BIOL999: Sample Class - Example Lecture Exam

Instructor Report for Dr. Franklund - Summer 2012

Introduction

Welcome! This is an automated report concerning the class' performance on a recent exam. This analysis is based upon classical test theory and is presented in several sections. A brief description of each section is provided below along with links to enable rapid navigation within this document. The document can be used to evaluate the performance of the class and the exam and to identify weaknesses and strengths therein.

Summary of exam scores: This section reports the overall distribution of class score for the assessment. Descriptive statistics are provided and some of their implications are discussed.

Part B Listing of student scores: The performance of each student in the class is reported in this section. The actual (observed) score is provided along with an estimate of the students' true score. The observed and true scores are plotted with a score band that represents the 90% confidence interval for the students' true score.

Outcome assessment: Some or all of the items in this exam have been mapped to specific course outcomes. The results of these questions have been compiled and analyzed to determine what progress the class has made so far toward achieving these goals.

Exam specifications: The balance of exam items corresponding to each content area and Bloom's category is tabulated here. This information can be useful for evaluating the validity of the assessment.

Part Performance by content area: The class scores are reported by content area here. This can facilitate the identification of weaknesses in different areas of the material presented.

Performance by cognitive level: The class scores are reported by levels of Bloom's taxonomy here. This can facilitate the identification of weaknesses in comprehension.

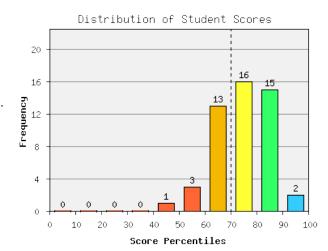
Part G Psychometric assessment of exam items: The performance of each exam item is evaluated in this section. This allows for the identification of weak or poorly functioning questions.

Part H Annotated reading list: A few useful references are included at the end of this report to aid those interested in learning more about the analyses performed by this program.

A) Summary of exam scores back to top

A total of 50 students took this assessment; the distribution of scores (as percentages) is plotted below. The exam had 75 questions worth a maximum of 150 points. The criterion of success (set at 70%) is indicated by the vertical dashed line. Some descriptive statistics for this assignment are also provided in this section.

- The class average was 74.2%, with students scoring 111.4 out of 150 possible points (a letter grade of a C).
- The standard deviation of the mean was 10.9% (16.4 points).
- The standard error of the mean was 1.5% (2.3 points).
- The median class score was 74.7%, or 112 points. (a letter grade of a C).
- This score distribution did not exhibit very much skew (skew = -0.32).
 The Z-score for this skew was -0.96, which is not statistically significant. Given the small magnitude of the skew value, this is not surprising.
- This score distribution did not exhibit very much kurtosis (kurtosis =
 -0.44). The Z-score for this kurtosis was -0.67, which is not statistically significant. Given the small magnitude of the kurtosis value, this is not surprising.



- The Z-values for the skew and kurtosis values from above were used to perform the D'Agostino-Pearson omnibus test. The graph of these scores **does not appear to significantly deviate** from the expected normal distribution ($K^2 = 1.37$, p > 0.05). Since the data seems to follow a Gaussian distribution, we are free to continue to analyze these results with a variety of parametric tests.
- The exam reliability, as measured by the Kuder-Richardson measure of internal consistency, was very good (KR-20 = 0.839).

- This is a terrific score for a classroom exam. The KR-20 score can be improved by increasing the class size, the length of the exam, or the consistency of the exam items.
- The standard error of measurement for this exam was 4.0%, or 6.0 points. This value indicates the degree of uncertainty associated with each observed test score due to the sampling error associated with the test instrument. Lower error values indicate that the observed scores are highly correlated with the students' true latent abilities. Higher error scores indicate a poorer correlation between the observed and true score.

Overall, 33 out of 50 students (66.0%) scored at or above the criterion of success defined for this assessment. The 95 percent confidence interval for the class average was $74.2\% \pm 1.5\%$ (111.4 points ± 2.3 points). This is **significantly better** than the criterion of success - t (49) = 2.7, p < 0.005). The magnitude of the effect size for this score was **small (Cohen's** d = **0.39)**. The class performance measurably exceeded the criterion of success.



BIOL999_Sample_Report.txt

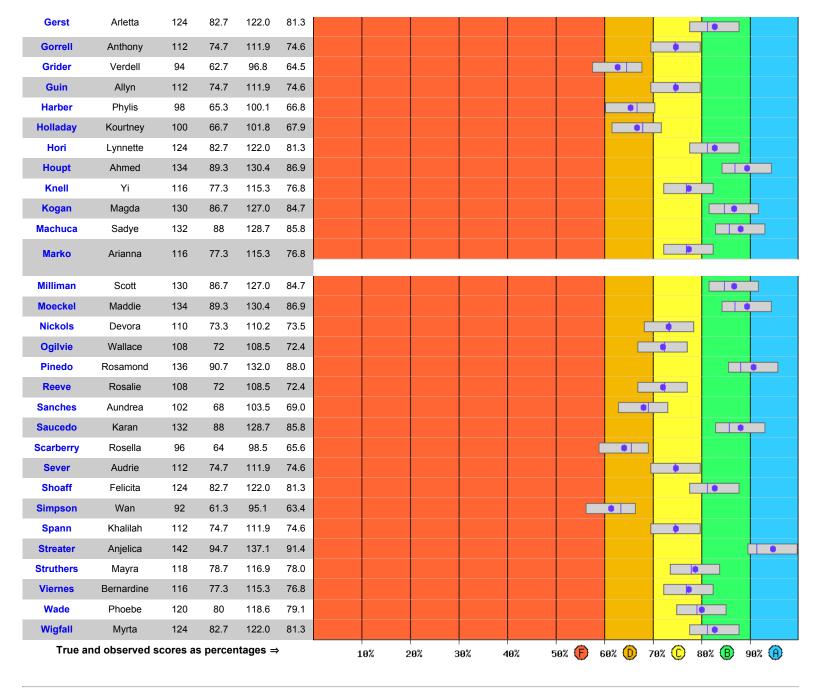
To facilitate additional statistical analysis of the results for this assignment, a simple tab-delimited file has been generated and may be accessed here. The scores for each exam item, exam subscores, and overall grades for each student are included. This file should be compatible with any statistics package that you might wish to use (i.e. Minitab, Microsoft Excel, SPSS, or even R).

B) Listing of student scores (back to top)

A complete listing of all student scores for this assessment are listed below in alphabetical order. In classical test theory, the earned score for an assignment relects both the students' true, or latent, ability and and error component due to sampling. This is formated as: X(observed) = T(true) - E(error). The true scores for students (the thing that we really want to know) can never be directly measured. What we have instead is a single estimate of their ability in the observed score. True score values can be estimated by correcting for the error of measurment for the exam. A 90% confidence interval for the true score of each student has also been calculated and is plotted as a gray rectangle below - if we were able to repeatedly give this test (or one exactly like it) to the students, their true score would fall within the gray box range nine times out of ten. The true score estimates are shown as vertical lines within each rectangle; the observed scores are plotted as filled blue circles.

Individualized reports have been created for each member of the class. You may elect to have these HTML documents emailed automatically to each student as an attachment if you enable the corresponding option in the configuration file. Clicking on the students' last name will link you to the formative feedback report for that student.



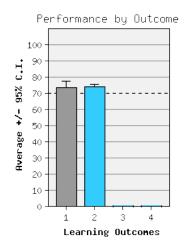


C) Outcome assessment (back to top)

Many or most of the questions in this exam have been mapped onto specific learning outcomes for the course. The class performance on each of these outcomes is plotted to the right, with the criterion of success (70%) indicated as a horizontal dashed line. The average values of each outcome were compared against the criterion of success using a series of two-tailed, one-sample t-tests. The Dunn-Sidak correction for multiple comparisons was used to limit the likelihood of type I errors. Each individual t-test was performed with a smaller α (p = 0.025) in order to keep the family-wise α low (p = 0.049). The sheer size of the sample size keeps the probability of a type II error reasonable under these conditions.

Blue columns denote content areas with performance that was statistically better than the criterion of success.

Red columns indicate content areas with performance that was statistically worse than the criterion of success.



Gray columns show the content areas with performances that were deemed to be not statistically different from the criterion of success.

The overall class results on the materials from each content area are summarized below. You may quickly navigate to a particular section by clicking on the desired column in the graph.

1. **Microbial Diversity** - By the completion of this course, students will be able to compare and contrast different types of microbes (including viruses, bacteria, fungi, and protozoa) with regard to their structure, function, and composition.

This learning outcome was assessed by items worth a total of 30 points (20.0% of the total exam). The class scored 22.1 ± 1.2 of these points (average \pm 95% confidence interval). That gives a subscore for this learning outcome of $73.6\% \pm 4.1\%$. A total of 30 of the students (60.0% of the class) met or exceeded the criterion of success defined for this section. The class average for this material was **not significantly different** from the criterion of success - t (49) = 1.74, p > 0.049. The magnitude of the effect size for this score was **small (Cohen's** d = 0.25). Although the scores exeed the criterion of success, the difference is not very impressive. We are unable to discount the null hypothesis - that the course performance equals the criterion of success. Therefore, we retain the belief that the stated goal for this learning outcome **has been met** by the class.

Most of the class performed well on this learning outcome. However, a sizable number of students had difficulty with this material. The class as a whole appears to have a good grasp of this material.

2. **Microbial Metabolism** - By the completion of this course, students will be able to explain the various strategies employed by microbes with respect to their metabolism and growth in relationship to their structure and composition.

This learning outcome was assessed by items worth a total of 120 points (80.0% of the total exam). The class scored 89.3 ± 3.7 of these points (average $\pm 95\%$ confidence interval). That gives a subscore for this learning outcome of $74.4\% \pm 2.0\%$. A total of 33 of the students (66.0% of the class) met or exceeded the criterion of success defined for this section. The class average for this material was **significantly better** than the criterion of success - t (49) = 4.34, p < 0.049. The magnitude of the effect size for this score was **small** (**Cohen's** d = **0.40**). The class scored measurably better than the criterion of success on this learning outcome. The stated goal for this learning outcome **has been exceeded** by the class.

