Total	130	882	1012
Percent	13%	87%	100%
Mean Score	25.3	32.5	31.6
Standard Deviation	5.42	6.3	6.65

Table 4

Scores on the UT CBE in BIO 311D in Relation to Student Performance in BIO 311D: Possible Decision Scores and Corresponding Accuracies of Placement

Placement category	Cumulative Number of Students			Percent of Students in Each Placement Category			Overall Accuracy of Placement		
	Unsatisfactory 0, 1 (N=130)	Satisfactory 2-4 (N=882)		Unsatisfactory	Satisfactory		Placement Accuracy	Number of Students	% of Students
30 - up	Too High 28	608	Correct	Too High 22%	69%	Correct	Too High	28	3%
Below 30	Correct 102	274	Too Low	Correct 78%	31%	Too Low	Correct	710	70%
29 - up	Too High 40	651	Correct	Too High 31%	74%	Correct	Too High	40	4%
Below 29	Correct 90	231	Too Low	Correct 69%	26%	Too Low	Correct	741	73%
28 - up	Too High 44	687	Correct	Too High 34%	78%	Correct	Too High	44	4%
Below 28	Correct 86	195	Too Low	Correct 66%	22%	Too Low	Correct	773	76%
27 - up	Too High 52	719	Correct	Too High 40%	82%	Correct	Too High	52	5%
Below 27	Correct 78	163	Too Low	Correct 60%	18%	Too Low	Correct	797	79%
26 - up	Too High 58	753	Correct	Too High 45%	85%	Correct	Too High	58	6%
Below 26	Correct 72	129	Too Low	Correct 55%	15%	Too Low	Correct	825	82%
25 - up	Too High 72	781	Correct	Too High 55%	89%	Correct	Too High	72	7%
Below 25	Correct 58	101	Too Low	Correct 45%	11%	Too Low	Correct	839	83%
24 - up	Too High 82	807	Correct	Too High 63%	91%	Correct	Too High	82	8%
Below 24	Correct 48	75	Too Low	Correct 37%	9%	Too Low	Correct	855	84%
23 - up	Too High 94	829	Correct	Too High 72%	94%	Correct	Too High	94	9%
Below 23	Correct 36	53	Too Low	Correct 28%	6%	Too Low	Correct	865	85%
22 - up	Too High 99	841	Correct	Too High 76%	95%	Correct	Too High	99	10%
Below 22	Correct 31	41	Too Low	Correct 24%	5%	Too Low	Correct	872	86%
21 - up	Too High 105	856	Correct	Too High 81%	97%	Correct	Too High	105	10%
Below 21	Correct 25	26	Too Low	Correct 19%	3%	Too Low	Correct	881	87%
20 - up	Too High 108	869	Correct	Too High 83%	99%	Correct	Too High	108	11%
Below 20	Correct 22	13	Too Low	Correct 17%	1%	Too Low	Correct	891	88%
19 - up	Too High 113	873	Correct	Too High 87%	99%	Correct	Too High	113	11%
Below 19	Correct 17	9	Too Low	Correct 13%	1%	Too Low	Correct	890	88%
							Too Low	9	1%

Item Analyses

There were test items that require closer attention. Items with less than 0.20 of R(IT), an item-total coefficient of correlation, are listed below. R(IT) indicates the item's discrimination value, and items with values less than 0.20 do not yield much information about differences among the abilities of the persons tested.

Item No. 6		Keyed Response = D					
Split	Omit	A	B	C	D	E	SUM
1	0	4	16	0	103	4	127
2	1	3	12	0	105	6	127
3	0	1	25	0	92	9	127
4	0	2	26	1	84	14	127
Sum	1	10	79	1	384	33	508
Mean	34.0	34.4	30.5	12.0	32.5	29.1	
P TOT = 1.00				P = 0.76		R(IT) = 0.14	

Item No. 21		Keyed Response = C					
Split	Omit	A	B	C	D	E	SUM
1	0	1	14	52	44	16	127
2	0	5	27	53	28	14	127
3	0	9	30	34	38	16	127
4	0	12	29	34	36	16	127
Sum	0	27	100	173	146	62	508
Mean	0.0	26.1	30.6	33.4	32.5	31.7	
P TOT = 1.00				P = 0.34		R(IT) = 0.15	

Item No. 46		Keyed Response = B					
Split	Omit	A	B	C	D	E	SUM
1	0	7	97	7	6	10	127
2	1	9	102	5	8	2	127
3	0	17	91	11	5	3	127
4	0	30	70	11	7	9	127
Sum	1	63	360	34	26	24	508
Mean	36.0	28.1	32.8	30.1	32.4	32.6	
P TOT = 1.00				P = 0.71		R(IT) = 0.18	

Item No. 50		Keyed Response = D						
Split	Omit	A	B	C	D	E	SUM	
1	1	1	4	0	120	1	127	
2	1	2	6	1	111	6	127	
3	0	1	4	0	118	4	127	
4	1	3	12	1	103	7	127	
Sum	3	7	26	2	452	18	508	
Mean	32.3	29.0	28.7	23.0	32.4	28.7		
P TOT = 0.99				P = 0.89		R(IT) = 0.17		

The easier an item is the higher P (an index of the item’s level of difficulty for a given group of persons) value is. Items with difficulty values close to 1 or at/below the guessing level (0.20 for an item with 5 choices) may need to be investigated. Items below show this problem.

