

Martha Fisher

Assessment Analysis and Critique

4/24/2013

Overview

This test was designed to assess 10th grade biology students' learning of a unit on evolution. Students began this unit the week before Spring Break and had one additional week of instruction after returning from break. During that second week, I was absent one day for the William & Mary job fair and my cooperating teacher covered the remaining material for the unit. Students finished with test material that Wednesday, had a related lab on Thursday, and started a mini-unit on classification Friday (a topic which I originally intended to be included on this test but pushed back to a separate quiz because of timing). The test was administered the following Monday. Students in this block were given a review packet on Friday that they could turn in prior to the test for five extra credit points on their test. In addition, students were given a list of relevant pages from the notebooks to study for the test. After the test, students were given the opportunity to complete test corrections independently in which they were to identify the correct answers for questions they missed and explain in full sentences why that answer is correct. For each correction, they will receive half the points back.

I have analyzed the results for first block, and the results of the test are shown below:

First Block Test Results – Percentage Correct Per Question				
1. 83%	8. 50%	15. 79%	22. 79%	29. 58%
2. 96%	9. 83%	16. 88%	23. 50%	30. 75%
3. 100%	10. 50%	17. 75%	24. 25%*	31. 92%
4. 67%	11. 33%	18. 67%	25. 67%	32. 71%
5. 42%	12. 75%	19. 79%	26. 50%	33. 71%
6. 83%	13. 54%	20. 79%	27. 75%	34. 58%
7. 79%	14. 96%	21. 46%	28. 46%	35. 58%
Mean: 78% C Median: 83% C Mode: 88% B Range: 51 – 103%				
*After reviewing the test, I omitted question 24 because so many students missed it. Looking at the question, I determined that the problem was not with the question itself, but that I did not cover it enough during class and in review. The other low percentage questions I reviewed and decided that the questions were fair and in my judgment that the students had instruction and practice enough with the material to be held responsible.				

Considering your students in the aggregate, what inferences do you draw about the nature and degree of student learning with regard to the learning objectives targeted by this test?

After the reviewing the test results, I mapped the test questions back onto the table of specifications and calculated summary figures of percentage correct for each content strand and each cognitive level. The figures are

shown in the table of specifications below along with color coded question numbers to indicate exceptionally low and high performance.

Table of Specifications (*Original ILOs attached at end)

Content	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation	Student Performance
Evidence for Evolutionary Relationships (30%)	1	Describe 17	Describe, Infer 3, 4	Infer 6, 9, 16	-	-	83%
Fossil Record (10%)		Differentiate between	Determine, Compare 2, 5, 7, 8		-	-	67%
Natural Selection (30%)		Explain how 22, 31, 33, 36	Predict 14, 19, 20, 34, 35	Analyze 10, 12, 23	-	-	72%
History of evolutionary theory (5%)		Describe how 18		Compare and Contrast 36	-	-	67%
Speciation and macroevolution (20%)		Describe, Discuss, Compare and Contrast 24, 25, 26, 27, 28, 29, 30	Apply 11, 21	Compare and Contrast 13, 15, 37	-	-	55%
Student Performance	83%	61%	68%	70%	-	-	

Learning of Content Strands:

Students performed the best on evidence for evolution. It was the first topic of the unit and got the most emphasis in terms of review, so it makes sense that students would do better on this topic than others. The second highest strand of performance was natural selection, which was the second topic in the unit and was reviewed, but not as much as I would have liked. The fossil record category was brought down by the questions about radioactive dating which I now recognize I did not spend enough time on initially or in review. Students performed the weakest on speciation and macroevolution. This was the last topic of the unit and I don't believe that I spent enough time on these concepts in instruction or review. Students didn't practice these concepts and I did not have any kind of formative assessment of these concepts before the test to gauge student understanding.

Performance of Cognitive Levels:

The weakest was comprehension. This category was definitely weakened because of the matching section with macroevolution terms. These were oversampled on the test and students weren't prepared for them. The true ILOs were primarily at the application and analysis level as well and should not have been tested so heavily at the comprehension level with a matching section. Students weren't prepared with enough practice with this vocabulary, and it was not completely fair of me to expect this from my students because they were not adequately prepared. Students did better (although overall low) on the application and analysis questions. This is more of what they were prepared for with practice questions and instruction than comprehension level questions.

Select at least two individual students and analyze their respective results on this assessment. What inferences do you draw about the nature and degree of their learning with regard to the intended learning outcomes?

Student A received a 95 on the test, having missed five multiple choice questions and received full credit on short answer questions. She received five extra credit points for turning in her review packet (without this her grade would have been a 90). She missed questions 1, 10, 11, 12, 23. She missed all three questions about analyzing the effects of natural selection. Overall, she performed well on the test with this one area of weakness or misunderstanding. Because this student is a good student who generally pays attention, studies and does her work in class, I can infer that there was probably a lack of adequate learning experiences to address this ILO (#8 – *Analyze the impact of reproductive and survival strategies on fitness of a population*). This student turned in test corrections that will bring her grade from a 95 to a 100 on the test.