Most of the class performed well on this learning outcome. However, a sizable number of students had difficulty with this material. The class as a whole appears to have a good grasp of this material.

3. **Microbial Genetics** - By the completion of this course, students will be able to explain the various genetic strategies employed by microbes to express, regulate and share genetic information.

This learning outcome was not assessed by any items in this exam.

4. **Host-Microbe Interactions** - By the completion of this course, students will be able to compare the different interactions possible between host and microbial cells and describe several different molecular strategies employed by specific microbial pathogens.

This learning outcome was not assessed by any items in this exam.

D) Exam specifications (back to top)

Every exam can be view as a composite of several parts (be they the content areas that were covered, different levels of critical thinking, or even just a collection of individual items - the questions themselves). In order to assure exam validity - the assignment actually tests the students on the materials that were covered - an exam blueprint has been generated here. This table is a record of the point distribution in this particular assessment with regard to content areas and levels of Bloom's taxonomy. You will have to decide for yourself is the balance of coverage in the exam matches that of the class presentations. The class' performance on these different dimensions (content and cognitive level) will be analyzed later in this report.

Clicking on one of the column or row headings in this table will link you to the corresponding analysis within this report.

Blueprint of the Point Distribution for This Assessment

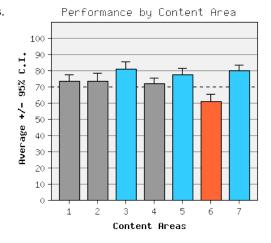
Content Areas	Remembering	Understanding	Applying	Analyzing	Evaluating	Creating	Total
Module A Review	10	10	8	2	0	0	30 20%
Microbial Growth	6	4	8	2	0	0	20 13.3%
Environmental Effects	6	8	4	2	0	0	20 13.3%
Catabolism	6	10	4	0	0	0	20 13.3%
Respiration and Photosynthesis	10	10	9	0	0	0	20 13.3%
Biosynthesis	12	8	0	0	0	0	20 13.3%
Applied Microbiology	14	4	2	0	0	0	20 13.3%
Total	64 42.7%	54 36%	26 17.3%	6 4%	9 9%	9 9%	150 100%

E) Performance by content area (back to top)

This exam evaluated the class' comprehension of materials drawn from 7 content areas. The class performance is plotted to the right, with the criterion of success (70%) indicated as a horizontal dashed line. The average values of each content area were compared against the criterion of success using a series of two-tailed, one-sample t-tests. The Dunn-Sidak correction for multiple comparisons was used to limit the likelihood of type I errors. Each individual t-test was performed with a smaller α (p = 0.005) in order to keep the family-wise α low (p = 0.034). The sheer size of the sample size keeps the probability of a type II error reasonable under these conditions.

Blue columns denote content areas with performance that was statistically better than the criterion of success.

Red columns indicate content areas with performance that was statistically worse than the criterion of success.



Gray columns show the content areas with performances that were deemed to be not statistically different from the criterion of success.

The overall class results on the materials from each content area are summarized below. You may quickly navigate to a particular section by clicking on the desired column in the graph.

1. **Module A Review** - The material in this section is a review of the Microbial Diversity content covered by the first lecture exam. I will be comparing the class' performance on this to the class average on the first unit exam.

The following questions probed the students' comprehension of this topic. The performance of each question is indicated as a superscript (P, poor; M, marginal; G, good; E, excellent). Click on a question number to skip down to the corresponding item analysis graph.

This content area was assessed by items worth a total of 30 points (20.0% of the total exam). The class scored 22.1 \pm 1.2 of these points (average \pm 95% confidence interval). That gives a subscore for this content area of **73.6**% \pm **4.1**%.

A total of 30 of the students (60.0% of the class) met or exceeded the criterion of success defined for this section. The class average for this material was **not significantly different** from the criterion of success - t (49) = 1.74, p > 0.034. The magnitude of the effect size for this score was **small** (**Cohen's** d = **0.25**). Although the scores exeed the criterion of success, the difference is not very impressive.

Most of the class performed well on this content. However, a sizable number of students had difficulty with this material. A bit of review concerning the key facts, concepts, and procedures would probably help these students to improve. The class as a whole appears to have a good grasp of this material.

2. **Microbial Growth** - This material covered the basic mechanism of microbial growth and the ways in which growth may be measured. Some of the factors affecting growth were covered along with applications to environmental and medical microbiology.

The following questions probed the students' comprehension of this topic. The performance of each question is indicated as a superscript (P, poor; M, marginal; G, good; E, excellent). Click on a question number to skip down to the corresponding item analysis graph.

```
16<sup>G</sup>, 17<sup>M</sup>, 18<sup>G</sup>, 19<sup>P</sup>, 20<sup>E</sup>, 21<sup>P</sup>, 22<sup>G</sup>, 23<sup>E</sup>, 24<sup>M</sup>, and 25<sup>E</sup>.
```

This content area was assessed by items worth a total of 20 points (13.3% of the total exam). The class scored 14.8 \pm 1.0 of these points (average \pm 95% confidence interval). That gives a subscore for this content area of **73.8%** \pm **5.1%**.

A total of 31 of the students (62.0% of the class) met or exceeded the criterion of success defined for this section. The class average for this material was **not significantly different** from the criterion of success - t (49) = 1.47, p > 0.034. The magnitude of the effect size for this score was **small** (**Cohen's** d = **0.21**). Although the scores exceed the criterion of success, the difference is not very impressive.

Most of the class performed well on this content. However, a sizable number of students had difficulty with this material. A bit of review concerning the key facts, concepts, and procedures would probably help these students to improve. The class as a whole appears to have a good grasp of this material.

3. **Environmental Effects** - The effect of environmental factors such as temperature, pH, and oxygen availability were discussed in this unit. Antimicrobial compounds (antiseptics and disinfectants) were also covered in this lecture.

The following questions probed the students' comprehension of this topic. The performance of each question is indicated as a superscript (P, poor; M, marginal; G, good; E, excellent). Click on a question number to skip down to the corresponding item analysis graph.

```
26<sup>G</sup>, 27<sup>M</sup>, 28<sup>P</sup>, 29<sup>E</sup>, 30<sup>G</sup>, 31<sup>G</sup>, 32<sup>G</sup>, 33<sup>E</sup>, 34<sup>G</sup>, and 35<sup>E</sup>.
```

This content area was assessed by items worth a total of 20 points (13.3% of the total exam). The class scored 16.2 \pm 0.9 of these points (average \pm 95% confidence interval). That gives a subscore for this content area of 81.2% \pm 4.7%.

A total of 44 of the students (88.0% of the class) met or exceeded the criterion of success defined for this section. The class average for this material was **significantly better** than the criterion of success - t (49) = 4.70, p < 0.034. The magnitude of the effect size for this score was **medium** (**Cohen's** d = **0.66**). The class performed meaningfully better than the criterion of success on these materials.

The vast majority of the class performed well on this content. A little reinforcement of these facts, concepts, and procedures is probably all that is required. The class appears to have mastered this material.

4. **Catabolism** - This material included a refresher in the overall balance between anabolism and catabolism in cells. A few of the core intermediary metabolic pathways important to microorganisms were also detailed.

The following questions probed the students' comprehension of this topic. The performance of each question is indicated as a superscript (P, poor; M, marginal; G, good; E, excellent). Click on a question number to skip down to the corresponding item analysis graph.

```
36<sup>M</sup>, 37<sup>M</sup>, 38<sup>G</sup>, 39<sup>G</sup>, 40<sup>M</sup>, 41<sup>G</sup>, 42<sup>P</sup>, 43<sup>G</sup>, 44<sup>P</sup>, and 45<sup>P</sup>,
```

This content area was assessed by items worth a total of 20 points (13.3% of the total exam). The class scored 14.4 \pm 0.8 of these points (average \pm 95% confidence interval). That gives a subscore for this content area of **72.0%** \pm **3.8%**.

A total of 36 of the students (72.0% of the class) met or exceeded the criterion of success defined for this section. The class average for this material was **not significantly different** from the criterion of success - t (49) = 1.04, p > 0.034. The magnitude of the effect size for this score was **tiny** (**Cohen's** d = **0.15**). The class average is essentially indistingishable from the criterion of success. The class performance was adequate.

Most of the class performed well on this content. However, a sizable number of students had difficulty with this material. A bit of review concerning the key facts, concepts, and procedures would probably help these students to improve. The class as a whole appears to have a good grasp of this material.

5. Respiration and Photosynthesis - The source and fate of electrons in the various metabolic schemes used by microbes were

described in this unit. Oxygenic photosynthesis was also compared and contrasted with other phototrophic lifestyles.

The following questions probed the students' comprehension of this topic. The performance of each question is indicated as a superscript (P, poor; M, marginal; G, good; E, excellent). Click on a question number to skip down to the corresponding item analysis graph.

```
46<sup>G</sup>, 47<sup>M</sup>, 48<sup>G</sup>, 49<sup>E</sup>, 50<sup>P</sup>, 51<sup>P</sup>, 52<sup>G</sup>, 53<sup>M</sup>, 54<sup>G</sup>, and 55<sup>M</sup>,
```

This content area was assessed by items worth a total of 20 points (13.3% of the total exam). The class scored 15.6 ± 0.8 of these points (average ± 95% confidence interval). That gives a subscore for this content area of 77.8% ± 4.2%.

A total of 41 of the students (82.0% of the class) met or exceeded the criterion of success defined for this section. The class average for this material was significantly better than the criterion of success - t (49) = 3.61, p < 0.034. The magnitude of the effect size for this score was medium (Cohen's d = 0.51). The class performed meaningfully better than the criterion of success on these materials.

The vast majority of the class performed well on this content. A little reinforcement of these facts, concepts, and procedures is probably all that is required. The class appears to have mastered this material.

6. Biosynthesis - Pathways essential for the assimilation of inorganic carbon and nitrogen were compared in this unit. Tie-ins were also made to the synthesis of important macromolecules (like proteins, lipids, nucleic acids, and peptidoglycan).