Item No. 2		Keyed Response = D						
Split	Omit	A	B	C	D	E	SUM	
1	0	0	1	2	122	2	127	
2	0	0	4	2	118	3	127	
3	0	1	1	2	115	8	127	
4	0	2	3	4	109	9	127	
Sum	0	3	9	10	464	22	508	
Mean	0.0	25.0	30.4	29.3	32.3	29.0		
P TOT = 1.00				P = 0.91		R(IT) = 0.13		

Item No. 25		Keyed Response = A						
Split	Omit	A	B	C	D	E	SUM	
1	0	126	0	0	0	1	127	
2	0	123	1	0	3	0	127	
3	0	121	1	3	1	1	127	
4	0	115	2	3	4	3	127	
Sum	0	485	4	6	8	5	508	
Mean	0.0	32.3	27.0	26.7	25.0	27.2		
P TOT = 1.00				P = 0.95		R(IT) = 0.18		

Item No. 41		Keyed Response = C					
Split	Omit	A	B	C	D	E	SUM
1	0	1	0	125	1	0	127
2	0	1	2	120	1	3	127
3	0	3	2	122	0	0	127
4	0	6	3	112	6	0	127
Sum	0	11	7	479	8	3	508
Mean	0.0	26.5	26.4	32.4	21.7	34.3	
P TOT = 1.00				P = 0.94	R(IT) = 0.22		

Item No. 42		Keyed Response = B					
Split	Omit	A	B	C	D	E	SUM
1	0	0	126	0	0	1	127
2	0	0	126	0	0	1	127
3	0	3	121	1	1	1	127
4	0	3	110	5	5	4	127
Sum	0	6	483	6	6	7	508
Mean	0.0	27.0	32.4	21.0	22.7	25.7	
P TOT = 1.00				P = 0.95	R(IT) = 0.26		

Item 38 warrants more investigation. Both P and R(IT) value are very low. It is possible the item's key is incorrect. A majority of students chose answer choice B as correct one, and the smallest number of students chose the correct answer E.

Item No. 38		Keyed Response = E					
Split	Omit	A	B	C	D	E	SUM
1	0	21	57	30	16	3	127
2	0	19	56	32	20	0	127
3	0	20	55	32	19	1	127
4	0	17	54	32	22	2	127
Sum	0	77	222	126	77	6	508
Mean	0.0	32.7	32.1	31.7	31.5	32.5	
P TOT = 1.00				P = 0.01	R(IT) = 0.01		

IV. Recommendation

Decision Score Determination

Several guidelines in choosing decision scores for course placement and awarding credit by examination can be found in the professional literature. Table 5 shows scores suggested by six guidelines for selecting decision scores for course placement and awarding credit by examination. The first column explains the rationale for each guideline and indicates specific reference tables used to get each decision score. Based on these guidelines, possible decision scores for UT CBE in BIO 311D range from 20 to 30.

Table 5
Scores on the UT CBE in BIO 311D in Relation to Student Performance in BIO 311D: Scores Suggested by Six Guidelines for Use in Selecting Decision Scores

	Guideline	Test Score
1	Expected Score for students whose performance in course was just minimally satisfactory (i.e., students with preliminary grades of C; see Expected Score row at bottom of Table 2)	30
2	Score for which Expected Grade was just minimally satisfactory (i.e., C; see Expected Grade column in Table 2.)	26
3	Score for which percents of errors of students in each academic performance category (Unsatisfactory, Satisfactory) were most nearly equal. (See % Too High and % Too Low values in middle columns of Table 4.)	29
4	Score for which overall percents of errors were most nearly equal. (See % Too High and % Too Low values in last column of Table 4.)	24
5	Score that would have cut off (or held back) approximately the same number of students as were in the Unsatisfactory performance group. (See Table 3 for number of students in the Unsatisfactory group and the test score that most nearly identifies that number of low-scoring students.)	23
6	Score that would have maximized overall accuracy of placement. (See number Correct in next-to-last column of Table 4.)	20

The recommendations for awarding credit by examination with a letter grade of C, B, or A were shown in Table 6.

Table 6

Scores Suggested for Use in Selecting Decision Scores for awarding Letter Grades C, B, or A Based on Expected Score Guideline

BIO 311C

30-34	CR	
35-38	B	After 12th class day in Fall 2006 ---> 30-50 CR only
39-50	A	

BIO 311D

30-32	CR	
33-36	B	After 12th class day in Fall 2006 ---> 30-50 CR only
37-50	A	

Other considerations

When the results of item analyses are considered, it is possible that some items need to be discarded or modified. All the items listed in the report should be reviewed for accuracy and clarity in the question and answer choices. In particular, item 38 needs to be modified if there is not a miskey problem. If item 38 is rescored or replaced, cut scores could increase.

V. References

Biological Sciences Course Descriptions. (2006). Retrieved May 25, 2006, from the University of Texas at Austin, College of Natural Sciences Web site:
<http://www.biosci.utexas.edu/bsac/syllabi/>

Testing Period and Test Descriptions. (2005-2006). Retrieved January 25, 2006, from the University of Texas at Austin, Division of Instructional Innovation and Assessment Website: <http://www.utexas.edu/academic/mec/cbe/testperiod.html>