POINT LOMA NAZARENE UNIVERSITY

Using Interview Data from Non-Major Biology Students To Improve the Conceptual Inventory of Natural Selection

A thesis submitted in partial satisfaction of the

requirements for the degree of

Master of Science

In General Biology

By

Danielle Maureen Dwyer

Committee in charge:

Dr. Dianne Anderson, Chair
Dr. Ryan Botts
Dr. April Maskiewicz

2011

The thesis of Danielle Maureen Dwyer is approved, and it is acceptable in quality and form for publication:

____

____

____
Chair

Point Loma Nazarene University
2011

Table of Contents

List of Figures and Tables

Acknowledgment

I would like to acknowledge Dr. Kathleen Fisher for taking the time to help with making some of the changes that were implemented for the revised version of the CINS. Your help was greatly appreciated and very valuable in this study. Furthermore, I would like to acknowledge Dr. Ryan Botts for helping me with all of the statistical data by running all of the numbers and explaining to me what they mean. I would also like to thank Dr. Dianne Anderson for being my advisor and working on this study with me, and for guiding me along the way on this long journey. Lastly, I would like to acknowledge, Dr. Mike Dorrell, Dr. April Maskiewicz and Professor Anita Plagge for allowing me the time to use their classroom and students for my data collection.

Abstract of the Thesis

Using Interview Data from Non-Major Biology Students

To Improve the Conceptual Inventory of Natural Selection

By

Danielle Maureen Dwyer

Master of Science in General Biology

Point Loma Nazarene University, 2011

Dr. Dianne Anderson, Chair

The conceptual Inventory of Natural Selection was first published in 2002 (Anderson & Fisher, 2002). Recent research by Nehm & Schonfeld (2008) has raised concern as to whether or not the Conceptual Inventory of Natural Selection (CINS) is effective in assessing students' understanding on concepts related to natural selection. The purpose of this study was to obtain feedback from non-major biology students, through semi-structured interviews, on 12 items that generated low scores or have not paired appropriately on factor analysis on the current version of the CINS. Interviews indicated that improvements needed to be made. The improvements made on the 12 CINS items included word substitutions and minor sentence restructuring to preserve the intent of the original CINS, yet to improve it. Field testing with both biology majors and non-majors showed that the changes were beneficial, although further testing with larger sample sizes is warranted.

Introduction

Natural selection is widely accepted as the most common mechanism of how species evolve. In order for individuals to have a scientific understanding of evolution, it is important to know what natural selection is and how it works. Even though the public has been exposed to the topic of natural selection, both through the media and in school, the persistent lack of understanding has been blamed on the media and educators (Alters & Nelson, 2002; Crawford, Zembal-Saul, Munford, & Friedrichsen, 2010).

Accordingly, teachers need ways to assess their students' current understanding of natural selection in order to evaluate the effects of instructional reform. Research has been conducted to assess secondary and post-secondary students' understanding of natural selection in the form of a diagnostic test called the Conceptual Inventory of Natural Selection, or CINS (Anderson, Fisher, & Norman, 2002). The problem is that recent research has raised concerns about whether or not the CINS is effective in assessing students' understanding of natural selection. This indicates the need for more research.

Conceptual inventories are designed for educators to obtain an understanding of what conceptions their students hold on a topic, by using distracters based on common student ideas as answer choices, thereby identifying the current alternative conceptions that the student holds so that educators can help foster conceptual change (Tanner & Allen, 2005). Conceptual inventories can either be in the form of criterion-referenced or norm-referenced tests. Criterion-referenced tests attempt to assess students' understanding of a given topic by creating a list of desired content, then creating test items to determine the students' understanding of the content. With these tests, all students in a group could score high. Norm-referenced tests, on the other hand, are tests that allow for only a few students to do well, which may be represented by a normal bell curve or distribution of scores.

Several researchers have indicated that the CINS is a valuable tool to measure student understanding of natural selection. Battisti (2009) studied how the distractors of the CINS test designated the level of understanding that the student has on the subject. Depending on the level of understanding that the student holds, he or she would either answer correctly or choose a specific distractor. Battisti et. al. used Thissen and Steinberg's (1984, 1997) multiple choice (MC) model and Item Response Theory (IRT) on the CINS to show at what level of understanding the student would choose the correct answer over the distractors provided. By using the MC model, they were able to estimate the probability of test guessing. Their results showed that some CINS item pairs were more difficult for students than others. For the students who had a lower level of understanding, it was seen that the distractors were the chosen over the actual answer (Battisti, Hanegan, Sudweeks, & Cates, 2009). Their research provided recommendations and suggestions that could help improve the CINS to be a more valuable test, which indicates the need for the current study.

There are certain CINS items that have proven to be especially difficult for students (Battisti, Hanegan, Sudweeks, & Cates, 2009; Nehm & Schonfeld, 2008). In addition, some item pairs which test the same concept do not load together on factor analysis (Nehm & Schonfeld, 2008). Research is needed to determine if there are problems with the items, or if the difficulty is truly due to students holding alternative conceptions. The purpose of this study was to allow for the improvement of the CINS based on feedback form student interviews, so that it can be a better tool to assess students' conceptions on the topic of natural selection. The CINS items that have been shown to be problematic in previous research (4, 6-10, 13, and 17-20) per Anderson, Fisher, and Norman (2002), along with the data collected in a pilot study by the present study's author, were studied in depth to see if they accurately assess students' understanding. This was achieved using semi-structured interviews, as well as think-aloud prompts (Southerland, Smith, & Cummins, 1999), which allowed students to express their knowledge and understanding of the topic verbally. The data analysis provided a starting point for the improvement of the wording of some questions so that the students correctly understand what the question is asking. This will give a more accurate assessment of conceptions the students actually have about natural selection by confirming that their answer choice that closely matches their understanding.

Theoretical Framework

Students come into the classroom with their own pre-existing knowledge about how they view the world around them. When an individual revises his or her own current beliefs, thoughts, or knowledge on a topic, a conceptual change may take place, thereby allowing for the learner to have a better understanding of the topic or theory (Hewson, Tabachnick, Zeichner, Blomker, Meyer, Lemberger, Marion, Park, & Toolin, 1998; Jensen & Finley, 1996; Larsson & Hallden, 2010; Tanner & Allen, 2005; Tsui & Treagust, 2005). Conceptual change is a learning process that essentially allows for the existing conception that is held by the student to be shifted and reconstructed, so that it becomes more like the current conception that is held by the "experts" in the subject.

Educators can help with this process of conceptual change by having the learner examine his or her own pre-existing conception. When the students are able to explore and analyze the evidence and information for the topic, their current conceptions will be able to be built upon, or countered, depending on the learner's existing knowledge that they already have (Tanner & Allen, 2005). Conceptual change can be a long-term process, due to the time invested in examining, evaluating, and testing the validity of the information that is being presented in the classroom. According to Taber (2001), the environment in which the student is learning and the schemas that the student has in place (which will be explained later), will play a role in whether or not the student's schema will be positively reinforced or rejected when the student fails to apply the information that is being presented, thus a conceptual change can occur.

Conducting investigations, identifying sources to help build knowledge, and applying the concept to situations can promote conceptual change in the classroom. These actions will be an internal process, with students having to consider the new information within the context of their prior knowledge and how they view the world (Tanner & Allen, 2005). Once the student has fully internalized the information that he or she has learned, understanding of the concept can be achieved. The students' internal process can also be influenced by the social components that can occur in the classroom (Vosniadou, 2007) such as learning in groups and/or pairs by seeing or hearing the information based on how their peers perceive and understand the content. Faced with new interactions and experiences, students will be able to use the prior knowledge that they have and build upon it so that they can construct new knowledge and understanding

(Tanner & Allen, 2005; Vosniadou, 2007; Zirbel, 2004). Once this occurs, a conceptual change can happen that allows for understanding of the topic.

Reasoning or understanding of the activity is based upon the situation and how it is experienced by the individual. Learning can occur with different experiences, both inside and outside of the classroom. These experiences are based on how individuals perceive the world around them and how their perceptions are internalized. Jean Piaget (1963) describes how individuals learn and how knowledge is constructed through experience. Students are always actively constructing knowledge, and Piaget (1963) described learners as “little scientists”, who are constantly exploring and making sense of their world (Meece & Daniels, 2008; O’Donnell, Reeve, & Smith, 2007). Each individual will develop a scheme or schema, which refers to cognitive structure(s) that will represent an experience, idea, or concept that has been identified and interpreted, based on an individual’s experience (Meece & Daniels, 2008; O’Donnell, Reeve, & Smith, 2007). Furthermore, teachers can facilitate the learning process by creating exploratory opportunities in the classroom. As a result of cognitive conflict, the student will either incorporate that idea into current schema (assimilation), modify existing schema in order to make sense of the new idea (accommodation), or reject the information altogether. Both assimilation and accommodation will account for the growth and developmental change in a student’s schema. Using a constructivist perspective, this study shows how a diagnostic test can be used to show what conceptions the students already hold.

Literature Review

Alternative Conceptions

Alternative conceptions are non-scientific ideas or knowledge that students hold based on prior experiences and attempts to make sense of the world around them. In this paper the term will be used to denote students' understanding of scientific conceptions not aligned with the current understanding of the scientific community (Bishop & Anderson, 1990; Tanner & Allen, 2005; Vosniadou, 2007). Knowledge or information that the student internalizes from experiences will be interpreted based upon his or her own reality. Sometimes the knowledge that is internalized by the student is incorrect; an alternative conception is formed because it was based on the student's prior knowledge and experience, rather than on a scientific explanation.

By the time they come to the classroom, students have already developed ideas, and alternative conceptions based on how they view the world and formed explanations to explain different phenomena. Those alternative conceptions could also have been due to insufficient or misinterpreted instruction from previous years. Consequently, those alternative conceptions that the students already hold may support the preexisting misconceptions of the subject being taught (Zirbel, 2004).

Natural selection is a topic that is often misunderstood by students (Anderson, Fisher, & Norman, 2002). The prior knowledge that students have and bring into the classroom makes a difference in how they understand natural selection (Jensen & Finley, 1996). This knowledge is often influenced by culture and beliefs, which can play an essential role in how individuals shape their generic descriptions about the theory of evolution (Dagher & BouJaoude, 2005; Jokayem & BouJaoude, 2008). Students who do not have any religious beliefs are not exempt from alternative conceptions about natural selection; nevertheless, having religious beliefs may provide additional complications for a student learning about natural selection.

Chinsamy and Plaganyi (2007) discussed why the concept of evolution often challenges the student's religious beliefs and values. Like most Americans, we are raised with morals, beliefs, and values that have been ingrained in us since birth. The topic and teaching of evolution in the classroom brings up facts and evidence which could lead a student to question what they learned from their church, thereby creating a spiritual and intellectual quandary about what to accept as true. This internal conflict can also lead to the formation of alternative conceptions.

Diagnostic assessments using alternative conceptions as distractors

Over the past two decades, there has been some disagreement over what types of "instruments" are most effective at assessing students' understanding, whether they are in the form of a multiple choice diagnostic test or open-ended questions. Many states use high-stakes assessments to test student performance on state standards, but this essentially encourages teachers to "teach to the test" (Blanchard, Southerland, Osborne, Sampson, Annetta, & Granger, 2010). In order to assess large numbers of students, a multiple-choice test is the most practical option (Anderson, Fisher, & Smith, 2010) given that they can be objectively scored (Battisti, Hanegan, Sudweeks, & Cates, 2009).

Unlike typical high-stakes multiple-choice assessments, the CINS was constructed by using the most common student alternative conceptions as distractors, as will be explained later. A concept inventory can measure conceptual change more effectively than a standardized test because conceptual inventories can be used in the classroom as a pre/post-test (Battisti, Hanegan, Sudweeks, & Cates, 2009; Nehm, Kim, & Sheppard, 2009; Sadler, 1998). When used as a pre-test, it allows the educator to pinpoint the students' alternative conceptions. Teachers can then guide students through constructivist activities to deepen their understanding of the given topic of natural selection.

Natural Selection

Natural selection is a mechanism of evolution that was defined by Charles Darwin in the 19th century. The underlying theme is that natural selection occurs by differential survival based on possession of certain traits, and when those traits are heritable, evolution will occur. Darwin believed that all species and individuals within a population face a 'struggle for existence', and that those that are better suited for the environment will survive (Gould, 2007_a; Gould, 2007_b; Lewens, 2010; Wesson, 1991; Zimmer, 2001). Species are the units of evolution and biodiversity (Catley, 2006; Ereshefsky, 2010) because they drive evolution which depends on the selection pressures that are constantly changing (Weiner, 1995). Darwin believed that all life has descended from a common ancestor, an idea that has been supported by the fossil record (Catley, 2006; Gould, 2007_a; Gould, 2007_b; Weiner, 1995; Wesson, 1991; Witting, 2008; Zimmer, 2001), which holds that over millions of years, the common ancestor has diverged, forming the tree of life (Grehan, 2001). Much of that divergence can be explained by natural selection.

Mayr (1982) identified the five facts and three inferences that he believed were the basis for Darwin's construction of the theory of natural selection:

Fact 1: All populations have the potential to grow at an exponential rate.

Fact 2: Most populations reach a certain size, and then remain fairly stable over time.