The following questions probed the students' comprehension of this topic. The performance of each question is indicated as a superscript (P, poor; M, marginal; G, good; E, excellent). Click on a question number to skip down to the corresponding item analysis graph.

```
56<sup>G</sup>, 57<sup>M</sup>, 58<sup>E</sup>, 59<sup>M</sup>, 60<sup>G</sup>, 61<sup>G</sup>, 62<sup>G</sup>, 63<sup>P</sup>, 64<sup>P</sup>, and 65<sup>P</sup>,
```

This content area was assessed by items worth a total of 20 points (13.3% of the total exam). The class scored 12.2 ± 1.0 of these points (average ± 95% confidence interval). That gives a subscore for this content area of 61.2% ± 5.0%.

A total of 20 of the students (40.0% of the class) met or exceeded the criterion of success defined for this section. The class average for this material was significantly worse than the criterion of success - t (49) = -3.45, p < 0.034. The magnitude of the effect size for this score was small (Cohen's d = -0.49). The scores were measurably lower than the criterion of success for this content area.

Less than half of the class met the performance standard for this material. A large number of students are apparently struggling with facts, concepts, and procedures covered in this section. This material should be reviewed online and probably reassessed with some sort of quiz to monitor improvements in understanding.

7. Applied Microbiology - This unit is a brief overview of environmental microbiology. Terrestrial and aquatic communities were compared and contrasted. The importance of microbes in the cycling of biologically important elements was emphasized.

The following questions probed the students' comprehension of this topic. The performance of each question is indicated as a superscript (P, poor; M, marginal; G, good; E, excellent). Click on a question number to skip down to the corresponding item analysis graph.

```
66<sup>P</sup>, 67<sup>M</sup>, 68<sup>G</sup>, 69<sup>P</sup>, 70<sup>P</sup>, 71<sup>G</sup>, 72<sup>G</sup>, 73<sup>G</sup>, 74<sup>M</sup>, and 75<sup>M</sup>,
```

This content area was assessed by items worth a total of 20 points (13.3% of the total exam). The class scored 16.1 ± 0.7 of these points (average ± 95% confidence interval). That gives a subscore for this content area of 80.4% ± 3.7%.

A total of 45 of the students (90.0% of the class) met or exceeded the criterion of success defined for this section. The class average for this material was significantly better than the criterion of success - t (49) = 5.55, p < 0.034. The magnitude of the effect size for this score was medium (Cohen's d = 0.79). The class performed meaningfully better than the criterion of success on these materials.

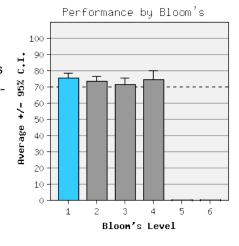
The vast majority of the class performed well on this content. A little reinforcement of these facts, concepts, and procedures is probably all that is required. The class appears to have mastered this material.



This exam also evaluated the class' ability to reason at different cognitive levels. For the purposes of this report, each question was mapped to one of the six levels of the modified Bloom's taxonomy of Anderson and Krathwohl. The class performance is plotted to the right, with the criterion of success (70%) indicated as a horizontal dashed line. The average values of scores for each level were compared against the criterion of success using a series of two-tailed, one-sample t-tests. The Dunn-Sidak correction for multiple comparisons was used to limit the likelihood of type I errors. Each individual t-test was performed with a smaller α (p = 0.01) in order to keep the family-wise α low (p = 0.039). The sheer size of the sample size keeps the probability of a type II error reasonable under these conditions.

Blue columns denote cognitive levels with performance that was statistically better than the criterion of success.

Red columns indicate cognitive levels with performance that was statistically worse than the criterion of success.



Gray columns show the cognitive levels with performances that were deemed to be not statistically different from the criterion of success.

The overall class results on the materials from each content area are summarized below. You may quickly navigate to a particular section by clicking on the desired column in the graph.

Remembering - Students were expected to respond to these questions by simply recalling memorized information. Although
these items may involve remembering a wide range of material from specific facts to complete theories, all that is required is
the bringing to mind of the appropriate information.

The following questions probed the students' ability to work using the course materials at this cognitive level. The performance of each question is indicated as a superscript (P, poor; M, marginal; G, good; E, excellent). Click on a question number to skip down to the corresponding item analysis graph.

1^M, 5^G, 6^E, 10^G, 14^P, 16^G, 17^M, 24^M, 28^P, 33^E, 35^E, 43^G, 44^P, 45^P, 49^E, 52^G, 53^M, 54^G, 55^M, 57^M, 59^M, 60^G, 62^G, 63^P, 65^P, 66^P, 67^M, 69^P, 70^P, 71^G, 72^G, and 74^M.

This content area was assessed by items worth a total of 64 points (42.7% of the total exam). The class scored 48.3 ± 2.0 of these points (average $\pm 95\%$ confidence interval). That gives a subscore for this content area of $75.5\% \pm 3.2\%$.

A total of 35 of the students (70.0% of the class) met or exceeded the criterion of success defined for this section. The class average for this material was **significantly better** than the criterion of success - t (49) = 3.42, p < 0.039. The magnitude of the effect size for this score was **small** (**Cohen's** d = **0.48**). The class scored measurably better than the criterion of success on these materials.

Most of the class performed well on this content. However, a sizable number of students had difficulty with this material. A bit of review concerning the key facts, concepts, and procedures would probably help these students to improve. The class as a whole appears to have a good grasp of this material.

2. **Understanding** - Students were expected to demonstrate an ability to grasp the meaning of material to answer these questions. The items may require students to identify examples of a category, classify specific items into groups, summarize concepts, or compare and contrast material. This level goes one step beyond the simply remembering facts.

The following questions probed the students' ability to work using the course materials at this cognitive level. The performance of each question is indicated as a superscript (P, poor; M, marginal; G, good; E, excellent). Click on a question number to skip down to the corresponding item analysis graph.

2^E, 3^P, 4^E, 9^P, 11^M, 18^G, 23^E, 27^M, 29^E, 32^G, 34^G, 36^M, 38^G, 39^G, 40^M, 42^P, 46^G, 47^M, 48^G, 50^P, 51^P, 56^G, 58^E, 61^G, 64^P, 68^G, and 73^G.

This content area was assessed by items worth a total of 54 points (36.0% of the total exam). The class scored 39.9 \pm 1.9 of these points (average \pm 95% confidence interval). That gives a subscore for this content area of **73.9%** \pm **3.4%**.

A total of 35 of the students (70.0% of the class) met or exceeded the criterion of success defined for this section. The class average for this material was **not significantly different** from the criterion of success - t (49) = 2.19, p > 0.039. The magnitude of the effect size for this score was **small** (**Cohen's** d = **0.31**). Although the scores exceed the criterion of success,

the difference is not very impressive.

Most of the class performed well on this content. However, a sizable number of students had difficulty with this material. A bit of review concerning the key facts, concepts, and procedures would probably help these students to improve. The class as a whole appears to have a good grasp of this material.

3. **Applying** - Students were expected to exhibit an ability to use the learned material in new and concrete situations material to answer these questions. These items may require the students to sequence or arrange concepts or procedures solve problems using simple calculations, or apply rules or methods to respond to a specific situation. Questions written at this level require a higher level of cognitive ability than those under understanding.

The following questions probed the students' ability to work using the course materials at this cognitive level. The performance of each question is indicated as a superscript (P, poor; M, marginal; G, good; E, excellent). Click on a question number to skip down to the corresponding item analysis graph.

This content area was assessed by items worth a total of 26 points (17.3% of the total exam). The class scored 18.7 \pm 1.1 of these points (average \pm 95% confidence interval). That gives a subscore for this content area of **71.8%** \pm **4.1%**.

A total of 23 of the students (46.0% of the class) met or exceeded the criterion of success defined for this section. The class average for this material was **not significantly different** from the criterion of success - t (49) = 0.88, p > 0.039. The magnitude of the effect size for this score was **tiny** (**Cohen's** d = **0.12**). The class average is essentially indistingishable from the criterion of success. The class performance was adequate.

Less than half of the class met the performance standard for this material. A large number of students are apparently struggling with facts, concepts, and procedures covered in this section. This material should be reviewed online and probably reassessed with some sort of quiz to monitor improvements in understanding.

4. Analyzing - Students were expected to show an ability to use critical thinking skills in order to break down a problems into its component parts. These items may require the students to recognize unstated assumptions and logical fallacies in reasoning, distinguish between facts and inferences, evaluate the relevancy of data, and analyze the organizational structure of the problem. Questions written at this level require the students to apprehend both the content and the structural form of the material presented in the problem.

The following questions probed the students' ability to work using the course materials at this cognitive level. The performance of each question is indicated as a superscript (P, poor; M, marginal; G, good; E, excellent). Click on a question number to skip down to the corresponding item analysis graph.

This content area was assessed by items worth a total of 6 points (4.0% of the total exam). The class scored 4.5 ± 0.3 of these points (average $\pm 95\%$ confidence interval). That gives a subscore for this content area of $74.7\% \pm 5.8\%$.

A total of 17 of the students (34.0% of the class) met or exceeded the criterion of success defined for this section. The class average for this material was **not significantly different** from the criterion of success - t (49) = 1.58, p > 0.039. The magnitude of the effect size for this score was **small (Cohen's** d = **0.22**). Although the scores exeed the criterion of success, the difference is not very impressive.

Only some of the class met the performance standard for this material. A large number of students are apparently struggling with facts, concepts, and procedures covered in this section. This material should be reviewed online and probably reassessed with some sort of quiz to monitor improvements in understanding.