Student B received a 65 on the test, a failing grade. She missed 14 multiple choice questions and got no credit for first short answer question and partial credit on second short answer question. She missed questions 5, 7, 8, 9, 10, 11, 13, 15, 19, 25, 27, 28, 29, 34. Mapping each of these missed questions back onto the table of specifications, it appears that most questions she missed were at the application and analysis, particularly with fossil record, natural selection, and macroevolution. This indicates to me that this student needed not only more instruction or practice in these specific learning objectives, but that she would probably benefit from working with general skills of applying and answering questions using concepts. She also missed a significant portion of the matching section, something that many students had difficulty with. As I mentioned previously, this is probably due more to a lack of adequate learning experiences from me than from a particular weakness in this student. Nonetheless it is a weakness in many students that will need to be addressed. This student turned in test corrections that will bring her test grade from a 65 F to an 83 C.

Given the results on this assessment and your reflection after having administered it, what instructional decisions are you led to and why?

In the short-term, I will need to re-teach and have students practice with vocabulary for speciation and macroevolution, because very many students missed these questions and I know I did not provide enough instruction for this. I also intent to have students practice answering application questions, as that seems to be the focus of most SOL questions and students had some difficulty with these types of questions. Students have their benchmark this coming Tuesday and evolution will be one unit on the benchmark, so I will be implementing various review strategies in the upcoming class periods.

In the long-term, next year for this unit and for all of my units that I cover, I want to make sure that ILOs are clearer to students. I would like to actually hand out ILOs with accompanying questions for students as a study guide. Though I did hand out a review, it was not one that I had created from scratch and was not as closely aligned with my ILOs and with the test as I would have liked. Though I believe that students were told what content would be on the test, I believe that they needed to be more clearly informed about what they would actually be asked to do with that information and held responsible for. I also plan to build in more opportunities

for ongoing practice and review of concepts throughout the unit. Finally, I will add in several formative assessments like quizzes built into the schedule of the unit in order to have ongoing remediation instead waiting to discover any weaknesses at the end of the unit. Additionally, this will encourage students to study in small chunks throughout the unit instead of waiting until the test.

Given the results on this assessment and your reflection after having administered it, what conclusions do you draw about the validity and reliability of this assessment? How would you edit, modify, or revise this test in order to improve it?

I realized after mapping the test back onto my table of specifications that some areas were oversampled and did not closely reflect the intended learning objectives. This weakens the content validity of the test because it does not align as strongly with the curriculum or the instruction as it ought to. Though the test still has acceptable validity and it does test concepts that were covered in class, there is much room for improvement in terms of improving the validity.

For the reliability of the test, there was no cultural bias or grammatical errors in the test that I could identify. Test items were written to be as reliable as possible. Particularly with multiple choice items because they make up the majority of the test I ensured that the question was in the item stem, the choices were in random or alphabetic order, the choices had similar formats and the distractors were plausible. For the SOL or other adapted questions that I did not write myself, I screened for these rules and modified them slightly if I found problems. Many test items include surplus information like graphs, images, or diagrams. I attempted to provide adequate explanation for this surplus information so that all the information students need to answer a question is provided for students if they understand the concepts being assessed. Some students may have had difficulty with some questions because of their reading ability. I tried to make all items appropriate to student's level, while also preparing them to answer questions that may require them to read and identify the questions like they will experience in the SOLs. Finally, one threat to reliability that I did notice while reviewing the tests were some instances where I suspect that students were sharing answers. This only occurred with multiple choice questions so there is no conclusive proof that cheating occurred. This negatively affects the reliability of the test.

To improve the test, I would eliminate the matching section, because I believe many students struggled with that not only because of inadequate instruction but also because of difficulty distinguishing between concepts that may not have been worded clearly. On previous tests and quizzes, I have noticed that matching sections consistently have very low scores. I think students may get thrown off because they put a guess for one and that throws everything off. In the future I would consider choosing an alternative way to test these concepts, even something related like fill in the blank. Though this is a very similar format, I believe that it allows students to actually think through the question somewhat more thoroughly and self-check their answers more clearly. I would also add additional test items to make the test more closely reflect the emphasis of the ILOs in the curriculum and instruction. For instance, I would add more questions about applying evidence for evolution and natural selection. Finally, to prevent cheating, I would create a form A and form B with scrambled questions and distribute them so students did not sit next to someone with the same form.

*ILOs

1. Describe relationships based on homologous structures, analogous structures, and vestigial structures
2. Infer evolutionary relationships based on similarities in embryonic stages and biochemical evidence
3. Compare structural characteristics of an extinct organism, as evidenced by its fossil record, with present, familiar organisms.
4. Differentiate between relative and absolute dating and determine relative age of a fossil given information about its position in the rock and dating by radioactive decay
5. Explain how genetic mutations and variation are the basis for selection and how natural selection leads to changes in gene frequency of a population over time
6. Predict the impact of environmental pressures on populations
7. Describe how the work of Lamarck and Darwin shaped the theory of evolution and compare and contrast their theories
8. Analyze the impact of reproductive and survival strategies on fitness of a population
9. Describe the mechanisms of emergence of species.
10. Apply the biological species concept to novel scenarios and discuss problems with the definition.
11. Compare and contrast punctuated equilibrium with gradual change over time.
12. Compare and contrast convergent evolution with divergent evolution
13. Construct and utilize dichotomous keys to classify groups of objects and organisms
14. Interpret a cladogram or phylogenetic tree showing evolutionary relationships among organisms.
15. Investigate flora and fauna in field investigations and apply classification systems