Fact 3: Natural resources are limited.

Inference 1: Not all offspring survive to reproductive age, in part because of competition for natural resources.

Fact 4: Individuals in a population are not identical, but vary in many characteristics.

Fact 5: Many of the characteristics are inherited.

Inference 2: Survival is not random. Those individuals with characteristics that provide them with some advantage over others in that particular environmental situation will survive to reproduce, whereas other will die.

Inference 3: Populations change over time as the frequency of advantageous alleles increases. These could accumulate over time to result in speciation.

In addition to the facts and inferences identified by Mayr, the concepts of origin of variation and origin of species can be added to complete the picture of how evolution works. All species have the potential to grow exponentially,(*biotic potential)*, but rarely do as a result of the amount of resources available (*natural resources)* which will keep the population stable, to reach carrying capacity (*populations are stable)*. All living organisms require certain essentials such as food, water, and shelter in order to survive. When resources are available and abundant, the population will be able to grow exponentially, but typically, resources are limited, and competition will occur either between individuals in a population or between two different populations (Anderson, Fisher, & Norman, 2002; Gould, 2007[a]; Gould, 2007[b]; Weiner, 1995; Wesson, 1991; Witting, 2008; Zimmer, 2001). In nature, populations often fluctuate depending on the resources that are available, levels of predation, and intraspecies and interspecies competition. Therefore, in most populations, only a fraction of offspring survive (*limited survival*).

Due to random mutation and genetic recombination (*origin of variation*), individuals within a population have different versions of traits, called alleles. Those traits will differ among the individuals (*variation*) within the population (Gould, 2007_a; Gould, 2007_b; Weiner, 1995; Wesson, 1991; Witting, 2008; Zimmer, 2001). Those alleles are inherited and passed down from parents to their offspring (*variation inherited*), allowing certain individuals to be better suited or adapted to their environment than others, and thus able to survive and reproduce (*differential survival*). Over time, the frequency of those alleles in the population can change from generation to generation, causing evolution (*change in population*) to occur. If a population becomes separated due to barriers, e.g. physical, temporal, etc., then once closely-related populations of one species will diverge causing those populations to become genetically different (*origin of species*) (Grehan, 2001; Weiner, 1995; Wesson, 1991; Zimmer, 2001). These two populations will still be similar, but will now be sub-species, which will usually lead these populations to further diverge from one another, causing mating between the two populations to cease resulting in the formation of new species.

Development and Use of the CINS

The CINS test is a 20-question multiple-choice test that was developed to assess students' conceptions, current knowledge, and understanding on natural selection (Anderson, Fisher, & Norman, 2002). Before the development of the CINS, an open-ended natural selection test was created by Fisher and a grad student, C. Sandifer, which was based on the research conducted by Bishop and Anderson in 1990 (Anderson, Fisher,

& Norman, 2002). This open-ended natural selection test was then administered to non-biology major undergraduates, and served as a starting point for Anderson and Fisher's development of the CINS test. Semi-structured interviews were then conducted to pinpoint the students' understanding of natural selection.

The CINS was constructed as a concept inventory by developing questions to assess understanding of each of the concepts identified by Mayr (Anderson, Fisher, & Norman, 2002; Anderson, Fisher, & Smith, 2010). Along with these concepts, questions to assess the concepts of origin of variation and origin of species were also developed, since a comprehensive understanding of natural selection would include these ideas, as well (Table 1).

Anderson and Fisher argue that the CINS test is a realistic and comprehensive test because it focuses on ten scientific ideas related to natural selection, along with the alternative conceptions they align with (Anderson, Fisher, & Norman, 2002). For each concept, the authors identified at least two to three alternative conceptions that students can possess (Table 2). The research has shown that students have a hard time with the concepts of variation, inheritance of variation, and the origin and survival of new traits (Jensen and Finley, 1996). These difficult concepts and related twelve CINS items were addressed during this study.

Table 1
Natural selection concepts on the CINS, definitions taken from Anderson, Fisher, & Norman, 2002.

	Concept	**Questions**	**Definition**
1	Biotic Potential	1, 11	The potential that all populations have the ability to grow exponentially, assuming no limits on food, space, etc.
2	Populations are stable	2, 14	Most populations do not show sustained exponential growth, but rather the population stays relatively stable or crashes.
3	Resources Limited	3, 12	Individuals in a population compete for resources.
4	Limited Survival	5, 15	Some do not survive when competing with one another for resources.
5	Variation	9, 16	Within a species organisms differ from one another's' inherited traits.
6	Variation Inherited	7, 17	Parents pass on some of their traits to their offspring.
7	Differential Survival	10, 18	Offspring that are better suited to their environment will be the most successful in reproducing.
8	Change in population	4, 13	Frequency of alleles (traits) can change in succeeding generations.
9	Origin of Variation	6, 19	Variations arise through mutations and genetic recombination.
10	Origin of Species	8, 20	Distinct species can arise when two populations of a species are separated for a period of time by barriers: physical, temporal, etc.

Before the CINS became a multiple-choice test, the researchers considered using the two-tiered test format because the design of the questions separates factual knowledge (Tier 1= facts) from reason for choosing a particular fact (Tier 2= mechanisms and beliefs). The reason why they did not continue with the two-tiered format was because with the large, complex, and multi-faceted concepts such as evolution, the two-tiered system breaks down (Anderson, Fisher, & Norman, 2002). Moreover, it was too complex for a two-tiered diagnostic test to be administered in a pencil-and paper format, thus they chose a single-tier multiple-choice formant instead.

Table 2
Alternative conceptions based on the research conducted by Anderson, Fisher, & Norman (2002)

	Concept	Alternative Conception
1.	Biotic Potential	A. Not all organisms can achieve exponential population growth. B. Organisms only replace themselves. C. Populations level off.
2.	Population Stability	A. All populations grow in size over time. B. Populations decrease. C. Populations always fluctuate widely/randomly.
3.	Natural Resources	A. Organisms can always obtain what they need.
4.	Limited Survival	A. There is often physical fighting among one species (or among different species) and the strongest one wins. B. Organisms work together and don't compete.
5.	Variation within a population	A. All members of a population are nearly identical. B. Variations only affect outward appearance, and don't influence survival. C. Organisms in a population share no characteristics with others.
6.	Variation Inheritable.	A. When a trait is no longer beneficial for survival, the offspring will not inherit the trait. B. Traits acquired during an organism's lifetime will be inherited by offspring. C. Traits that are positively influenced by the environment will be inherited by offspring.
7.	Differential Survival	A. Fitness is equated with strength, speed, intelligence, or longevity. B. Organisms with many mates are biologically fit.
8.	Change in population	A. Changes in a population occur through a gradual change in all members of a population. B. Learned behaviors are inherited. C. Mutations occur to meet the needs of the population.
9.	Origin of species	A. Organisms can intentionally become new species over time. B. Speciation is a hypothetical idea.
10.	Origin of variation	A. Mutations are adaptive responses to specific environmental agents. B. Mutations are intentional. An organism does it because it wants, needs, or tries to genetically change.

The CINS was constructed by using many of the alternative conceptions expressed by students during interviews as shown in Table 2.The test items were field-tested in both community college and university classrooms. Version 1 composed of four sets of five questions and was used as a pre- and post-test. Per Anderson, Fisher, and Norman (2002), the interviews that were conducted after the pre- and post-test were used to determine whether the scores on the test reflected the students understanding of the concepts of natural selection. After the first version was analyzed, modifications to the test questions were made based on student feedback from the interviews. For example, the salamander questions that were on the original version were replaced with Venezuelan guppies, and there were now three passages in the headings that preceded specific questions. Version 2 was then field-tested and the results showed that it was necessary to have two questions that assess each concept. More corrections were made by deleting words from the passages and questions, expanding to the current 10 concepts, and creating the final version (Appendix A).

The research by Anderson, Fisher, and Norman (2002) determined whether or not the students understood all of the factors that contributed to natural selection. This proved to be efficient by showing what conceptions those students held. Research from this thesis author's pilot study conducted spring of 2010, involved getting feedback, via interviews, on the CINS test from four college graduates. This study showed that certain questions were not accurate in showing what students understood. From the research, it has been concluded that the CINS is a valuable tool for teachers to use in their classrooms to help assess their students' understanding, but it was also clear that it could benefit from some improvements.

Criticisms of the CINS

Nehm and Schonfeld (2008) reported problems with the CINS questions, claiming they do not accurately assess students' understanding of alternative conceptions about natural selection. In turn, Nehm and Schonfeld developed their own instrument, the Open Response Instrument, or ORI. The ORI is an open-ended essay test created by taking questions from Bishop and Anderson's (1990) and Nehm and Reilly's (2007) earlier studies, which focused on seven key concepts, described in Table 3. Nehm and Schonfeld developed the ORI in order to measure potential learning gains.

Table 3
Seven key concepts of natural selection from Nehm and Schonfeld's ORI

Concepts tested in the ORI
1. The causes of phenotypic variation (e.g., mutation, recombination, sexual reproduction).
2. The heritability of phenotypic variation.
3. The reproductive potential of individuals.
4. Limited resources and/or carrying capacity.
5. Competition or limited survival potential.
6. Selective survival based on heritable traits.
7. A change in the distribution of individuals with certain heritable traits.

For Nehm & Schonfeld's research, they used a discriminate validity instrument in the form of a multiple choice test on the topic of rocks, which was developed from the New York State Regents earth science exams from 2001 to 2006 (Nehm & Schonfeld, 2008). This was done for a number of reasons: (1) its structure was similar to the CINS, so it might expose those students who are good at taking multiple-choice tests; (2) the majority of participating undergraduates would not have taken earth science; and (3) the questions were appropriate for college level (Nehm & Schonfeld, 2008). This test was used to reveal whether or not the questions on the CINS or the ORI were well constructed by examining the correlation between the rock test, the CINS, and the ORI.

Nehm and Schonfeld (2008) interviewed the participants and, in order to analyze their transcripts, they constructed their own coding rubric. Using that rubric, the information gathered from the interviews measured the key concepts and alternative conception for each student. They tested two classes of biology majors who were in their second semester course. One class was able to complete both the ORI and CINS, while the other class only completed the ORI due to time constraints. The researchers took two different theoretical approaches in the analysis of their data: the Classical Test Theory (CTT) and the Item Response Theory (IRT).

Like Anderson, Fisher, and Norman (2002), Nehm and Schonfeld used CTT to evaluate and validate the CINS because it explores item attributes in context by focusing on what distractors are chosen, which is in line with concept inventories. The CTT analysis provides the opportunity for analyzing a student's performance based directly upon the total score of the subject by looking at the point biserial values, with these measurements highly dependent on the test population (Sadler, 1998).

Along with CTT, Nehm and Schonfeld used Item Response Theory (IRT), in the form of Rasche Analysis. For the purpose of this study, the type of theories used and analysis will be mentioned only to identify how the researchers approached and analyzed their results, but will not be explained in depth. Rasche Analysis is conceptually analogous to measurements of reliability, validity, and item difficulty, and is a common IRT approach (Nehm and Schonfeld, 2008; Sadler, 1998). IRT appears to provide a better framework than CTT for understanding performance on conceptual inventories.

Nehm and Schonfeld concluded that the ORI produced a richer description and diversity of alternative conceptions than the CINS. They believed that Anderson, Fisher, and Norman (2002) had a small sample size, that their interviews were problematic, and that the sample size did not provide enough evidence to support their validation of the CINS (Nehm and Schonfeld, 2008). According to Nehm and Schonfeld, the biology majors who participated in their study did poorly on the CINS because the students found the alternative conception distractors so compelling (Nehm & Schonfeld, 2008) that they chose the wrong answers. In terms of the items on the CINS that tested for the same concept, the scores showed that certain items on the CINS were not loading in the appropriate pairs. The researchers concluded that the CINS test failed to provide any measure of students' abilities in regard to assembling the pieces and forming coherent and functional explanatory structure (Nehm and Schonfeld, 2010), so essentially from their perspective, the CINS test fell short.

Nehm and Schonfeld concluded that a combination of both the CINS and the ORI would be more beneficial. From their 2010 paper, they stated that numerous CINS items display unacceptable levels of discriminability and difficulty values using any psychological measurement standard or methodology (CTT or IRT) (Nehm & Schonfeld, 2010), so the items were not loading appropriately on factor analysis when the tests were performed. The item pairs that were not loading together for the different concepts were: 2 and 14 (*populations are stable*), 3 and 12 (*resources are limited*), 6 and 19 (*origin of variation*), 7 and 17 (*variation inherited*), 8 and 20 (*origin of species*), and 9 and 16 (*variation*). Overall, Nehm and Schonfeld strongly believed that the CINS needs to be improved, and is better suited for biology majors over non-majors due to the difficulty of the test items, even though Anderson (2002) presented evidence that the CINS scores were significantly correlated with interview scores obtained with non-majors.

Anderson, Fisher, and Smith (2010) responded to the research that was conducted by Nehm and Schonfeld in their 2010 response article. According to Anderson, Fisher, and Smith, the CINS test was criticized because it did not fully assess students' conceptions on natural selection, but the importance of the CINS is that it is a diagnostic measure that serves to monitor the progress of understanding in natural selection. Anderson, Fisher, and Norman (2002) also tested the CINS on undergraduate ecology majors, who had knowledge of evolution and performed well. From this, it was demonstrated that undergraduates with comprehensive of and experience in biology can identify the correct responses on the CINS. Anderson, Fisher, and Smith concluded that more research needs to be conducted to see if the items in the CINS test load together on the Principal Components Analysis (PCA), which will indicate whether the concept pairs assess the understanding of the same concept.