- 5. **Evaluating** Students were expected to display the ability to judge the appropriateness of a particular action or potential solution for a given purpose. Their conclusions would be based on definite criteria, which may be internal (organization) or external (relevance to the purpose). The student may determine these criteria or they might be provided in the prompt. *This level of cognitive activity was not assessed by this exam.*
- 6. **Creating** Students were expected to put several independent parts together to form a new whole. This may involve the production of a unique paper or talk, a research proposal, or a set of abstract relations (like a concept map). *This level of cognitive activity was not assessed by this exam.*

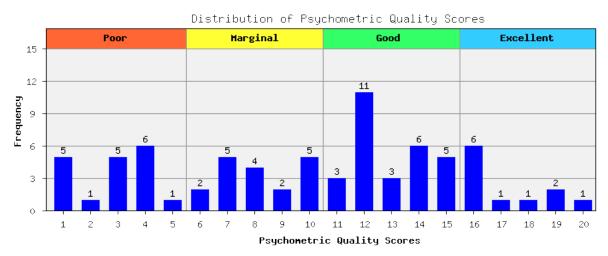
G) Psychometric assessment of exam items back to top



This portion of the report summarizes how well each item performed in this assessment. In classical test theory, the quality of an item's function is affected by three factors: facility, discrimination, and distracter effectiveness. Facility (or difficulty) represents the fraction of the class that correctly answered the question. Very high (easy) and very low (hard) scores are generally undesirable. Since most of the students either correctly or incorrectly respond in these situations, the questions do not allow us to differentiate between students very well. Discrimination is a measure of how well the item score reflects or predicts the overall exam score. In this report, the point biserial correlation is calculated. Higher scores indicate a better relationship between the class performance on the item and the overall exam. Discrimination is the most important factor to consider when evaluating an item's performance. The last factor affecting item performance is the quality of each distracter (incorrect option). Good distracters should be plausible responses (chosen by at least 3% of the class). However, they must not be so ambiguous as to outperform the actual key. In addition, good distracters ought to be more appealing to low-proficiency students (quintiles 1 and 2) than to high-proficiency students (quintiles 4 and 5). As a result, their quintile plots should exhibit a strong negative slope.

Psychometric Score Factors												
Facility			Discrimination			Distractors (X3)						
Description	Range	Pts	Description	Range	Pts	Description	Range	Pts				
Very easy	p ≥ 0.90	2	Excellent	Rpb ≥ 0.4	7	Active	n ≥ 3%	1				
Easy	0.74 ≤ p < 0.9	3	Good	0.3 ≤ Rpb < 0.4	4	Dead	n < 3%	0				
Moderate	0.52 ≤ p < 0.74	4	Adequate	0.2 ≤ Rpb < 0.3	2	Defective	n > key	-2				
Difficult	0.40 ≤ p < 0.52	2	Weak	0.1 ≤ Rpb < 0.2	1	Strong effect	m ≤ -4.0	2				
Very difficult	p < 0.40	0	Poor	Rpb < 0.1	0	Medium effect	-4.0 < m ≤ -2.0	1				
						Weak effect	-2.0 < m ≤ 0.5	0				
						Defective	m > 0.5	-2				

As a means of quantitating item quality, I have created a unique 20-point pysychometric rating scale. Each question on the exam is given a quality score (based upon the item's difficulty, discrimination, and distractor performance). The parameters employed and their relative values are set out in the table above. Items that are rarely selected are identified with a caution (orange). Defective distractors either due to out-competing the key or giving a positive slope (poorer students excel more than better students) are identified with warnings (red). Summing all three of these factors produces a quality score for each question. A histogram of the quality scores for this assessment is presented below. The score range has been divided into four categories - poor, marginal, good, and excellent performance.



Excellent items (16 to 20 points) – These items discriminate very well between students of differing abilities. The item's key and most of its distracters are performing as expected. A few minor improvements may be possible at the lower end of this range.

Good items (11 to 15 points) – These items also discriminate between students of differing abilities and are contributing to the overall reliability of the assessment. The lower scores are often due to sub-optimal facility scores and one or two poorly functioning distracters. A bit of editing may improve the performance of these items.

Marginal items (6 to 10 points) - These items still discriminate between students of differing abilities, but not very well. Facility scores tend toward the extremes (very easy or very hard) and the distracter performance is usually not very robust. These questions should be seriously evaluated and may require major editing or rethinking to improve their performance on future exams.

Poor items (1 to 5 points) – These items are not functioning as intended and they are not adding to the reliability of the overall assessment. Such questions may need major edits or possibly be eliminated altogether. A few unsatisfactory items may just be *really* easy. If such items exist and are deemed to be important for the assessment, they can be retained. Unsatisfactory items should probably not comprise more than around 5 to 10 percent of all assessment items. Too many poorly-functioning items may reduce the reliability of the exam.

1.0

The relative performance of each question can also be visualized by plotting the item facility versus its discrimination as shown to the right. The color of each point on this plot corresponds to a psychometric quality category described above. The plot area has been color-coded into a "target" for best item performance. The light blue "bullseye" indicates an area with moderate facility and high discrimination. This is the location of near-optimal question performance. The surrounding green area contains strong (but not optimal) items. The yellow area contains marginal items, while the red area contains poor items. The goal is to have most of the exam items plot in the blue and green areas. Poor items may occur for many reasons. Some items are just too easy and give poor discrimination. Others may give poor discrimination because they are too difficult (misleading or ambiguous). Quality color mismatches help to reveal problematic distracters. For instance, a green (good) point in the yellow (marginal band) indicates that the question is performing better than expected - its distracters must be working pretty well. Conversely, a green (good) point in the blue (excellent) bullseye indicates that the question is under-performing. Its distractors are probably pulling the item quality score down.

The average item facility (the average exam score - since the exam score is actually just a composite of the individual item scores) and item discrimination are also plotted as dashed lines. Think of this as

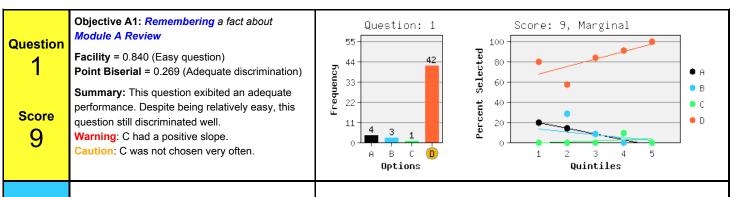
Click on a point to go Excellent questions Quality to that question. - Good questions 0.9 Marginal questions Poor questions 0.8 0.7 biserial) 0.6 **O**23 **0**6 835 **•**60 (point 0.5 **Q**20 **Ф**33 **\$71** 0.4Discrimination **ု_{နှာ့**န္တန္တ **့**53} **01630**32 **0**5 0.3 418 mean=0.27 --85 **Ģ**75 •37 31 •15⁷⁴ 040 07 067 0.2 **O**11 69 0.1 **\$**59 4442 64 **0**65 0.0 **©**14 -0.1-0.2 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 0.0 1.0 Facility (fraction correct)