Another contributing criticism that was stated was the linguistic complexity of the items on the test, which can confound scores on achievement tests (Abedi, Courtney, Mirocha, Leon, & Goldberg, 2005). Students may have a hard time understanding certain words on tests because of their diverse backgrounds, cultures, and languages. Identifying and possibly replacing words or phrases on the CINS questions that are under review may help provide a more accurate measure of students' content knowledge (Abedi et al, 2005). This brings us to the current study which sets out to improve those controversial CINS items.

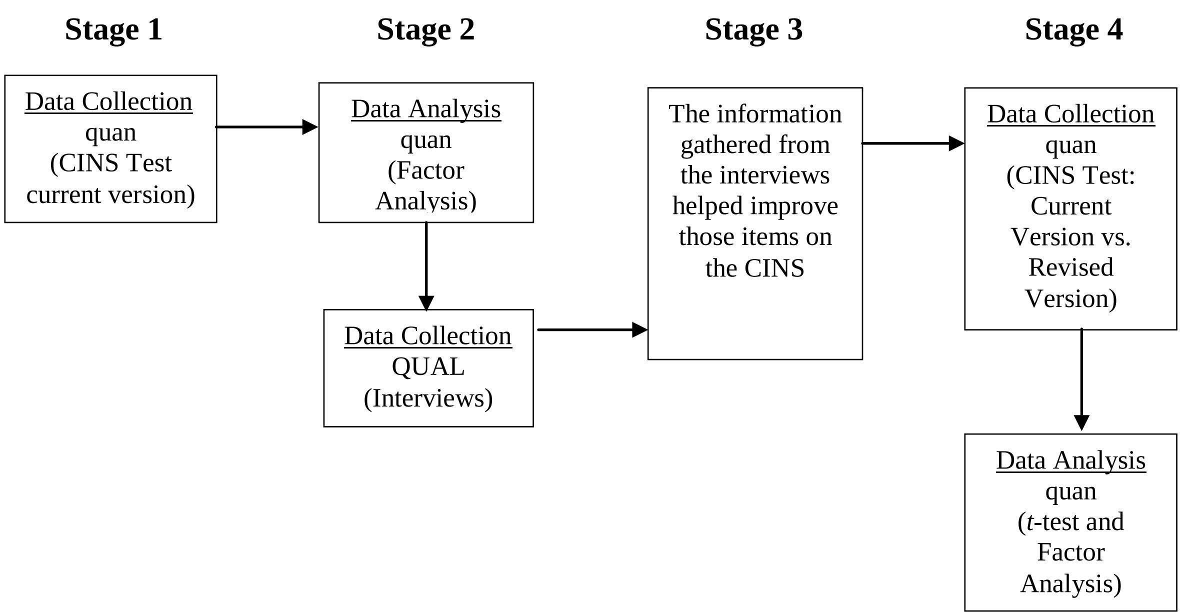
The main focus of this current research was to gather feedback on those CINS items that have been shown to have generated low scores or to have not paired appropriately, and to determine if changes are warranted. These changes included reducing the linguistic complexity by using words and phrases that students are more likely to know and have been exposed to as determined from interviews revealing common prior knowledge This study focuses on concepts five through ten (*variation, variation inherited, differential survival, change in population, origin of variation, and* *origin of species*), and determines if the pairs of items equally assess the same concepts by analyzing whether the student answers both questions either correctly or incorrectly. Therefore, the research questions guiding this study include:

1. Does the linguistic complexity in the CINS items hinder student understanding of the item by using words and phrases to which they may have been exposed to but do not have tangible sense of what the terms mean?
2. For the revised version of the CINS, are the two items that assess one concept equivalent in terms of how the students answer them? If the student answers only one item of a particular item pair correctly, does this mean that they do not fully understand the concept?

Methodology

Research Design

A sequential mixed methods design was used for this study:



In stage one of the study, the current version of the CINS was given to a group of students in a non-majors biology class. Stage two consisted of the qualitative data collection, which involved individual interviews with students from the class. In these interviews, the students verbally went through the problematic items 4, 6-10, 13, and 16-20 from the current CINS. These items were chosen based on the research that was completed by Anderson, Fisher, and Norman (2002) and Nehm and Schonfeld (2008), as well as Dwyer (2010) in her pilot study. The data analysis for stage two included the coding of the interview transcripts, along with running a factor analysis based on the quantitative data from the two lab classes that took the current version. Stage three consisted of examining the words and/or phrases that were difficult for the students in any of the 12 items. During stage three, existing items were refined based on the student feedback. Stage four consisted of the second quantitative portion of this study in which two groups of students who were majors (freshman and junior/seniors), took the revised version of the CINS, along with a group of non-majors. Additionally, one group of freshman majors took the original version of the CINS test. A factor analysis was used to see if the item pairs that were changed performed equally on both versions, which indicates an improvement, and comparisons were made for non-majors vs. majors, as well as original vs. revised CINS.

Study Site and Participants

The study took place at a small Christian teaching university in Southern California. Participants were between the ages of 18 and 22 and were current students enrolled in either a non-majors' general biology course, an ecology/evolution course for freshmen biology majors, or a research methods course for junior or senior biology majors. The IRB approval for this study is located in Appendix A. Students were given consent forms to sign, and their involvement was strictly voluntary if they chose to have their data included in this study (Appendix B).

The first stage of the study involved 45 students who took the current version of the CINS test. These students were given the entire CINS as part of the class activity. One student chose not to participate in this study, and her score was removed from the results. Eighteen of those students who took the original version were interviewed for stage two. Those students who volunteered indicated on the consent form that they would like to be interviewed and were compensated for their time with a ten dollar gift card. For the final 4th stage of the study, students from another non-majors' biology class of approximately 43 students took the revised version of the CINS test. Five students chose not to participate, and their data was taken out, as well. The reason why both the original and revised versions of the CINS were given to different classes of majors' was to indicate how students who have knowledge and understanding of the concepts within natural selection perform on the test vs. students, non-majors', who have been minimally exposed to natural selection. It allowed for a comparison with the non-majors', as well as to help see if the item pairs were assessing the same concept.

Data Collection

Qualitative Data Collection. The semi-structured interviews were conducted with 18 non-majors during stage two. Interviews spanned 26 to 62 minutes, depending on how students responded to each items and to the interview questions, with an average interview time of approximately 39 minutes. The individual interviews were conducted in a faculty office in the biology department and were audio recorded.

The semi-structured questions asked during the interview are listed in Table 4. These questions allowed for insight into how the students interpreted and understood each item. During the interviews, each student had a copy of the questions and was asked to think aloud while working through each item, and state why he or she chose their answer. The student was the one talking throughout the interview, with the interviewer asking more questions only if the student's response was unclear. By interviewing the students, it was possible to identify the concerns that they had about the test itself and with each question. Therefore, students were able to express verbally why and how they felt the way they did about each question, which allowed their thought processes and knowledge on the topic to become explicit. This provided feedback on which to base any revisions of the CINS.

Table 4
Interview questions

Questions will asked for each of the problematic items on the CINS
1. Please tell me which answer you would chose and why.
2. Please tell me why you did not choose each of the other answers.
3. What did you like about the question?
4. Did you understand what the question is asking?
5. Did you have an issue with any of the words within the question? If so, what words would you take out or change?
6. Thinking scientifically, how would you change the question so that it would make more sense to you?

The items that were included in the interviews included 4, 6-10, 13, and 17-20. As previously explained, these items covered the concepts of variation, variation inherited, differential survival, change in population, origin of variation, and origin of species. Investigating these item pairs allowed for the second research question related to item pairing to be answered. The data obtained from the interviews demonstrated how the students verbalized their understanding and expressed their knowledge about the specific topics within natural selection (Ercikan, Arim, & Law, 2010). The second goal of the interviews was to examine how the students went through both the linguistics and content of the questions that were being assessed, in order to answer the first research question.

Quantitative Data Collection. The quantitative data collection took place during the first and the fourth stage of the study. The data collected in Fall 2010 with the existing version of the CINS was compared with the results from the revised version based on the student feedback in stage 2. In stage 4, both versions of the test were given to an undergraduate biology class of freshmen, and a junior/senior majors' class took the revised version. A class of non-biology majors also participated in taking the revised version of the CINS. The revised version of the test was composed of some of the original items (1-3, 5, 11-12, and 14-15) from the CINS test published by Anderson, Fisher, and Norman (2002) and revised in 2004, as well as those revised during the study.

Data Analysis

Analysis of Quantitative data, Part One, Stage one. The students' answers to the CINS items were bubbled in on Scantron® forms. The reports that were used for this study were the standard item analysis, score distribution, test results, and class response. The original, current version data was collected and analyzed using factor analysis. Since two lab sections from the same class took the same test, the factor analysis was able to determine any statistical difference in the items on the test. Comparing the performance on the same concept item pairs showed if there is a reduction in the number of factors, as well as alignment within the same factor. This allowed for any discrepancies to be seen between how the student answered one item over the other on the same concept.

Use of Qualitative Data to Explain Quantitative Data Collection, Stages two and three. The qualitative interview data was coded and summarized to identify the key words or phrases that were problematic for the students. Interview questions one and two were also scored using a coding rubric (see Table 5). Since these students had already taken the current version of the CINS test prior to the interview, they were familiar with the test. Asking these two questions provided insight into whether or not the students changed their original answers and allowed the researcher to see their rationale for choosing each of their answers.

The answers to the other interview questions were organized into a chart format to summarize the feedback and suggestions given by each interviewee. For questions five and six, a list was created to show which words and phrases were confusing or problematic for students, along with their recommendations for improving them. These two questions, five and six, facilitated the construction of better items for the revised version of the CINS.

Table 5
Coding rubric for interview questions 1 and 2. Rubric was based on the research conducted by Anderson, Fisher, and Norman (2002).

Score	**Nature of response**
+2	Response is a clear statement, showing the use of scientific conception.
+1	Response shows the student has understanding of the concept, but is unable to verbalize it in the form of a clear statement.
0	Response is not fully conceptualized, and needs to be prompted by the interviewer.
-1	Response is unclear and shows no understanding of the concept.

Analysis of Quantitative data, Part Two, Stage four. Comparing the results from the fall biology class with the newer data from the spring showed whether or not the changes that were made were beneficial. The qualitative data from the interviews supported the need for changes to the questions. The quantitative data obtained from the revised version distributed in the spring provided information as to whether those changes made were beneficial in assessing students' understanding of natural selection.

Following the data collection in Stage 4, the results from the two versions of the CINS test were compared, using a *t*-test (independent samples). In addition, a factor analysis was conducted to see if the two items for each concept being tested were more reliable in terms of measuring understanding. Additionally the factor analysis helped determine whether or not there was less variability in the answers between the revised version and original version.

Results

Quantitative Data, Part One, Stage one.

All of the data that was collected during this study helped answer both of the research questions. The results will be divided based on the section that the data was collected in, either quantitative or qualitative. The quantitative data from non-majors using both the original and revised version of the CINS, addressed question two, and both the *t*-test and factor analysis results indicated that the improvements that were made to the CINS yielded statistically significant data sets.

Two lab sections of a non-major's biology class (n=45) took the original, version of the CINS in its entirety as a post-test, to conclude the evolution unit that was being taught in the class. Figure 1 (below) represents the distribution of the scores and shows that there was a large range, between 4 (20%) and 20 (100%), on how the students performed, with majority of students scoring between 55 and 65 percent. The median value was 13.50, with a mean of 12.93. The standard deviation for the class was +/- 3.39, which is relatively high, compared to the other values (Table 9). Table 9 indicates a great diversity in scores among all of the classes that participated in this study. Table 9 is located towards the end of the results section because it incorporates all of the data for each of the different classes that participated in this study.