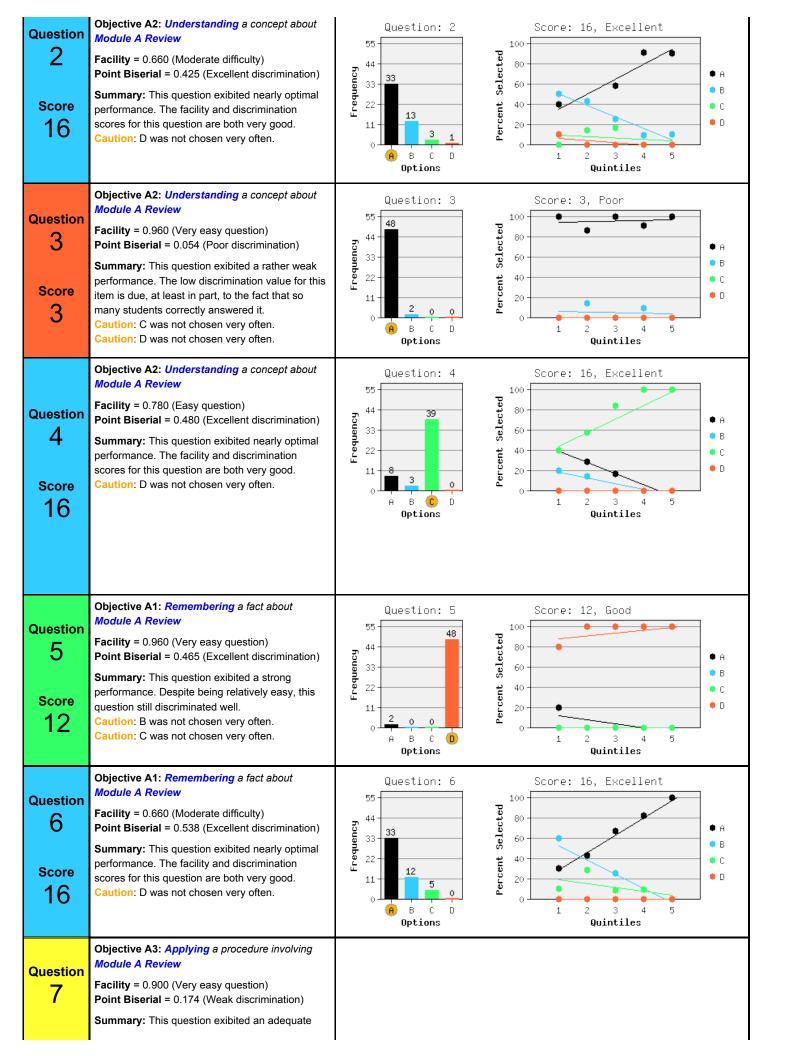
Plot of Item Performance

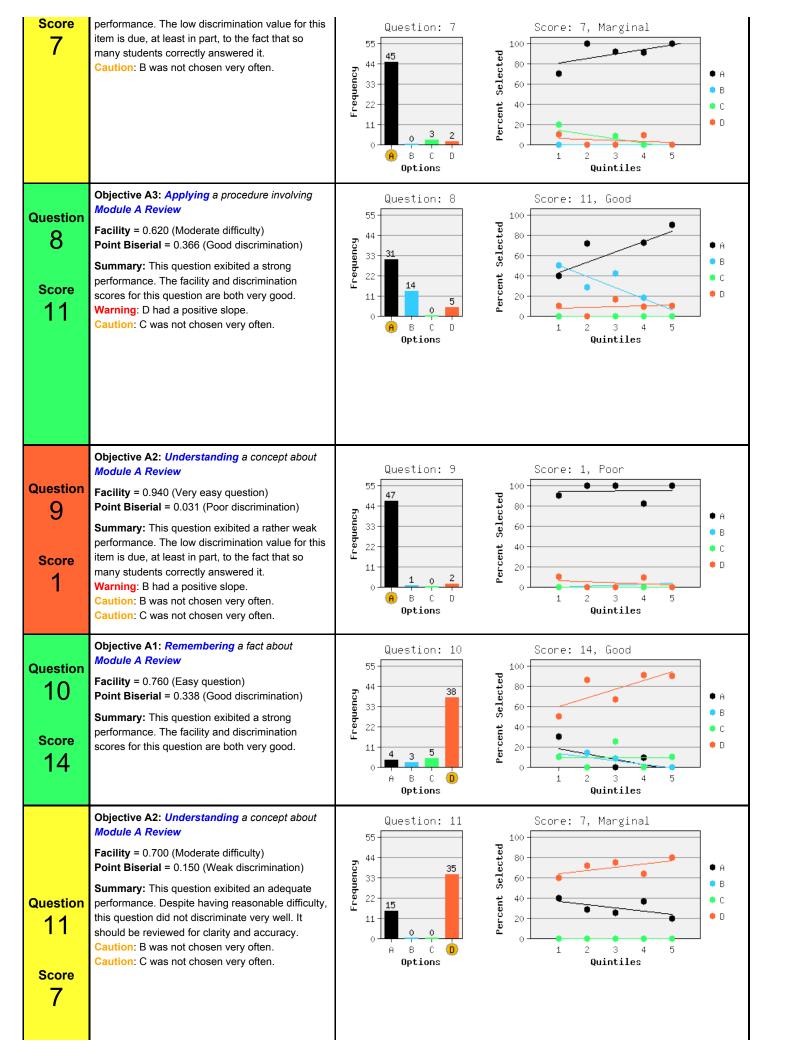
a rifle cross-hairs. You are aiming for the bullseye; this plot will let you rapidly assess how far you are from the mark.

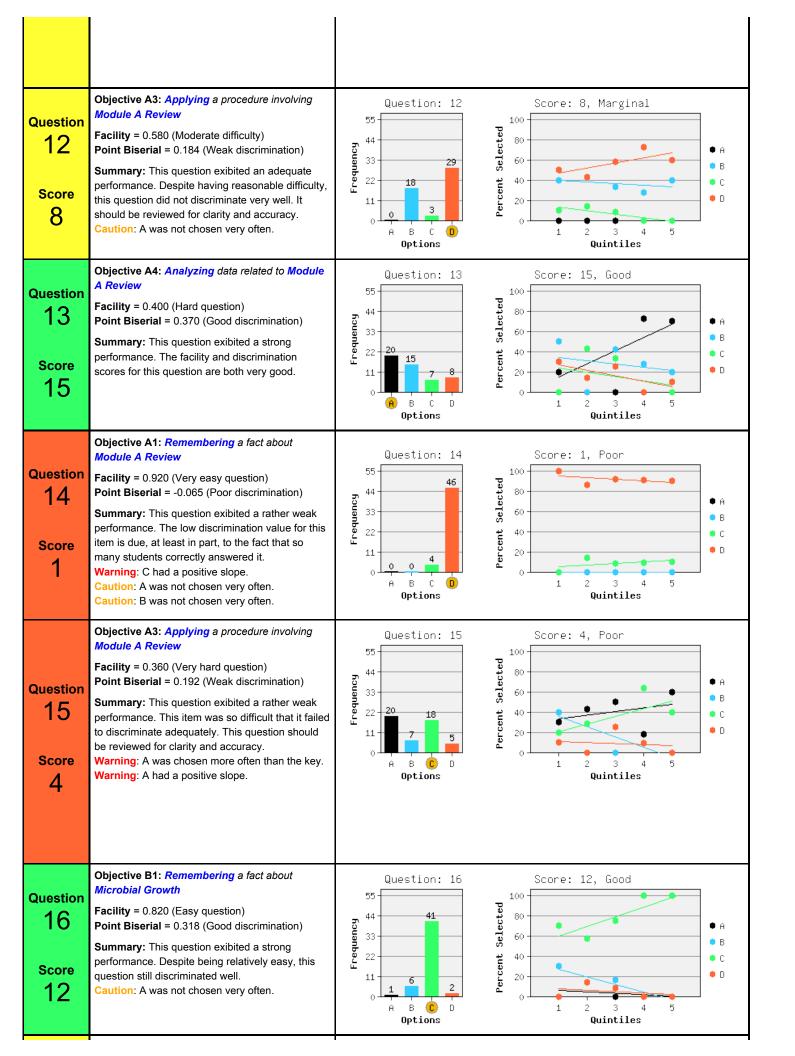
You may quickly navigate to the item analysis for each question by clicking on the points plotted in this graph. (closely spaced points may make this difficult at times).

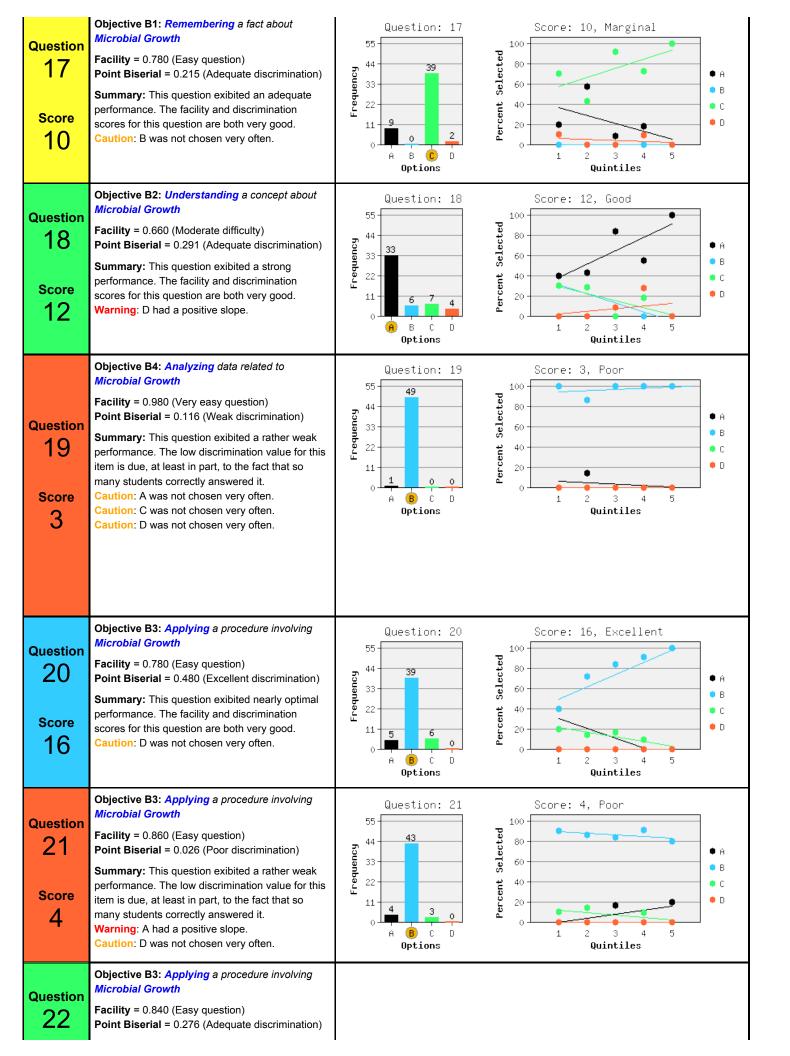
Each item of the exam is scored, color-coded to match the four levels of psychometric performance described above, and analyzed below. The facility and discrimination for each question is also given. To facilitate the evaluation of the exam questions, a brief summary of their performance is given. Warnings and cautions are provided to direct rational efforts to improve the performance of the exam for future classes. The left panel of the graph contains a color-coded histogram of the option frequencies for each question. The correct response is indicated with an orange circle. The frequency that each option was selected by different quintiles of the class is indicated in the right panel. Quintile 1 is the lowest 20% of all the exam scores, and quintile 5 is the highest 20% of the scores. The correct response ought to be selected more often by higher-scoring students and less often by those that performed poorly (that is, the plots should have a positive slope). The distractors should show the opposite trend (negative slopes). The color scheme for the options is the same in both panels.