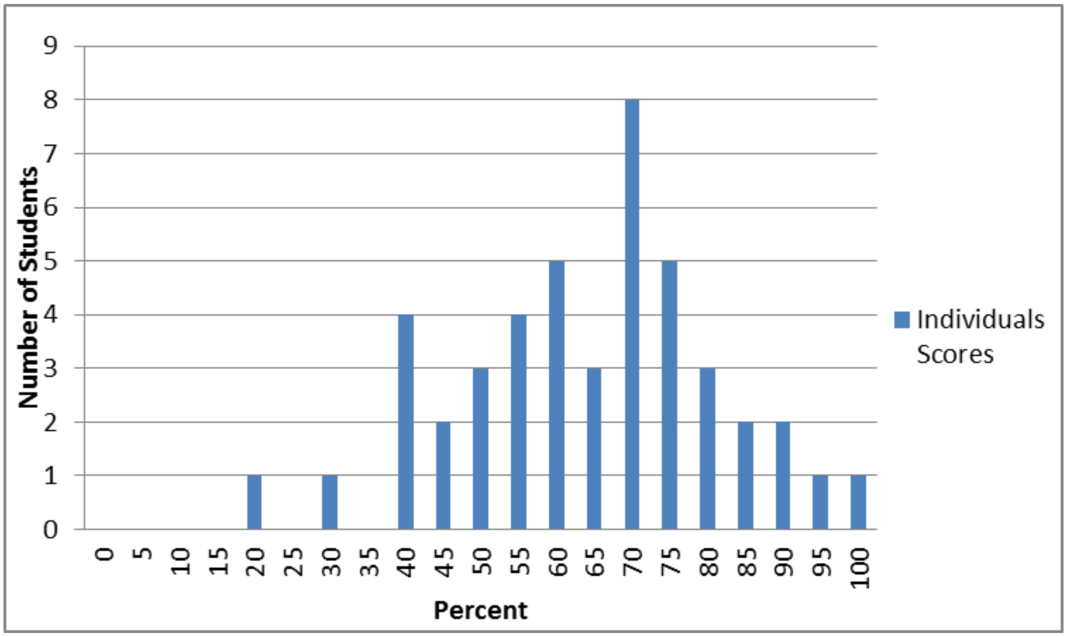


Figure 1. Score distribution for the non-major's class

Kuder-Richardson Formula 20 (KR20) is a reliability coefficient that measures the internal consistency for reliability for dichotomous data sets. It focuses on how the first half of the test matches the second half of the test, which is beneficial for this study, since it is Anderson and Fisher's desire to use the first ten CINS items as a pre-test and the last ten as a post-test. The reliability coefficient of .71 suggests good reliability between the first half (items 1-10) and the second half (items 11-20) of the CINS (Table 9).

Table 6 describes how the class scored for the 20 CINS items and indicates how many students chose the correct answer, or other distractors, for each item. Items 2, 3, 7, 8, 11, and 15 were items that had at least one distractor that was not chosen by the students. With all of the other items on the CINS, all of the distractors were chosen by at least one of the students. A closer look at the 12 focus items for this study reveals that certain item pairs, which are testing the same concept, show differences between how the upper and lower 27 percent answered, and how many distractors were chosen.

Table 6
Class data for non-majors, existing CINS, Fall 2010, standard item analysis

Concept	No.	Correct Group Responses			Correct	Response Frequencies -*Indicates correct answer					Non
		Total	Upper 27%	Lower 27%	Answer	A	B	C	D	E	Distractor
Biotic Potential	1	86.36%	100.00%	75.00%	C	1	2	*38	3	0	E
	11	72.73%	83.33%	66.67%	B	2	*32	0	10	0	CE
Population Stability	2	97.73%	100.00%	91.67%	A	*43	1	0	0	0	CDE
	14	63.64%	83.33%	41.67%	D	3	12	1	*28	0	E
Resources Limited	3	93.18%	100.00%	75.00%	B	2	*41	1	0	0	DE
	12	72.73%	83.33%	41.67%	A	*32	2	2	8	0	E
Limited Survival	5	84.09%	83.33%	58.33%	D	1	2	4	*37	0	E
	15	75.00%	83.33%	33.33%	D	0	5	6	*33	0	AE
Change in Population	4	34.09%	58.33%	25.00%	B	9	*15	5	15	0	E
	13	45.45%	91.67%	16.67%	B	3	*20	7	14	0	E
Origin of Variation	6	27.27%	83.33%	0.00%	B	19	*12	5	8	0	E
	19	68.18%	91.67%	41.67%	C	6	2	*30	6	0	E
Variation Inherited	7	90.91%	100.00%	75.00%	C	2	2	*40	0	0	DE
	17	56.82%	91.67%	41.67%	D	1	8	10	*25	0	E
Origin of Species	8	59.09%	75.00%	33.33%	A	*26	0	10	8	0	BE
	20	59.09%	83.33%	33.33%	B	6	*26	2	10	0	E
Variation	9	52.27%	75.00%	50.00%	D	4	12	5	*23	0	E
	16	50.00%	75.00%	25.00%	C	8	13	*22	1	0	E
Differential Survival	10	61.36%	75.00%	25.00%	C	10	2	*27	5	0	E
	18	43.18%	66.67%	16.67%	B	7	*19	4	14	0	E

This table also shows the difficulty (% correct) for upper and lower 27 percent of the students for each of the CINS items, depicting how the higher and lower-performing students answered each item. Items 1, 2, 3, and 7 on the chart indicate that most students in both groups are choosing the correct distractor for these items. With the items 5, 8, 10, 12, 13, 14, 15, 16, 17, 19, and 20, there was approximately a 50 percent difference between how the upper 27 percent answered, in comparison to the lower 27 percent. The other items, 4 and 18, had lower numbers for the upper 27 percent, and for item 6, none of the lower 27 percent of students even chose the correct answer.

Out of the 12 items, item 7 was the only one to have a higher percentage of students answering that question correctly. This means that the item must have been easy for the students to comprehend and answer. The other 11 items showed a percentage difference of 25 to 75, between the upper and lower 27 percent. With the lower 27 percentage students, guessing is a possibility due to the students' lack of understanding and knowledge on the different concepts of natural selection. The possibility that the students might guess would be 25%, since there are four answer choices for each item. Items 13 and 18 were items shown to be harder for those students who were on the lower end. About 16.67 percent of the lower 27 percent answered this question correctly. For these two items, guessing could have occurred, which would explain why there was a low percentage of students answering that item correctly, but with only 16.7% of the students choosing the correct answer, it is more likely that students intentionally chose an alternative conception answer.

Another interesting observation from Table 6, was item 6. Item 6 had 0.00% of the lower students (27%) choosing the correct answer. The students were in fact choosing the wrong answer, which indicates that random guessing was not even an option for this item. None of those students chose the correct answer, which could imply that this item is difficult for students who hold any conception or understanding. Item 18 was difficult for all of the students because the average for that question was low for both the higher and lower 27 percent of the students.

The other 8 items that were not focused on had concepts that seemed to be more easily understood by all of the students (Table 6). Items 1, 2, and 3 had high scores, and items 2 and 3 had either two to three distractors that were not chosen. Items 11 and 15 also had distractors that were not chosen. For the lower students items 12 and 15 seemed to be more difficult for them to answer, with both having scores of 33.33%. This may also indicate that guessing and/or lack of understanding of the concept occured.

For the item pairs on the six focus concepts, it was revealed that items that test the same concept were answered differently. For example, both items 6 and 19 assess, the concept of origin of variation, but item 6 seems to be the harder of the pair. For the concept of variation inherited, students answered item 7 more accurately than they did item 17. In contrast, the concepts of differential survial (10 and 18), origin of species (8 and 20), and change in population (4 and 13), had lower scores for each of the items meaning that certain answer choices were being picked based on how the student was understanding the item and their conception on the topic.

Qualitative Data

During the interviews, students respond to each of the 12 items being investigated. For the six concepts, Figure 2 represents the two items on each concept that were answered incorrectly by each of the students. For each item pair, it can be seen that one item was answered incorrectly more often than the other item in its pair. The origin of species concept is an example of this issue. Out of 18 students, 17 answered item 8 correctly and only 11 answered item 20 correctly. Student 18 was the only individual who missed both of those items, 8 and 20 (Table 7). The 10 other students who answered item 20 incorrectly did not miss item 8 when they were verbally stating their reasoning for the answer. Another indicator was revealed in the factor analysis in which items 8 and 20 did not align within the same factor, indicating that these two items did not seem to test the same concept. Variation was a concept that was missed by a significant number of the students who were interviewed. Item 9 had the most wrong answers. Three students answered both items 9 and 16 incorrectly, as shown in Table 7. The remaining students who missed these items on variation only missed one out of the two.

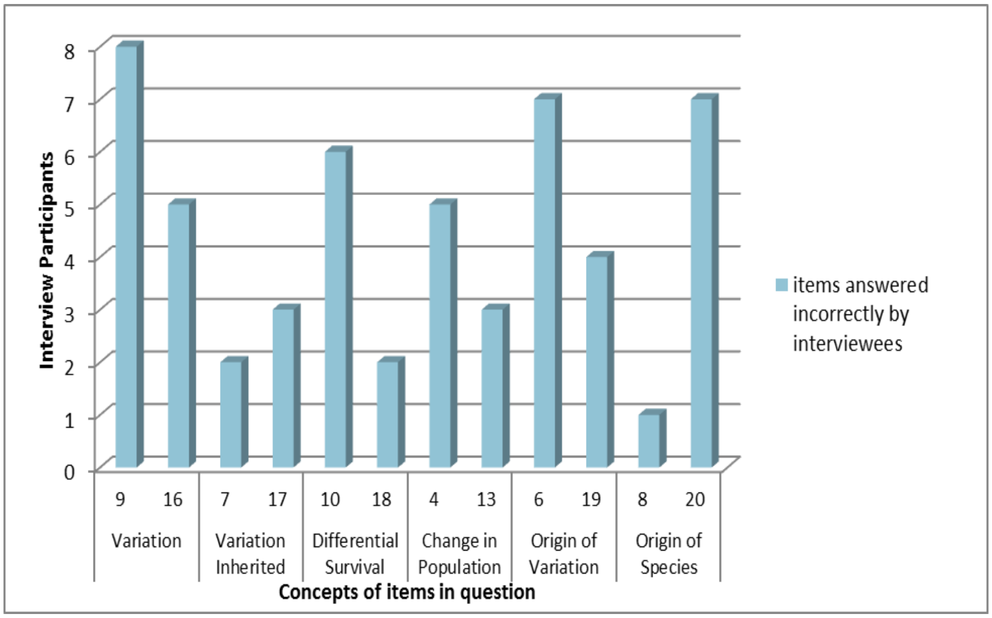


Figure 2. Items answered incorrectly in the interviews, grouped by concept.

Table 7, shown below, represents how each of the 18 interviewees performed on the 12 items of the original CINS, both in the class and during the interview data. It shows which items each student answered incorrectly the first time the test was taken in the class, as well as when they were interviewed and asked to verbally explain their answers. There are some differences between the answers that the students chose when they were in class compared to when they gave their answers in the interview. Table 7 shows that the majority of the students who were interviewed scored higher during the interview than when they originally took the CINS in class. Twelve out of the eighteen students interviewed chose correct answers on more of the 12 items during the interview than when they took the CINS in the class. The other five students either answered the same number of items incorrectly or answered more of the 12 items wrong (Table 7).

Table 7
Comparison of items answered correctly for CINS taken in the class with the items answered correctly during the interviews.

		Change in Population		Origin of Variation		Variation Inherited		Origin of Species		Variation		Differential Survival	
Student		4	13	6	19	7	17	8	20	9	16	10	18
1	C	x	x	x	x	x	x	x	x	x	x		x
	I	x	x			x	x	x	x	x	x	x	x
2	C	x				x				x		x	
	I	x	x			x		x		x	x	x	x
3	C		x	x	x	x	x		x	x		x	x
	I	x	x	x	x	x	x	x	x			x	x
4	C				x	x		x		x			
	I		x		x	x		x	x				x
5	C	x		x	x	x	x	x	x			x	x
	I	x	x	x	x	x	x	x	x		x	x	x
6	C	x	x		x	x	x			x	x		
	I	x			x	x	x	x		x	x		
7	C		x			x			x	x		x	x
	I	x	x	x		x	x	x	x	x	x		x
8	C		x	x	x	x	x		x	x	x		
	I		x	x	x	x	x	x	x	x	x		x
9	C	x			x	x			x			x	x
	I	x	x	x	x	x	x	x	x	x		x	x
10	C	x	x		x	x		x			x	x	x
	I	x	x		x	x	x	x		x	x	x	x
11	C				x			x		x	x	x	
	I							x		x	x	x	x
12	C		x		x	x	x	x	x				x
	I		x	x	x	x	x	x	x		x		x
13	C	x	x	x	x	x	x	x	x	x	x	x	x
	I	x	x	x	x	x	x	x	x		x	x	x
14	C					x	x	x	x		x	x	
	I	x	x	x	x	x	x	x	x			x	x
15	C				x	x	x		x			x	
	I	x	x	x	x	x	x	x			x	x	x
16	C		x		x	x	x	x		x	x	x	x
	I	x	x		x	x	x	x		x	x	x	x
17	C	x	x		x	x		x	x			x	
	I	x	x	x	x	x	x	x	x	x	x	x	x
18	C	x	x		x				x	x	x		
	I						x			x	x		

C=CINS taken in class I=CINS items answered during the interviews X= correctly answered
Shaded boxes= students whoanswered only one item correctly.

Results for differential survival, change in population and the origin of variation also revealed some differences between how many students answered the items incorrectly. As shown in Table 7, for the concept of differential survival, only two students answered both items 10 and 18 incorrectly. The other four students answered only item 10 incorrectly. For the concept of change in population, items 4 and 13, five students answered item 4 incorrectly, while two of those five students answered item 13 incorrectly, as well. Figure 2 illustrates that 11 students missed this concept. Three out of the four students who missed item 9 also missed item 6. Variation inherited was the concept that was least misunderstood, with only five students answering items 7 and 17 incorrectly (Figure 2). One of those students missed both of the items within the concept, which is seen in Table 7. The shading in Table 7 represents the students who answer one of the items out of the pair correctly for each concept.

Based on Figure 3 and Table 7, it appears that the majority of students chose more of the correct answers during the interviews than when taking the CINS in class. The red bars in Figure 3 depicts how each student answered during his or her interview. Students 2, 3, 4, 5, 7, 8, 9, 10, 12. 14, 15, 16, and 17 all achieved a higher percentage on the interview than on their in-class performance. The other students, 1, 13, and 18, all earned lower percentages than they had in class. Only two students, 6 and 11, had no change in the percentage they earned in answering the 12 items in either setting, with Student 6 answering 7 correctly and Student 11 correctly answering 5.

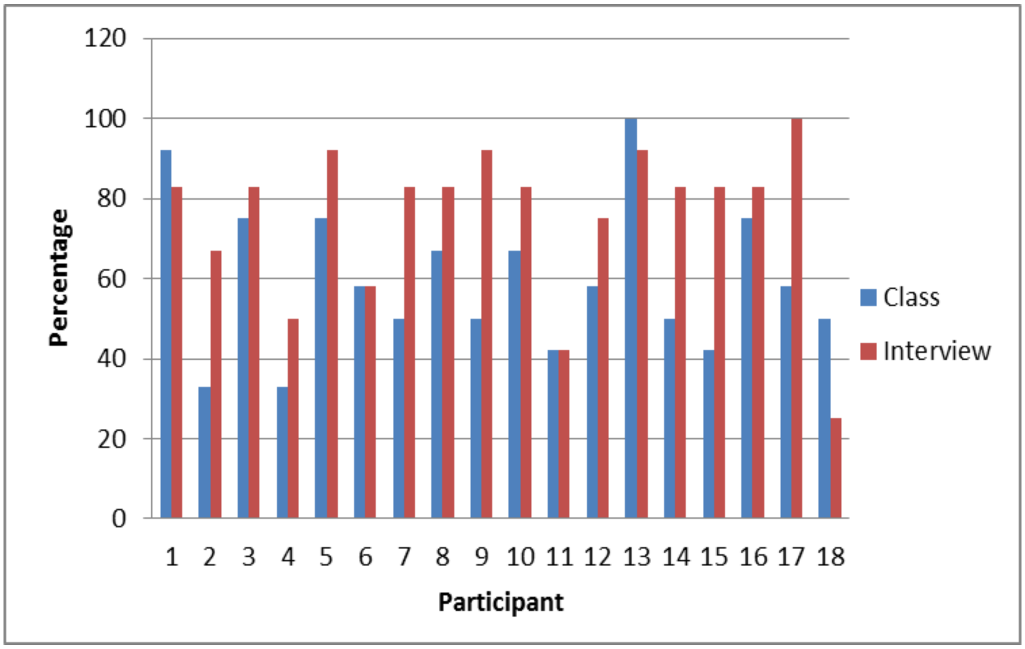


Figure 3. A comparison of the 12 CINS items that were answered correctly by the 18 students who were interviewed, data from both class and interview.

The students' scores on the 12 items during the interviews ranged from 3 to 12, with the average score of 9.05. The percentages of how well the students performed between the CINS taken in class and the 12 items that were answered during the interviews also differed. This data is based on how many of those 12 items were answered correctly, for both the class and the interviews, by the 18 students. The average percentage achieved for the class was 59.72%, while 75.39% was the average percentage for the interviews. This represents a significant difference $t(70)=1.99$, $p < .05$ in students' performance, dependent upon the setting, see Table 11.

Using the coding rubric from Table 5, numbers were assigned to each of the students' responses to interview questions one and two. The scores each student obtained for questions one and two were averaged, as shown in Table 8. The score range for all 18 interviewees was .375 to 1.833, with a mean of 1.278. The majority of the students received a score of 0 or higher, while only three individuals received a -1 (Table 8). This table, Table 8, shows each students' score based on their responses during the interview. If a student received a 2, it was due to his or her correct use of terminology and application of conceptual knowledge in giving examples. A score of 1 would mean that the student was on the right path, but did not articulate or give reasons why that option was chosen and why it was or was not correct. Students who received 0's or -1's did not give an accurate reason for their answers. These students also had to be prompted or guided with either a question or a statement by the interviewer. Below is an example of how the students responses were coded, based on the rubric in Table 5, with an excerpt from student 2 on item 7.

(Item 7)

7. What type of variation in finches is passed to the offspring?
   a. Any behaviors that were learned during a finch's lifetime
   b. Only characteristics that were beneficial during a finch's lifetime
   c. All characteristics that were genetically determined
   d. Any characteristics that were positively influenced by the environment during a finch's lifetime.

Student 2: I would of chose [*sic*] C and here is why.

With A it doesn't make sense because it talks about being learned. All right, you have to learn and that is personal and is not passed on genetically.

B [*reading the answer*] that is like saying that I am only going to pass on blonde hair and not going to give my children the height gene. That is like saying I am only going to give them parts of who I am.

C [*reading the answer*] makes sense, because you are passing off the same genetic material to your children and you are not changing anything.

D says any characteristics that are positively influenced, again that is saying that you are going to give the good genes to my kids, which is impossible to do, so that is why I would choose C.

For item 7, student 2 received 2's for his reponses for interview questions one and two. He was able to articulate why his answer choice C was correct, and why the other answer choices were not by going through each one and stating why. Student 2 clearly understood that genetic material is the only way variation can be passed on to the offspring, which is why he received a coding score of 2. His reponses were clear and understandable by giving examples showing the use of scientific conception. More examples are discussed later.

The codes assigned to each participating student's responses were averaged for both questions one and two to provide a clear picture as to how well each student was understanding the six concepts related to natural selection. According to Table 8, some students' averages between the two questions were very similar, while others' were quite different. A couple of reasons might be due to a lack of knowledge. Another reason may be that the student was a good test taker and was able to narrow down the answer choices but not know why the answer is correct, or the student has some understanding, but is unable to articulate his or her reasoning. The lowest recorded average for the two questions was for Student 11, who scored a .3333 on question one and .4167 on question two, respectively. The highest average for question one was 1.75 for Student 13, who also earned the highest score on question two, 1.9167. The lowest average score for question two was .3333, from Student 10. These averages for both questions one and two indicate the degree to which students are actually understanding the concept, as opposed to guessing on the 12 items, based on the verbal feedback that was given during the interviews. The verbal feedback that was given by the students showed whether or not the student understood what the question was asking. The students who had no idea as to what the answer was received lower scores because they were not able to articulate their reasoning, see Table 8.

Table 8
Individuals' coded results for the 12 items based on questions 1 and 2, and comparison of their overall CINS score.

Participants	**Questions**	**12 item score**	**Average of 12 items**	**Average of Questions 1 and 2**	**CINS 12 Item Score**
1	1	2, 2, 2, 2, 0, 1, 2, 0, 2, 2, 2, 2	1.583	1.583	10
	2	1, 2, 2, 2, 1, 1, 2, 1, 2, 1, 2, 2	1.583		
2	1	2, 0, 2, 2, 2, 1, 2, 1, 1, 2, 0, 0	1.25	1.416	8
	2	2, 0, 2, 2, 2, 1, 2, 1, 1, 1, 0, 0	1.167		
3	1	2, 2, 2, 1, 0, 1, 1, 0, 2, 2, 2, 2	1.417	1.459	10
	2	2, 1, 2, 2, 0, 1, 2, 1, 2, 2, 1, 2	1.5		
4	1	0, 0, 1, 2, 0, 1, 2, 0, 1, 1, 1, 0	.75	.584	6
	2	1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0	.417		
5	1	1, 2, 2, 2, 1, 2, 1, 1, 2, 2, 2, 1	1.583	1.583	11
	2	2, 1, 2, 1, 2, 1, 2, 1, 2, 2, 1, 2	1.583		
6	1	1, 2, 2, 2, 1, 1, 0, 2, 1, 1, 1, 0	1.167	1.209	7
	2	1, 2, 2, 2, 1, 1, 1, 2, 1, 0, 1, 1	1.25		
7	1	2, 2, 1, 1, 1, 1, 1, 1, 2, 1, 1, 2	1.333	1.292	10
	2	1, 1, 1, 1, 2, 1, 1, 2, 2, 1, 1, 1	1.25		
8	1	0, 2, 2, 1, 2, 1, 2, 1, 2, 1, 1, 1	1.333	1.292	10
	2	1, 1, 2, 2, 2, 0, 2, 2, 2, 1, 1, 2	1.5		
9	1	2, 2, 1, 1, 1, 1, 2, 0, 2, 2, 1, 1	1.333	1.25	11
	2	2, 1, 1, 1, 1, 1, 2, 1, 2, 1, 1, 0	1.167		
10	1	1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0	.5833	.4583	10
	2	0, -1, 1, 1, 0, 1, 1, -1, 1, 0, 1, 0	.3333		
11	1	-1, -1, 0, 1, 1, 2, 0, 1, 0, 2, -1, 0	.3333	.375	5
	2	0, 0, -1, 0, 1, 2, 0, 1, 0, 2, 0, 0	.4167		
12	1	1, 1, 1, 2, 1, 1, 2, 2, 2, 1, 2, 1	1.4167	1.21	9
	2	1, 0, 1, 1, 1, 0, 1, 1, 2, 1, 2, 1	1		
13	1	1, 2, 2, 2, 1, 2, 2, 2, 2, 2, 1, 2	1.75	1.833	11
	2	2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2	1.9167		
14	1	2, 2, 2, 2, 0, 2, 1, 0, 2, 2, 1, 1	1.4167	1.583	10
	2	2, 2, 2, 2, 1, 2, 2, 1, 2, 2, 1, 2	1.75		
15	1	1, 2, 2, 2, 1, 2, 2, 1, 2, 2, 2, 1	1.6667	1.5834	10
	2	2, 1, 2, 2, 1, 1, 2, 1, 2, 1, 2, 1	1.5		
16	1	1, 0, 2, 1, 1, 2, 1, 2, 2, 2, 1, 1	1.3333	1.2082	10
	2	0, 0, 1, 0, 1, 1, 2, 2, 2, 1, 2, 1	1.0833		
17	1	2, 2, 1, 2, 1, 2, 2, 1, 1, 2, 2, 2	1.6667	1.625	12
	2	2, 1, 1, 1, 2, 2, 2, 1, 1, 2, 2, 2	1.5833		
18	1	-1, 0, -1, 0, 2, 0, 1, 1, 2, 0, -1, -1	.1667	.375	3
	2	0, 0, 0, 0, 2, 1, 1, 1, 2, 1, -1, 0	.5833		

During the interviews, each student gave some type of reponse for each of the 12 items in question, their feed back on each of the 12 items is shown in Appendix D. The only items that none of the students seemed to have an problem with were items 16, 17, and 20, as none of the students expressed any concern or desire to change or modify these three items. The suggestions for the other nine items allowed for insight into how well they actually understood what each question was asking. Below is an example of the type of feed back that was given by the students for item 4.

Words or phrases that were heard/noticed/stated by the students:

1. Primary Changes
2. Proportions, which seemed to be new to some of the interviewees.

Possible solutions stated by students:

1. What type of primary changes are you referring to? Morphological, ability, behavioral, etc. Define/state what you mean by that specifically, explain, and expand. Also is it each finch or each species of finch? Maybe say ‘Main Changes’
2. Maybe use another word. The interviewees who had concern could not think of another word to replace it.

The correlation between how students answered during the interview, as compared to how well they answered the CINS items during a class setting are shown in Figure 4. There is a significant positive correlation ($r = 0.5848$, $p < 0.01$). The students who scored higher on the CINS, also performed well on the interview, while lower scores on the CINS correlated with lower interview scores.

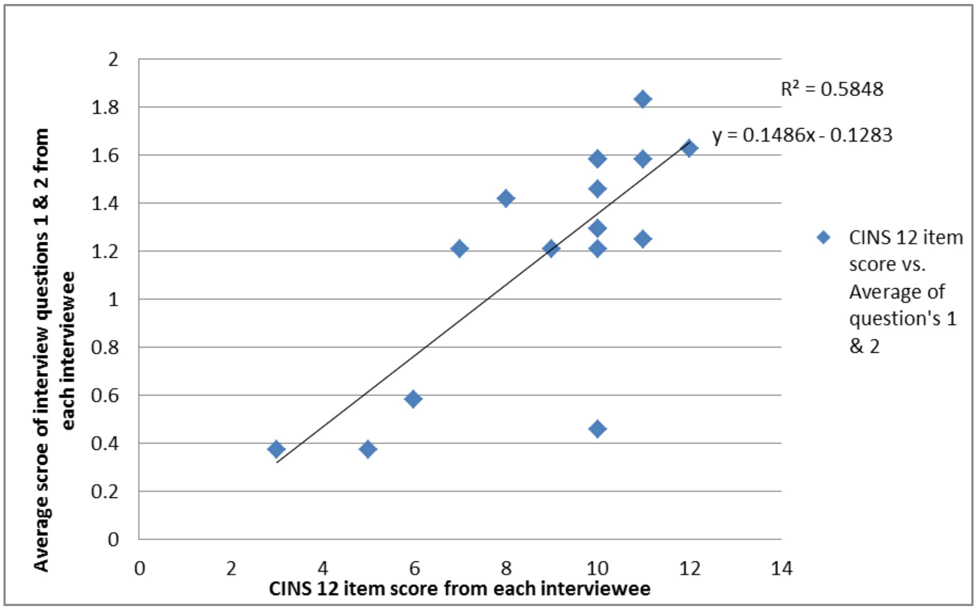


Figure 4. Correlation between the 12-item score with the average coded score for each student interviewed.