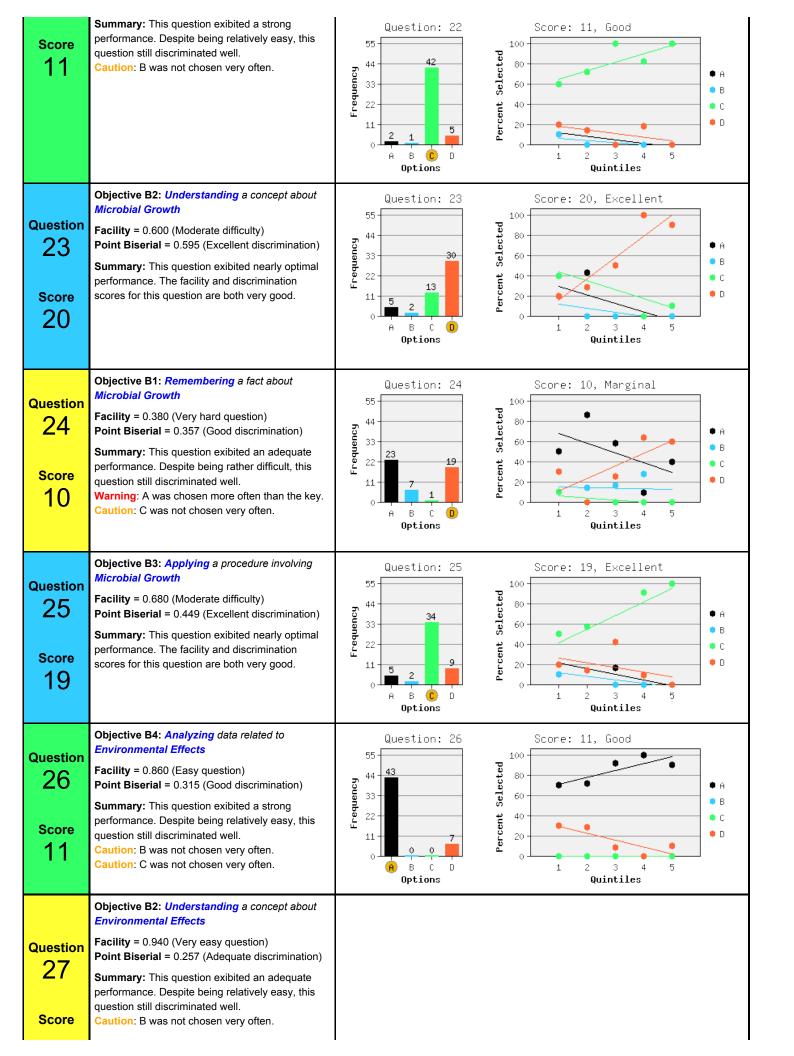


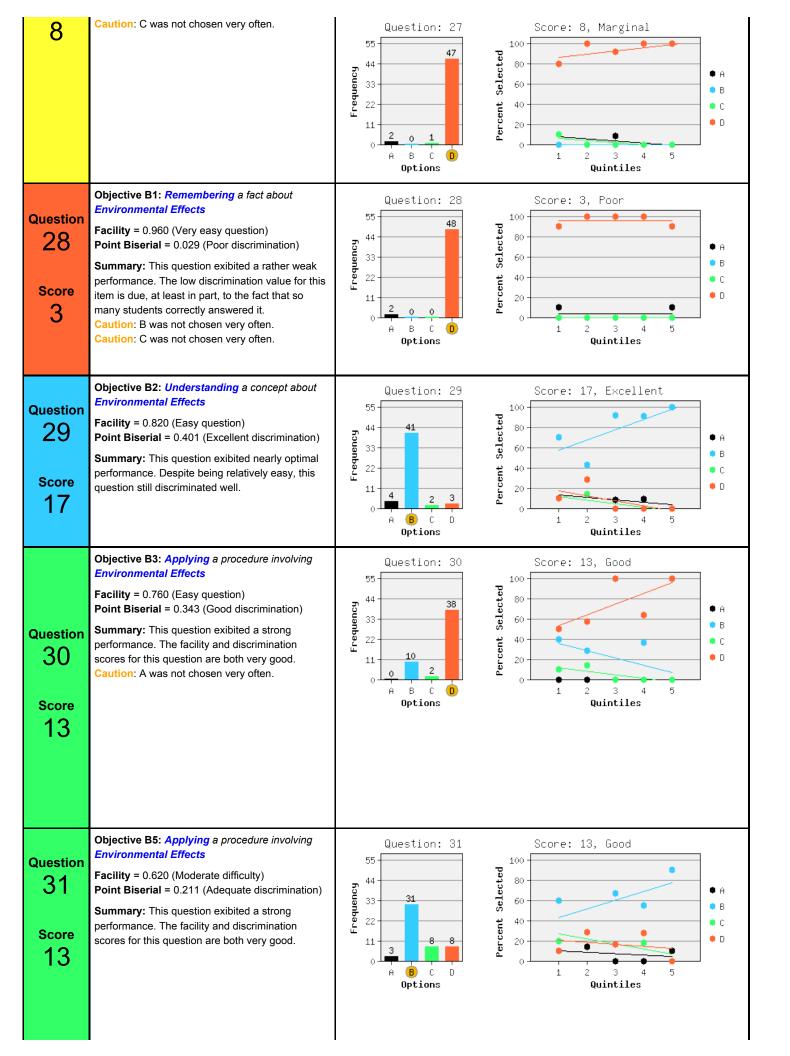


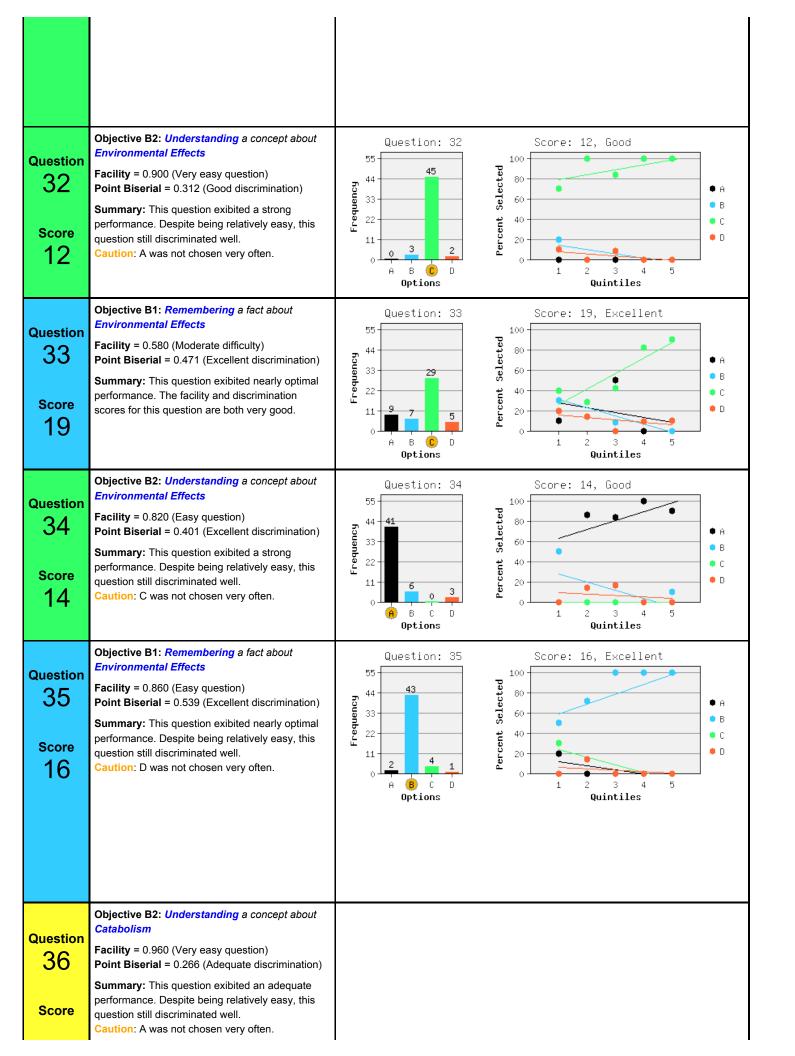


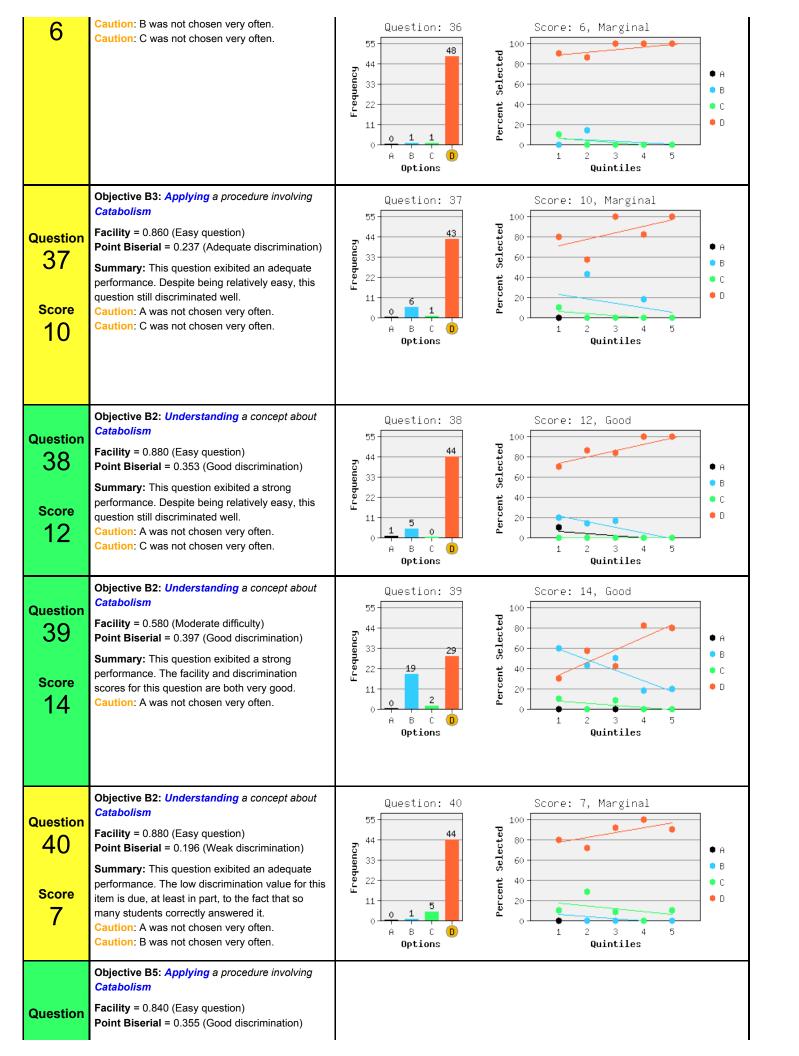


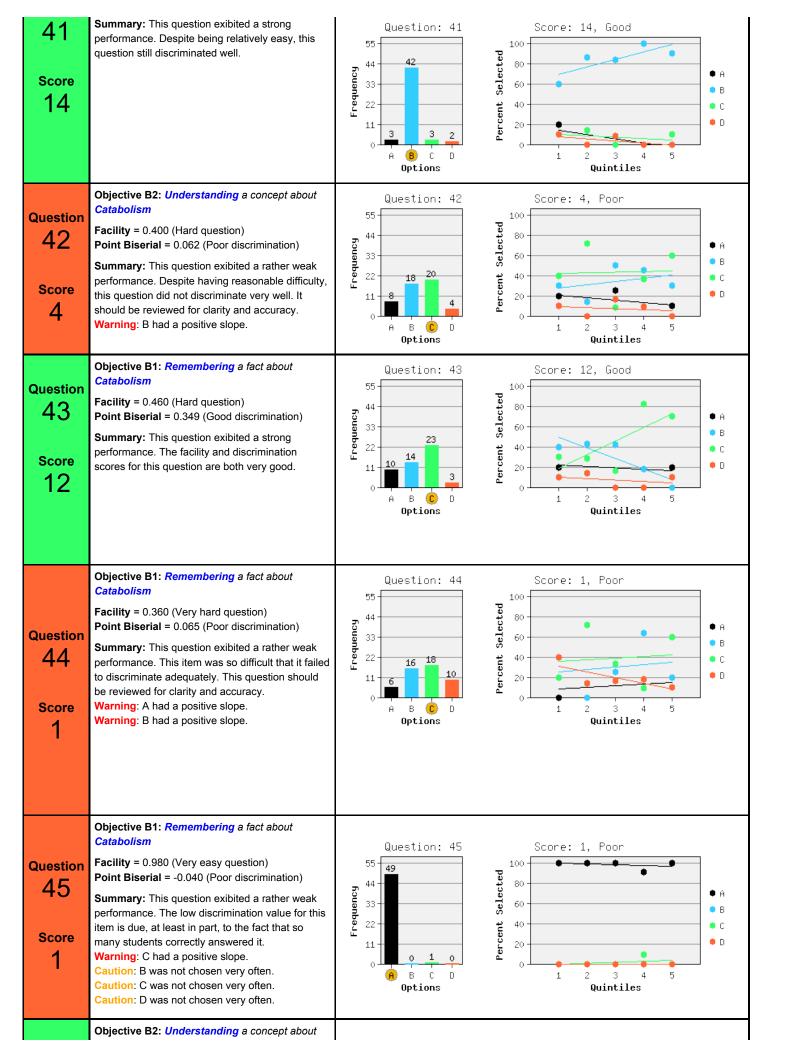


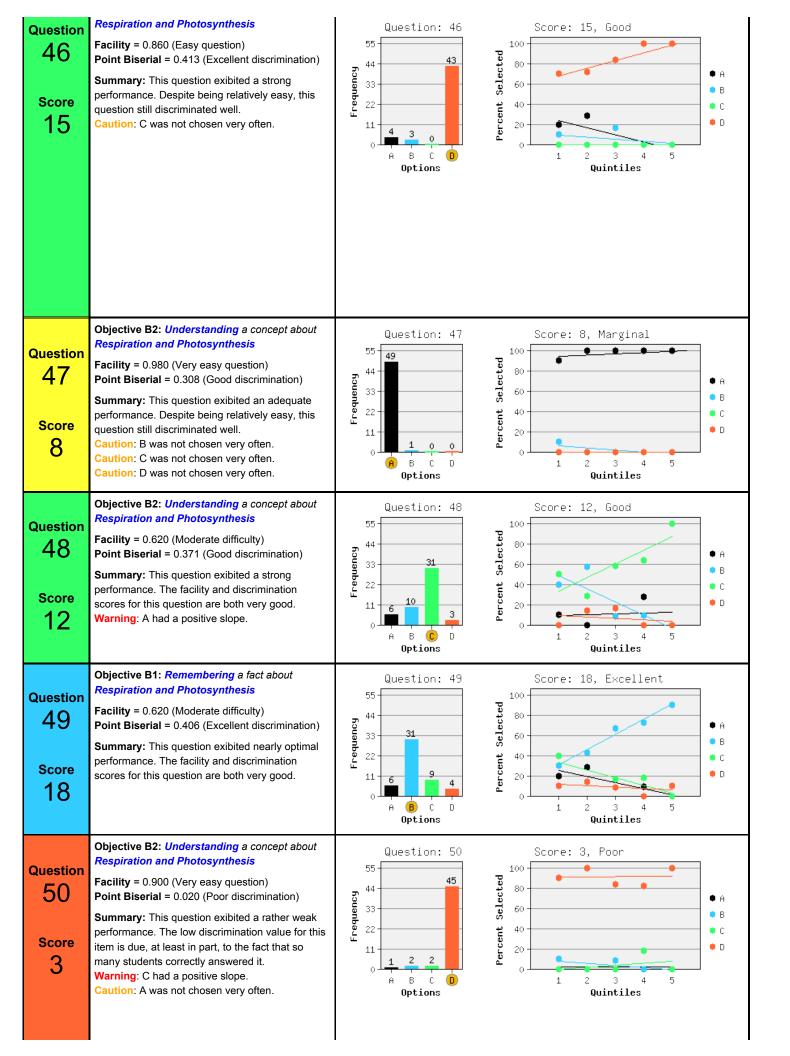


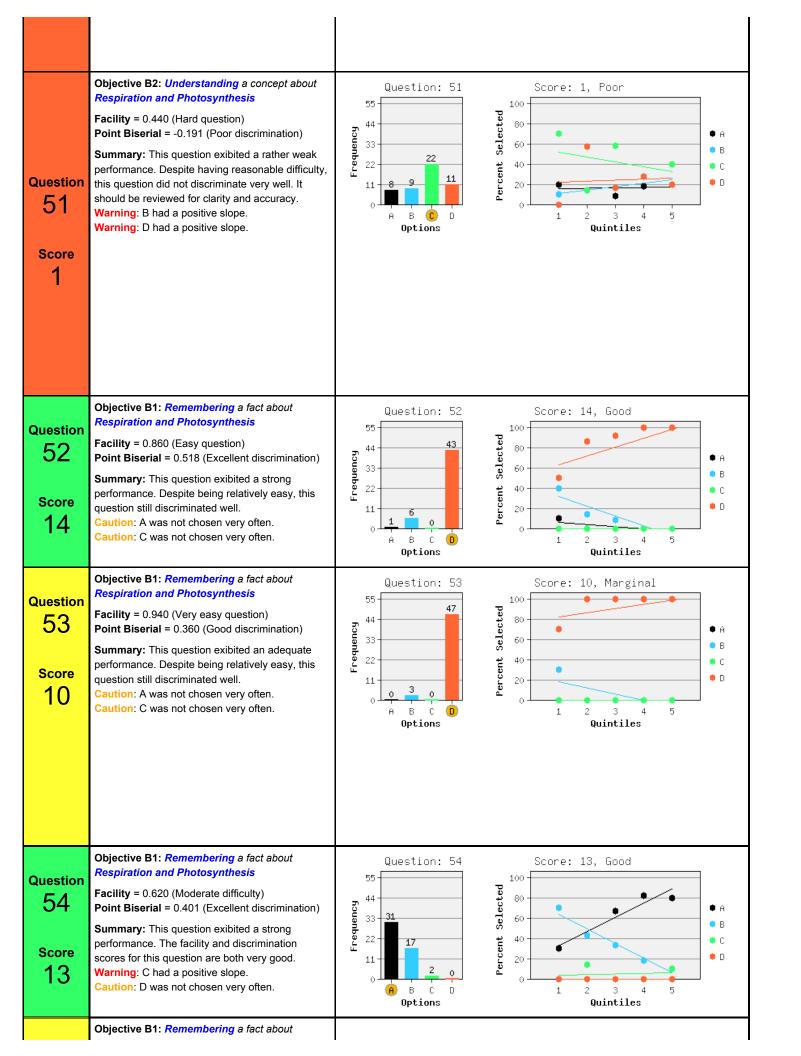


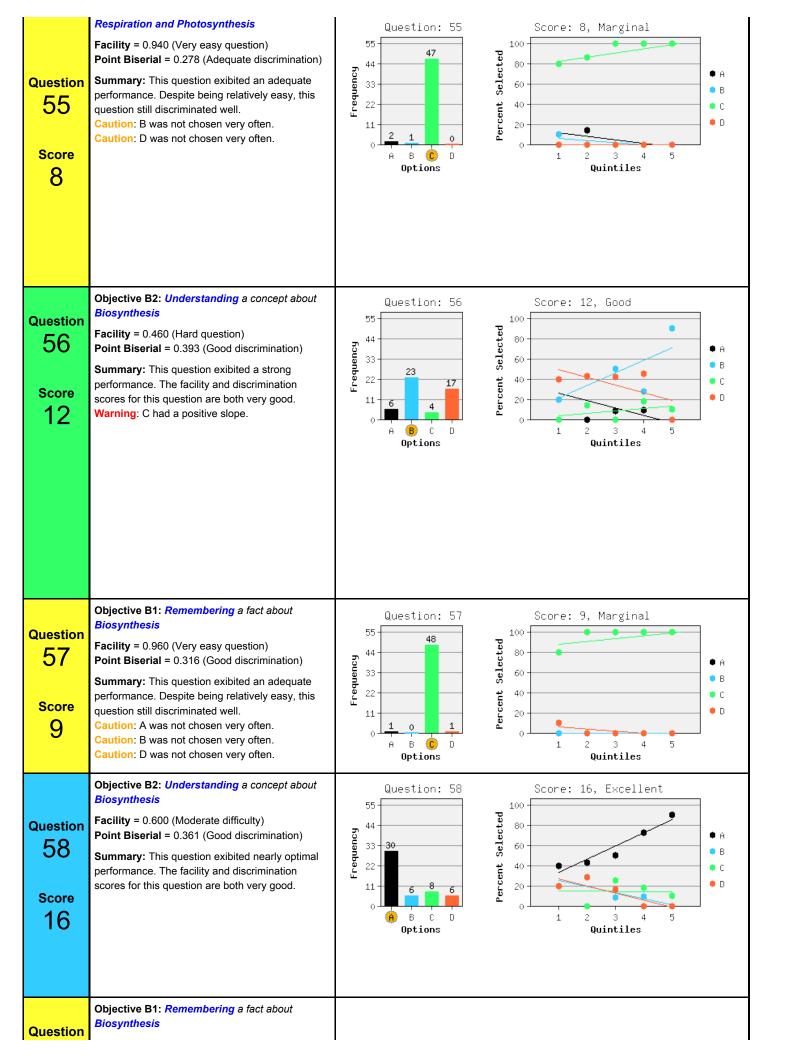


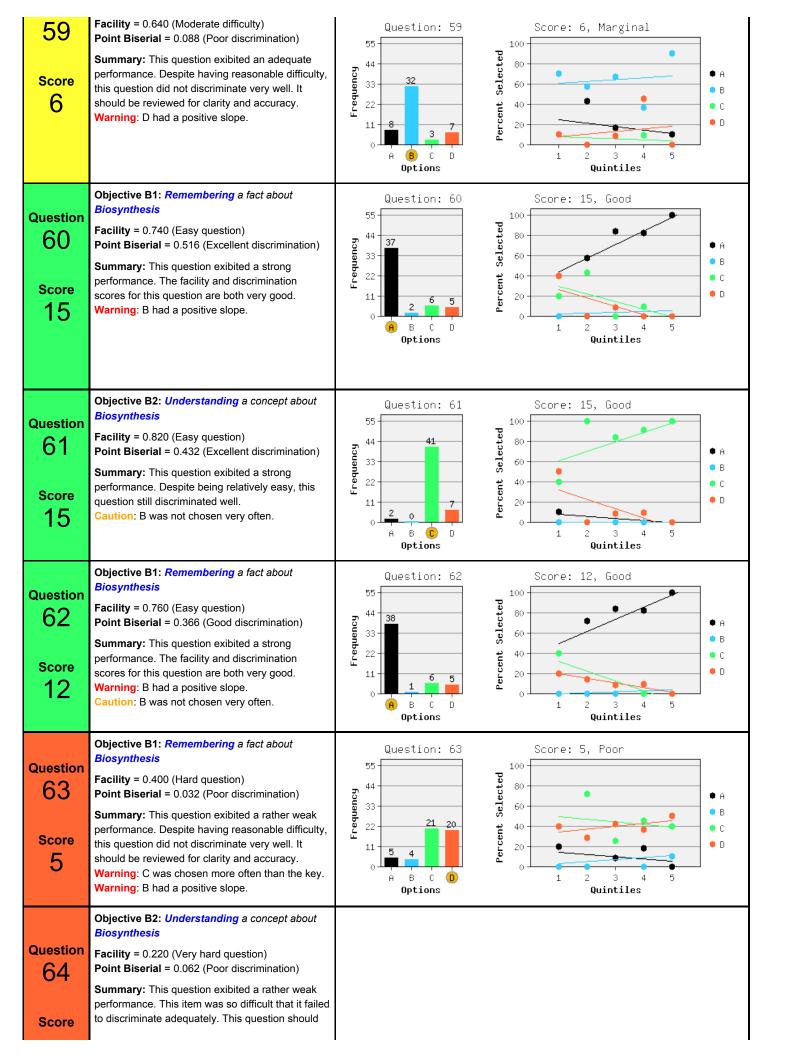


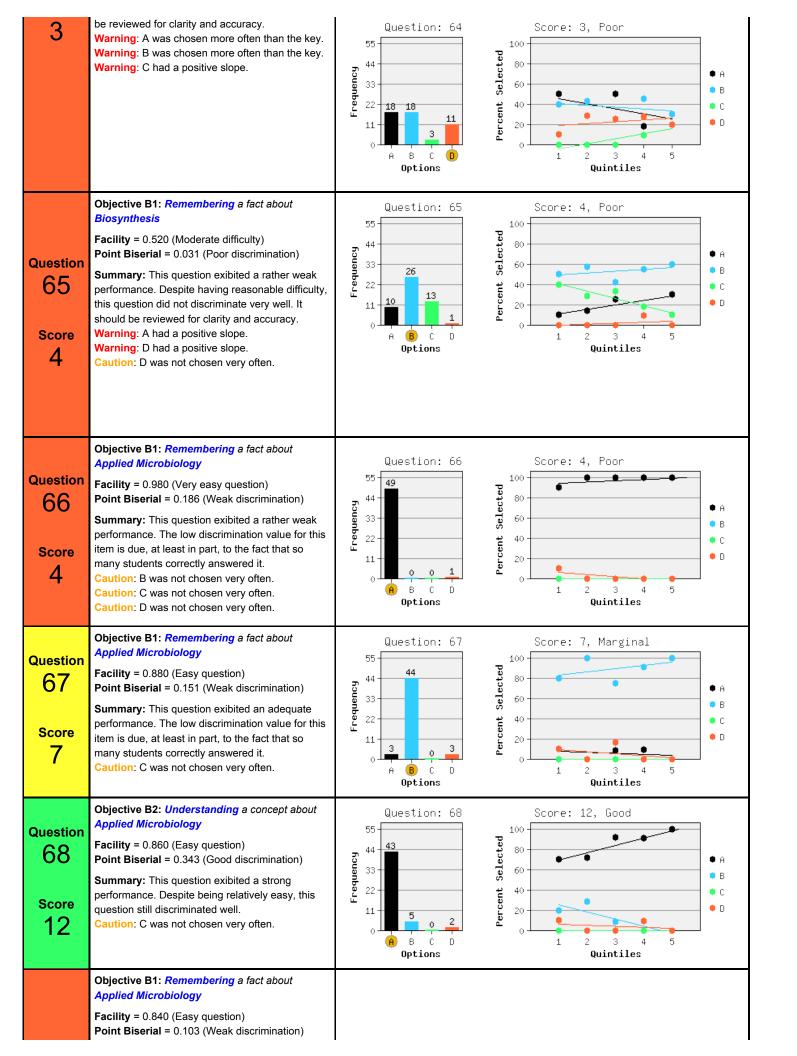


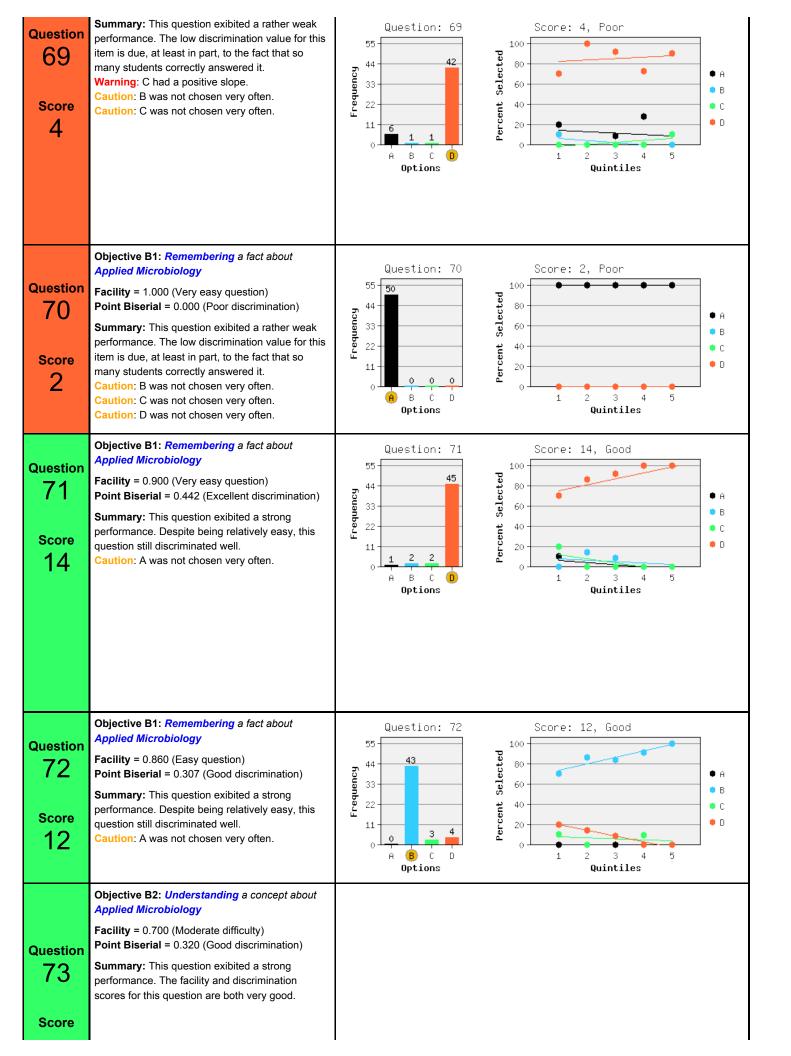


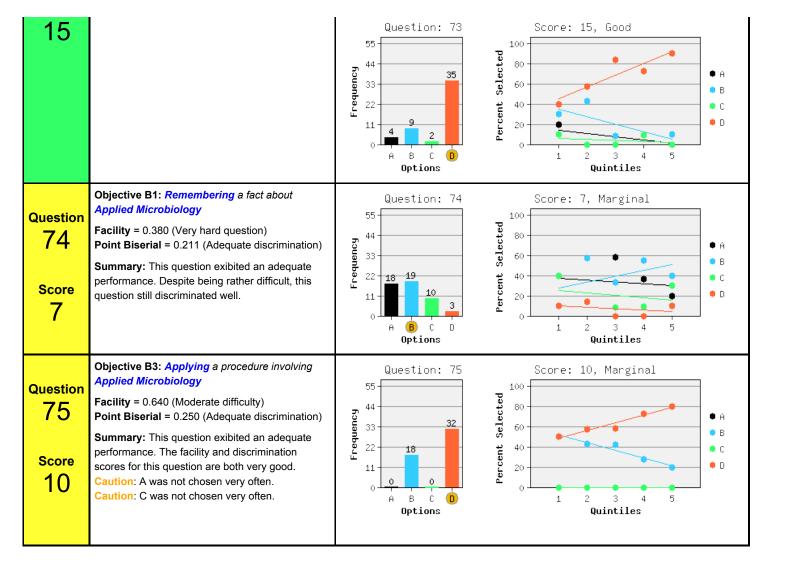












H) Annotated reading list back to top

This format and content of the feedback provided by this program reflects the prior work and knowledge of many different groups. In the space below, I have created a short list of some of the more helpful materials that I've found during the time spent developing this project.

Anderson., L W., & Krathwohl, D. R. (eds.) (2001). A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York: Longman. [Find it in Amazon]

This book gives is an excellent update to the original taxonomy defined by Bloom in the 1950's. They use a two-dimensional taxonomy (cognitive process dimension and knowledge dimension). I have slightly tweaked this to make it cognitive process vs content area.

Crowe A., Dirks C., Wenderoth M. P. (2008). Biology in Bloom: implementing Bloom's taxonomy to enhance student learning in biology. *CBE Life Sci. Educ.* **7**:368-381. [Find it online]

This article is an excellent example of mapping multiple-choice exam items to course outcomes and Bloom's taxonomy. Many specific examples are given for various fields of Biology. Furthermore, they address the importance of formative feedback to enhance student comprehension.