Information gathered from the interviewees is summarized in Appendix D. This table represents all of the changes that were suggested on the 12 items, based on the student feedback. The changes that were made to the 12 items in question are located in Appendix E. Certain words that were indicated to be "troublesome" according to the students were changed. For example, for the item pair 4 and 13, the word 'proportions' was changed to 'percentages' because it was found throughout the interviews that 'proportions' was confusing to students. That example, for item 4, can be seen below which shows the original item along with the changes made.

Original CINS Question:

4. In the finch population, what are the primary changes that occur gradually overtime?
   a. The traits of each finch within a population gradually change.
   b. The proportions of finches having different traits within a population change.
   c. Successful behaviors learned by finches are passed onto offspring.
   d. Mutations occur to meet the needs of the finches as the environment changes.

Revised CINS Question:

4. In the finch population, what are the ~~primary~~ changes that occur gradually overtime?
    a. The traits of each finch within a population gradually change.
    b. The percentages of finches having different traits within a population change.
    c. Successful behaviors learned by finches are passed onto offspring.
    d. Mutations occur to meet the needs of the finches as the environment changes.

Also, certain sentence structures for some of the items were changed to improve the grammar. The 8 items not included in the interviews were minimally altered (Appendix F). In both Appendices E and F, the original item is located on the left side, and all of the changes that were made are on the right.

Quantitative Data, Part Two

The last stage of the data collection consisted of comparing the performance of both majors and non-majors on the two CINS versions. The revised version of the CINS that was tested can be found in Appendix G. This data addresses research question 2 of this study to see whether or not the two item pairs are equivalent in terms of how the students answers them, and if a correct answer on only one item of a particular pair indicates that the student does not fully understand the concept. Figure 5 represents the scores for 32 junior and senior biology majors who took the revised version of the CINS. The median value was 18, with a mean of 18.35. The majority of the junior and senior students performed well, as evidenced by their class average of 86%. It can also be seen from Figure 5 that none of the students scored below 50%.

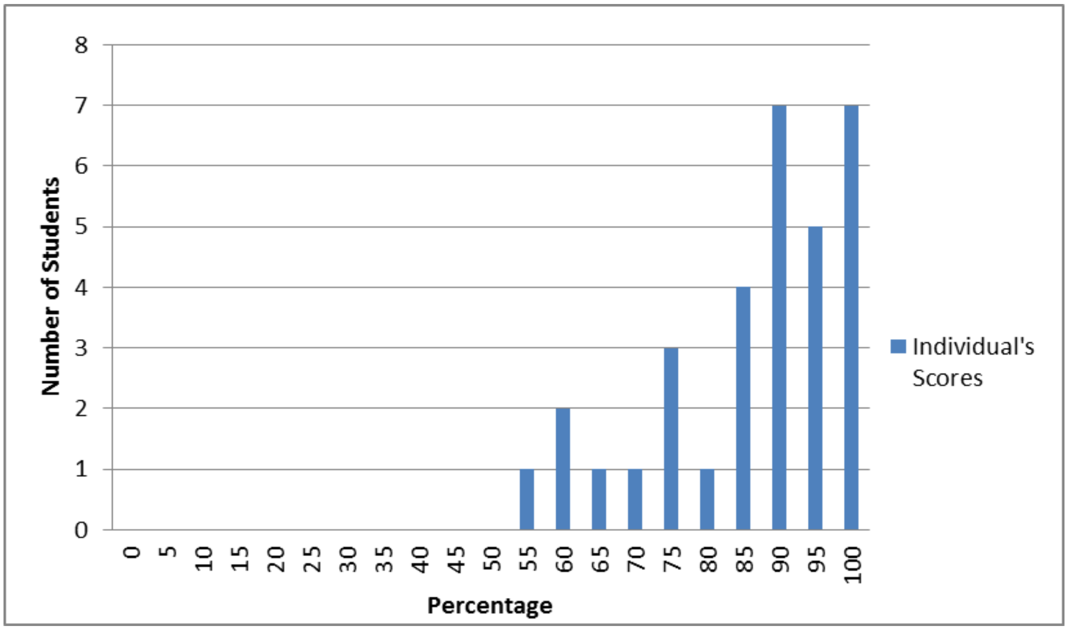


Figure 5. Score distribution for the junior and senior major's on the revised CINS.

Table 9 summarizes the descriptive statistics for all of the class data. Table 9 also represents the standard deviations for all of the classes who participated in this study, along with the mean, median, class range, reliability coefficient, and number of students. It can be seen that the higher scores were obtained by the majors. This can be explained by the increased number of biology classes that the junior and seniors have taken, as compared to the non-majors. The reliability coefficient for the junior/senior majors is .71, which means that the revised CINS is reliable between the two halves (i.e. items 1-10 and items 11-20) (Table 9).

Table 9
Student distribution of scores, original vs. revised version.

Class:	Non-Majors		Freshman Majors		Junior/Senior Majors
CINS Version:	Original Version	Revised Version	Original Version	Revised Version	Revised Version
Highest Score:	20	20	20	20	20
Lowest Score:	4	2	14	10	11
Mean:	12.93	14.32	18.35	17.67	17.22
Median:	13.5	15	19	18	17.83
Standard Deviation:	3.39	4.25	1.82	2.8	2.57
Reliability Coefficient (KR20):	0.71	0.85	0.64	0.78	0.71
Number of Students	44	38	20	13	32

The student who took the revised version did score significantly higher than those who took the original. This could indicate that the non-majors did poorly on the original CINS due to a lack of understanding that the student had, along with the conceptions that they held. There also seemed to be more variability among the students who took the revised version. The standard deviaions alone show how spread out each class answered the items, which is an indication that knowledge and understanding is involved.

In addition to the data collected from the upperclassmen, data with the new CINS was also collected in a freshman majors' course. Of the 13 freshmen who took the revised version, 11 scored at or above 80%. As shown in Figure 6, like the upper-classmen, this freshmen class scored higher than the non-majors for both versions of the CINS, (also see Table 9). The median value was 18, with a mean of 17.67. The standard deviation was +/- 2.80, Table 9. The reliablity coefficient for this group was a .78, which is .07 higher than that for both the non-majors' and the majors' who took the original version, as well for as the upper-classmen who took the revised version (Figure 6). Even though this freshman class was smaller in size (n=13), compared to the other junior and senior majors (n=32), the mean, median, standard deviation, and reliability coefficient were similar, as shown in Figures 5 and 6, along with Table 9.

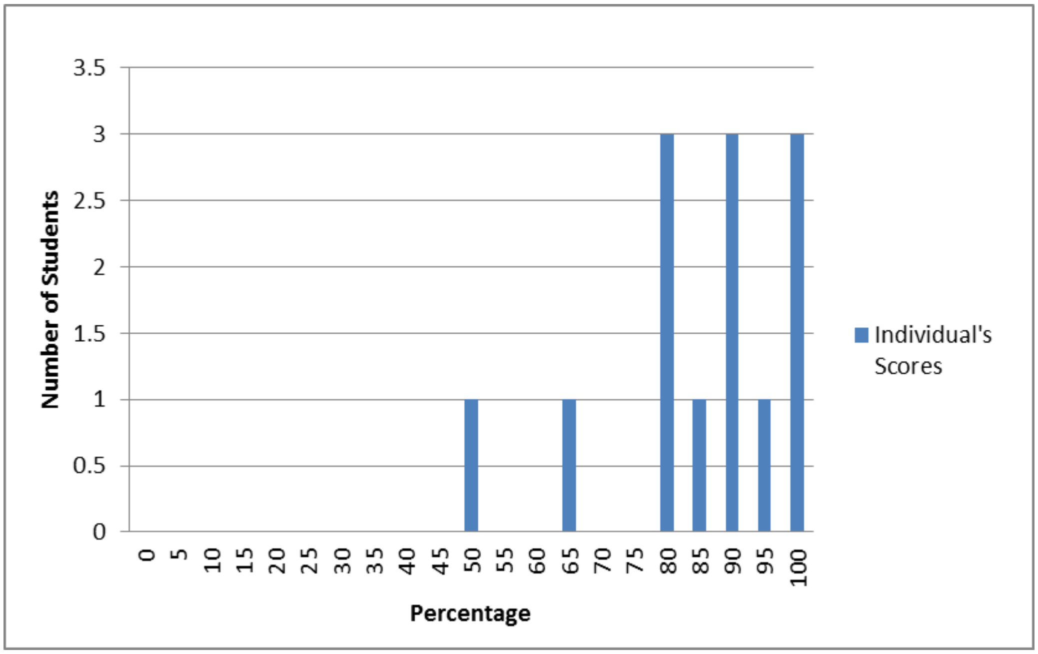


Figure 6. Score distribution for the freshman majors' class on the on revised CINS

As described in the methods section, a group of freshman majors took the original version of the CINS. All 20 students performed well on this version, and as expected, higher than the non-majors who took the original version in the fall, (see Figure 7 and Table 9). The median value was 19, with a mean of 18.35. None of the students received below a 70 percent (14/20), and more than half of the students scored either a 95% or 100%. Comparing Figure 7 with Figure 1 reveals that there is a large difference in scores between non-majors and majors, which was expected because of the difference in the number of biology classes taken by each group.

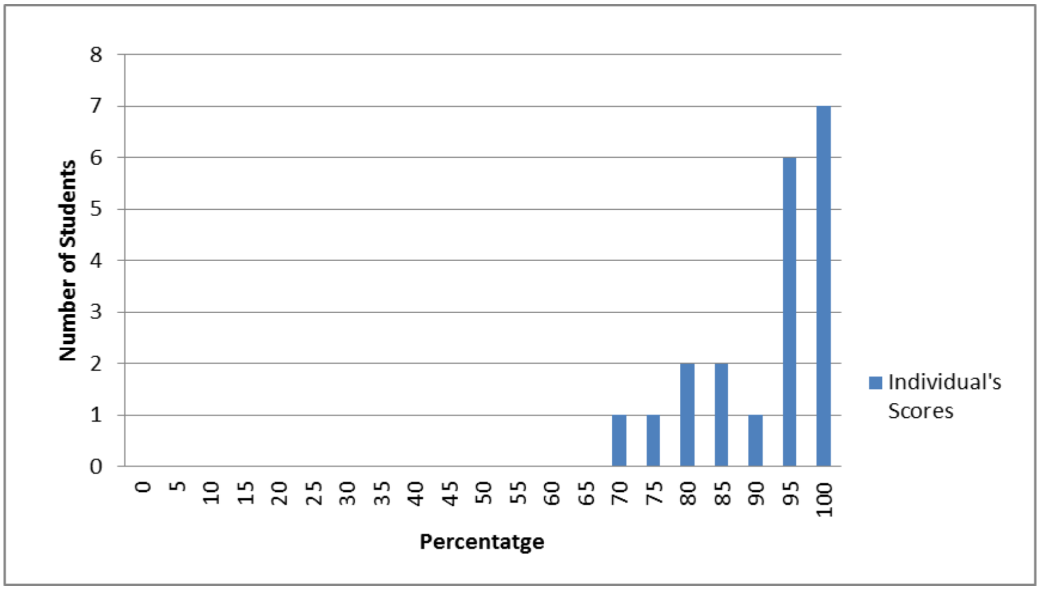


Figure 7. Score distribution for the freshman major's class on the original CINS

The standard deviation for the freshman majors class was +/- 1.82, which was lower than that for the other majors, who were upper classmen. This could be due to the fact that the freshmen, who took the original CINS, had just finished the unit on evolution and had exam questions that were similar to questions to the original CINS. Additionally, the reliability coefficient for this freshman class was lower than that of the other majors' class, Table 9, and indicates that there is a difference between the freshman majors who took the original version and the majors who took the revised version. The reliability coefficient for the freshman majors are shown in Table 9.

Figure 8, (below) shows the score distribution for the non-majors who took the revised version of the CINS. The distribution for this class seems to be bimodal, with half the class scoring in the mid 50's to low 70's, and the other half scoring 80% and above in Figure 8 and in Table 9. The range of the scores was from 2 (10%) to 20 (100%). The median value was 15, with a mean of 14.32. When Figure 1 and Figure 8 are compared, it is clear that the non-majors' class that took the revised version (Figure 8) had higher scores overall. The standard deviation for the non-majors class was +/- 4.25, Table 9. The reliability coefficient shows improvement for those who took the revised CINS.

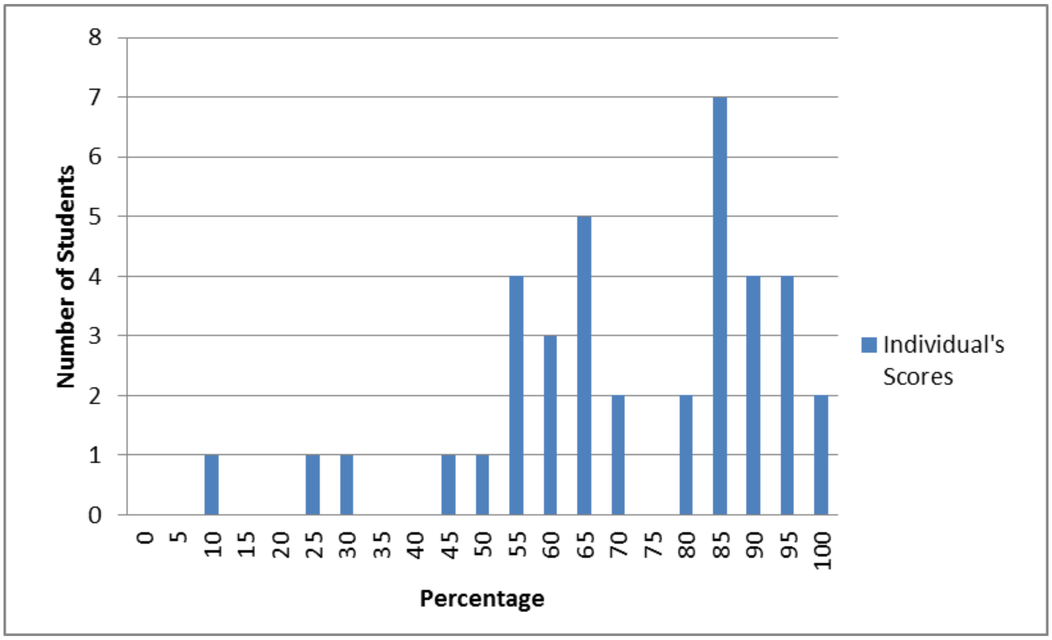


Figure 8. Score distribution for the non-majors' class on the revised CINS

Comparison of non-majors classes between the original and revised version

For this study, a group of non-majors took the original version of the CINS in its entirety. Three classes of biology majors participated in taking either the revised version (freshmen and junior/seniors) or the original version (freshmen). Lastly, a group of 43 non-majors took the revised version of the CINS. To compare the two non-majors' classes with one another in terms of specific student answer choices Table 10 and Table 6 can be compared. A difference is seen between how many distractors were chosen for each item on the revised version, with more students choosing the correct answer. For the group of non-majors who took the revised version, items 5, 10, 14, 16, and 17, all had distractors that were not chosen, along with 2, 3, 7, 8, 11, and 15 (Table 6), which were items that were not focused on in the study. Interestingly, when comparing Tables 6 and

10, item 7 had two distractors not chosen for the revised version, while the data from the original had only one not chosen, and that distractor was not even chosen for the revised version. This is interesting because the only changes that were made to the actual answer choices for item 7 was changing the word ‘characteristics’ to ‘traits’, which seemed to have an impact in how the students, taking the revised version, answered item 7.

The difference between the upper and lower (27%) performing students, as well as the percentage of students answering the 20 items correctly, increased on the revised version of the CINS. There were still some students performing lower on items, but the majority increased, except for item 8, which went from 59.09% to 47.37%. Item 6 still had zero percentage of lower-performing students answering this item correctly, which indicates that this item remains difficult for the lower-performing students. Moreover, the lower-performing non-majors tended to have a higher chance of answering the CINS items correctly, compared to the non-majors who took the original. Compared with Table 6, the results for item 18 for the revised version showed better results for the higher performing students. The difference between scores for the original and the revised version amounted to almost 40% for item 18.

Table 10

Class data for non-majors, revised version, Spring 2011, standard item analysis

Concept	No.	Correct Group Responses: Total	Upper 27%	Lower 27%	Correct Answer	Response Frequencies -*Indicates correct answer: A	B	C	D	E	Non Distractor
Biotic Potential	1	86.84%	100.00%	70.00%	C	1	2	*33	2	0	E
	11	65.79%	80.00%	30.00%	B	0	*25	1	12	0	AE
Population Stability	2	92.11%	100.00%	80.00%	A	*35	2	1	0	0	DE
	14	60.53%	90.00%	50.00%	D	3	12	0	*23	0	CE
Resources Limited	3	89.47%	100.00%	70.00%	B	3	*34	1	0	0	DE
	12	76.32%	100.00%	40.00%	A	*29	3	4	2	0	E
Limited Survival	5	89.47%	100.00%	70.00%	D	0	3	1	*34	0	E
	15	81.58%	100.00%	60.00%	D	0	5	2	*31	0	AE
Change in Population	4	34.21%	60.00%	30.00%	B	12	*13	4	9	0	AE
	13	50.00%	100.00%	20.00%	B	5	*19	4	10	0	E
Origin of Variation	6	55.26%	100.00%	0.00%	B	5	*21	3	9	0	E
	19	71.05%	100.00%	30.00%	C	5	2	*27	4	0	E
Variation Inherited	7	89.37%	100.00%	90.00%	C	0	0	*37	1	0	ABE
	17	68.42%	100.00%	40.00%	D	0	7	5	*26	0	AE
Origin of Species	8	47.37%	90.00%	20.00%	A	*18	8	9	3	0	E
	20	68.42%	100.00%	40.00%	B	2	*26	2	8	0	E
Variation	9	60.53%	80.00%	20.00%	D	2	4	9	*23	0	E
	16	68.42%	80.00%	30.00%	C	6	6	*26	0	0	DE
Differential Survival	10	86.84%	100.00%	50.00%	C	4	1	*33	0	0	DE
	18	81.58%	100.00%	40.00%	B	1	*31	3	3	0	E

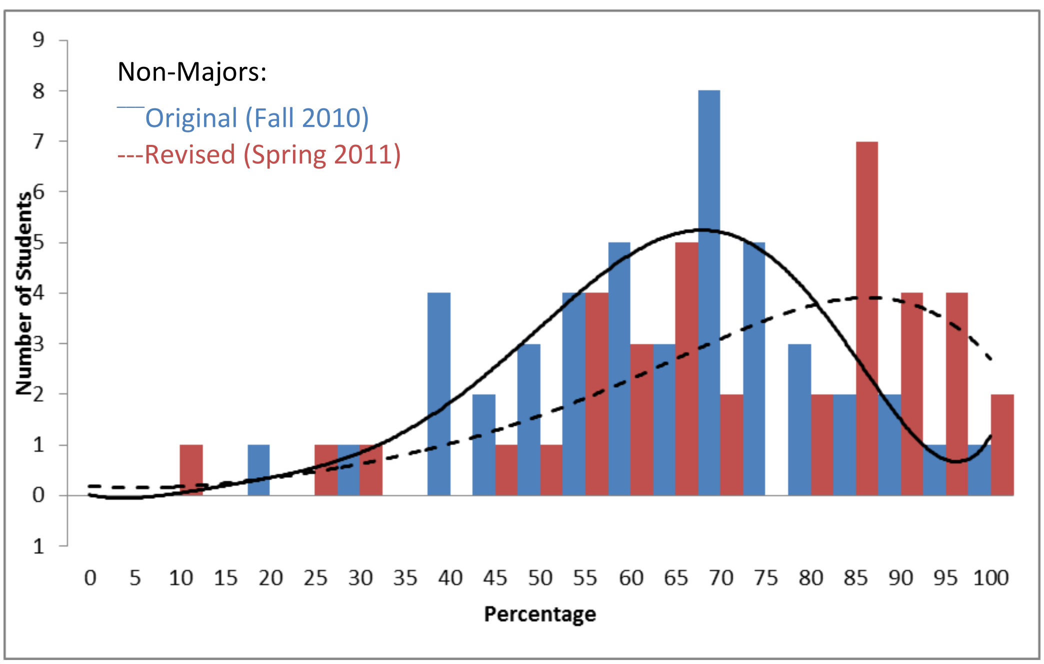


Figure 9. Histogram of non-majors scores on both the original version and the revised version, with trend lines.

Figure 9 (above), represents the overall class performance on the original and revised versions of the CINS by two groups of non-majors. The solid line represents the non-majors who took the original version, while the dotted line shows results on the revised version these are approximations of how the fall 2010 class of non-majors did compared to the spring 2011 class of non-majors. On the original test, the majority of student scores fell in the range between 50 and 80 percent, compared to the revised version's range of 70 to 95 percent. There was a shift in how well the students performed between the two versions of about 15 percent, which was attributed to the changes that were made to the 12 items.

An independent-samples *t*-test, assuming unequal variances, was conducted to compare the means of the two non-majors' classes that took the two different versions of the CINS. There was a significant difference (t (70)= 1.994, p=.002) in the scores that the students achieved, on the original version (Mean= 12.93, Standard Deviation= 3.39) and the revised version (Mean= 14.32, Standard Deviation= 4.25). Assuming that these two groups of students had the same level of understanding of natural selection, these results suggest that the changes made on the revised version allowed for the students to have a better understanding of what each item was asking relative to the concepts of natural selection.

A *t*-test was conducted to determine if the mean scores between the original and revised version to answer the first research question of whether the item improvements influenced overall student performance, by comparing the scores. To answer the second research question, a factor analysis was conducted for each set of data to see if the two items for each concept loaded together and align within the same component. The rotated component matrices that were conducted for each class are located in Appendix H. These depict the best alignment among the data, within a reduced set of factors that could explain the students' answers on the test. The items that showed strong correlations paired within the same component, which suggest that those items test the same concept. For all of the rotated matrices, the minimum correlation value that was reported was 0.40. The inverse values also show strong correlations as well. Individual items may pair with more than one component; these pairings may also indicate a strong relationship. Furthermore, some items might not associate with any of the components, which could indicate that the data could be explained using many components. These results are unlike those originally reported by Anderson, Fisher, & Norman (2002) in which nearly all of the item pairs loaded together. This could be due to the relatively low sample size of the present study.

Table 11 illustrates which of the 12 pairs loaded together from the non-majors and freshman majors who took the original version. The number of components that were extracted for this factor analysis was 7. The only item pair to load together was items 4 and 13, which fell under component 2. This is expected, because the group that has the lower scores would also have the lower and poorest pairings, which is observed with the one matching pair. All of the other items had one or more value in different components, which are the columns in Tables 11-13. Item 20 did not strongly pair with any of the components. This could mean that the performance on item 20 is random or that it is associated with many of the primary components.

Table 11
Condensed version of the rotated component matrix, for the original version, taken by non-majors and freshman majors.(n=64.)

CINS concept and item numbers	Component						
	1	2	3	4	5	6	7
Biotic Potential 1 11				0.73359			0.877317
Population Stability 2 14	0.760144				0.408479	0.577597	
Resources Limited 3 12	0.834513			0.673031			
Limited Survival 5 15			0.664036		-0.63998		
Variation 9 16		0.515311	0.553699			0.80103	
Variation Inherited 7 17	0.60566		0.76027				0.459993
Differential Survival 10 18	0.474479			0.40134			
Change in Population 4 13		0.842847 0.698866					
Origin of Variation 6 19		0.452743	0.420562		0.444145	0.437031	
Origin of Species 8 20		0.427912		0.528227	0.723737		

For the revised version, Table 12 represents the 8 components that were extracted from the data collected in freshman, and junior/senior majors' classes. There were four pairs that loaded together: 5 and 15, 4 and 13, 10 and 18, and, 6 and 19. Components 1, 2, 3, and 5 represented those item pairs, and half of the pairs that were being investigated, had high values >.73. It was also observed that the majority of the items had higher values than 0.40 within each of the components. None of the item pairs that loaded together failed to load within the same component.

Comparing this data to the data obtained from the non-majors and freshman majors, it can be seen that an improvement has been made with the number or pairs that loaded together. It was to be expected that more pairings would occur with the majors because of their increased exposure and knowledge derived from taking more than one biology class, so they should be more consistent in their answers. From Table 11, the concept of change in population item pairs loaded together in component 2, but compared to Table 12, the same concept loaded within component 3. This might indicate that component 2 of Table 11 and component 3 of Table 12 might be similar to one another.

Table 12

Condensed version of the rotated component matrix, for the revised version taken by freshman, junior, and senior biology majors (n=45)

	Component							
CINS concept and item numbers	1	2	3	4	5	6	7	8
Biotic Potential 1 11		0.599561					0.607877	
Population Stability 2 14				0.56148		0.775993		
Resources Limited 3 12	0.567959						0.796416	
Limited Survival 5 15	0.84681 0.858688							
Variation 9 16				0.755185				0.683374
Variation Inherited 7 17			0.565318			0.807532		
Differential Survival 10 18					0.73919 0.878453			
Change in Population 4 13			0.899591 0.877305					
Origin of Variation 6 19		0.806926 0.879771						
Origin of Species 8 20				0.804641				0.812654

Table 13

Comparison of non-majors' factor analysis: Original (n=44) vs. Revised (n=38).

CINS concept and item numbers	ORIGINAL VERSION Component								REVISED VERSION Component					
	1	2	3	4	5	6	7	8	1	2	3	4	5	6
Biotic Potential 1								0.656				0.868		
11						0.84						0.485		0.618
Population Stability 2	0.756											0.709		
14	0.431						0.448							0.804
Resources Limited 3	0.856										0.837			
12	0.405										0.447	0.416		
Limited Survival 5				-0.67							0.832			
15	0.481		0.429		-0.42				0.624					
Variation 9							0.828		0.61					
16			0.911						0.8					
Variation Inherited 7	0.742										0.627	0.478		
17		0.511	0.419						0.435	0.432				
Differential Survival 10	0.497					0.429			0.751		0.456			
18								-0.68	0.727					
Change in Population 4					0.874					0.845				
13		0.531			0.491					0.848				
Origin of Variation 6		0.717							0.494	0.589			0.51	
19		0.694											0.727	
Origin of Species 8				0.869						0.8				
20		0.775											0.839	

Table 13 (above), illustrates the differences between the original and revised version and how the item pairs loaded. For the original version, eight components were selected to explain the 20 items, while six components were selected for the revised version. Besides a two component difference, the number of item pairs that loaded together was different as well. The four item pairs that loaded together for the original version were populations are stable (2 and 14), resources limited (3 and 12), change in population (4 and 13), and origin of variation (6 and 19), compared to the six from the revised version, biotic potential (1 and 11), resources limited (3 and 12), variation (9 and 16), differential survival (10 and 18), change in population (4 and 13), and origin of variation (6 and 19). This indicates a greater match, although still not for all item pairs, on the revised version of the CINS.

The factor analysis on the data from the revised version shows component 1 with two sets of item pairs, variation (9 and 16) and differential survival (10 and 18), while component 6 did not show any item parings. The lowest paring was item 3 (.447) and item 12 (.837) in component 3, while the highest pairing was represented in component 2, items 4 (.844) and 13 (.847). Additionally, with the majors' factor analysis from Table 12, four item pairings were also observed out of the six that were being investigated. This is important because it indicates that, when a student has an understanding of a concept, he or she can apply that knowledge and answer the item correctly. The more items that are answered correctly, the stronger the individual item will be represented within a component.

Another observation from Table 13 is that of the six concepts that were improved, there seems to be a stronger or higher pairing of the values that were reported in the rotated matrices, with the exception of items 6 and 19, for the revised version. The other four concepts that were not focused on for this study had some pairings, as well. With the original version, item pairs 2 and 14, along with item pairs 3 and 12 were represented and aligned under component 1. For the revised version, those pairs did not load in the same concept, and item pair 2 and 14 did not even load together. This could mean that the non-majors who took the original version related those two concepts together, causing them to load together within the same component thereby when one pairing is answered, the other one will be too. Component 1, for the revised version, had other item pairs load together, 9 and 16 with 10 and 18. Overall, the revised version of the CINS supported more pairing between the item pairs, which represents that those six item pairs are appropriately testing the same concept.

Discussion

The results reported above presnt the number obtained by various analysis, but in order to understand and attempt to explain the numbers, the interviews are valuable. The interviews gave great insight into how students were interpreting and essentially understanding each question, and showed that certain items were more susceptible to guessing, depending on how much the student understood the concept. The interviews conducted with non-majors showed that particular words did trigger certain students to think of different things, based on prior conceptions. The qualitative data from the interviews allowed for a look at the linguistic complexity of the actual words and phrases within each item on the original CINS.

For the majority of each interview, the student was the only one talking unless what they stated was unclear, in which case the interviewer stepped in and asked for clarification. The following are different students' ideas and statements related to the 12 CINS items that were being investigated.

One of the most interesting students to be interviewed for this study was Student 11. Her own conception of evolution was so strong that her reason for why anything changed was due to "the environment being the cause for the change and affecting the population". Anytime the CINS answer included the word "environment", this student indicated that it was correct, which was the case for items 4, 6, 7, 13, 19, and 20. Below is an example of Students 11's response for item 4.

(Item 4)

4. What is the best way to describe the evolutionary changes that occur in a finch population over time?

a. The traits of each finch within a population gradually change.
b. The percentages of finches having different traits within a population change.
c. Successful behaviors learned by finches are passed on to offspring.
d. Mutations occur to meet the needs of the finches as the environment changes.

Student 11: I did not choose [sic] C because it talks about learned behaviors and I don't think behaviors are passed down, it is more of the "traits". So it could be A, because it says traits.

Interviewer: What do you mean by traits?

Student 11: Um…the traits are like the genes….

Interviewer: What do you mean by genes? Like Gap jeans?

Student 11: No, like the picture of the different beaks. The trait for [*sic*] like a specific finch to have like a big beak, to crunch… I don't know… acorns. Maybe their environment that they live in, they would have to adapt to that because that is the main source of food. Rather than another finch would be smaller, in a different area.
I was confused with the word 'primary changes'. I would change that.

Interviewer: What does 'primary changes' mean to you?

Student 11: I think of something like major factor that happened really fast and has impacted them.

Clearly, this student's previous experiences and the conception that she holds about the influence of the enironment has an impact on how she views evolution. Student 11 received a -1 and a 0 for her reponses to item 4. This was due because her response to why she chose answer D, which was unclear and showed no understanding that she understood the concept, and when she verbalized why she did not chose the other answer choices her thoughts were not clear and needed guidance. Recall that Table 5 is the coding rubric that was used to code all of the students responses to each item.

Other students did not understand that factors like mutations, the environment, and traits can not "meet" the needs of the animal to allow for changes to occur in the population. Student 18 is an example of this; she missed almost every one of the 12 items. Her conception of natural selection was that a mutation or traits will "meet the needs of the individual so it can survive". Throughout her interview, she struggled with explaining why certain processes occur. The concept of origin of variation was challenging for her. Below is an excerpt from her interview for item 6, which tests this concept. The excerpt for item 19 will follow.

(Item 6)

6. How did the different beak types <u>first</u> appear in the Galapagos finches?
    a. The changes in the finches' beak size and shape occurred because of their need to be able to eat different kinds of food to survive.
    b. Changes in the finches' beaks occurred randomly, and when there was a good match between beak structure and available food, those birds usually had more offspring.
    c. The changes in the finches' beaks occurred because the environment caused the desired genetic changes.
    d. The finches' beaks changed a little bit in size and shape with each successive generation, some getting larger and some getting smaller because of environmental changes.

Student 18: I'm going to go with A, again because they [the finches] needed to survive because of the scarcity of seeds on the island. I didn't choose B because it occurred by 'chance'….

Interviewer: What do you mean by 'chance'?

Student 18: Just randomly happened. Then with C, [*reading the answer out loud*] 'The changes in the finches' beaks occurred because the environment induced the desired changes'. Desired has to do with want, and that would have nothing to do with the beak changes, and than all of a sudden their genetic structure changes, doesn't happen.

And then, [*Reading answer D*], the beaks changed a little bit in size with each successive generation, which is progression. So no to that one.

So I went with A, because it's due to the survival on the island, which depends on the food supply and the beaks that they have.

From student 18's response for item 6, the point that she seemed to be making is that since the finches need to survive and the type of food that they can eat will depend on the beak structure that the finch already has. Student 18 did not believe that chance had any impact on how the finches will survive. An interesting thought that she had was expressed when she was discussing her reasons for not choosing answer C. Student 18 clearly understood that the finches can not want or desire their beaks to change, but it is evident that she is not making the connection that individuals or species can not want or need to change themselves in order to survive. Below is student 18's thoughts and answer on item 19.

(Item 19)

19. According to the theory of natural selection, where did the variation in body size of the three species most likely come from?
    a. The lizards needed to change in order to survive, so beneficial new traits developed.
    b. The lizards wanted to become different in size, so beneficial new traits gradually appeared in the population.
    c. Random genetic changes and new combinations of genes produced through mating both produce variation.
    d. The island environment caused genetic changes in the lizards.

When answering item 6, student 18 felt that individuals can not want to change which is why, when she was reading through the answer choices for item 19, she decided immediately that B was out, because it stated that the lizards <u>wanted</u> to become different in size.

Student 18: I don’t think it’s D, I don’t think the environment can cause genetic changes, maybe behavioral, like for survival they will change [*sic*] a little bit on the surface, but not their whole DNA. I think A, because [*sic*] it was like a forceful change for them to survive, so then new traits were developed.

Interviewer: So the environment forced them to change?

Student 18: Yeah.

Interviewer: Okay, so how do you define the theory of natural selection?

Student 18: Hmm. It’s like a change that occurs within a species, because of their necessity to survive. So it is like a competition. [*Within a species*] They all have to [*sic*] like change, and depending on who gets the better change, out of luck, that particular part of the species will like survive. So it really depends on how well they are developed in that change.

Interviewer: Their genetic change?

Student18: I think it’s more like their performance, like how well their performance is.

Interviewer: Okay, so the organisms’ performance is what causes them to change?

Student 18: No, that determines like [*sic*] which one will be, like the fittest one. But definitely their survival will influence their change.

Interviewer: Change to a new species?

Student 18: No like they [*sic*] will better develop some traits, or they will lose some traits, depending on their survival. It really depends on how [*sic*] well they can get their food, how well they can avoid predators, to reproduce and have viable offspring. If they can do that, then they can be the fittest one. That is what I [*sic*] think the theory is.

Student 18 did not understand what the biological explaination of the term 'fitness' means because she was only able to articulate the non-biological definition. Overall, she had a hard time understanding what natural selection actually is. She appeared to be mixing different ideas to make the answer that she choose sound correct. A lot of other students who were interviewed did this as well. Student 18 received 0's for item 6 and -1's for item 19. Clearly this student has no understanding of the concept of origin of variation based on her responses.

Students 6 and 7 had the same reasoning for items 10 and 18, which tested the concept of differential survival. Both students participated in the ROTC program at the university, so their conception for the term "fitness" was based on working out, which shows that from their prior experiences, working out leads to an individual becoming more fit for their environment so that they can get away from predators, etc. This perception was also seen during the pilot study that was done by the author.

In contrast to these examples of students who held alternative conception, there were individuals who performed extremely well on the 12 CINS items and were able to articulate their reasons as to why a specific answer choice was correct. There were also certain times that students would answer the question correctly, but would not be able to explain why. This would be due mainly to the process of elimination that the student was using when going through all of the answer choices. For example, Student 17 answered all 12 items correctly, but did not receive the highest coding score for understanding, compared to Student 13. The difference between these two students was in how they articulated their knowledge and understanding of the concepts within natural selection. Below are examples of two of the students who had high performance on the 12 CINS items, and scored a 10 out of 12.

(Item 18)

18. Fitness is a term often used by biologists to explain the evolutionary success of certain organisms. Below are descriptions of four fictional female lizards. Which lizard might a biologist consider to be the "most fit" according to Darwin's Theory?

	Lizard A	Lizard B	Lizard C	Lizard D
Body length	20 cm	12 cm	10 cm	15 cm
Offspring surviving to adulthood	19	28	22	26
Age at death	4 years	5 years	4 years	6 years
Additional Facts	Lizard A is very healthy, strong, and clever	Lizard B has mated with many lizards	Lizard C is dark-colored and very quick.	Lizard D has the largest territory of all the lizards.

a. Lizard A
b. Lizard B
c. Lizard C
d. Lizard D

Student 1: B, probably because fitness would give them more opportunities to pass on and have more offspring survive. They had 28 survive to adulthood, while lizard A only had 19… um..that's a lot, [*sic*] but it doesn't mean they are most fit, because the population of B is growing, so they are best fitted for that area. The chart is set up well and clear. I wouldn't change anything for the question.

Interviewer: What is your definition of fitness?

Student 1: Like how many offspring they have, which evolve best to their surrounding, evolutionary success. Those would be successful in having more offspring.

Student 1 earned a 10 on the CINS because she was able to chose the correct answer for 10 of the items during the interview. The concept that she struggled with was variation, items 9 and 16, which were the items that she missed to received the 10 out 12. Student 1 did answer item 18 correctly and when this section was coded from the interview, she received a 2 for question one and a 1 for question two. Item 18 was chosen to represent that the student did have an understanding, but also to show that she sometimes did not articulate why the answer that was chosen as correct. This was also seen during other interviews, with student 1 tending to use the process of elimination to help answer the question. For her interviews she scored a 1.6, which is a good score and shows that she has knowledge, but also indicated that she might not fully understand some of the CINS items.

(Item 8)

8. What caused populations of birds having different beak shapes and sizes to become distinct species distributed on the various islands?
   a. The finches were quite varied, and those whose features were best suited to the available food supply on each island reproduced most successfully.
   b. All finches are essentially alike and there are <u>not</u> really fourteen different species.
   c. Different foods are available on different islands and for that reason, individual finches on each island gradually developed the beaks they needed.
   d. Different lines of finches developed different beak types because they needed them in order to obtain the available food.

Student 15: The answer is A, [Read the answer]. Um.. just [*sic*] because I think that there was like a variety of finches to begin with. So I think the one that had the beaks that were able to find food that they could use their beaks to eat were more successful with living and reproducing than the ones that couldn't.

I didn't choose B, C, or D because B is wrong, there are really 14 different species of finches. And…I think the other two (C & D) say something about how the beaks are being formed because they needed to be that way so the finches can survive. And we have talked about in our class, genes don't come just because they are needed.

I liked how the question was clear, and I understood what it was asking. I wouldn't change any of the words to make more sense.

Student 15's response to item 8 was chosen to represent when the student has the knowledge and understanding. She was able to articulate why the answer was A and why answers B, C, and D were not right. This student is different from student 1 because she was able to articulate her understanding on the concept of origin of species. This excerpt shows how a student might verbalize their reasoning as to how they understand the item. Both students 1 and 15 received the same scores for both the 12 CINS items that were tested along with the average of the interview questions one and two, which were coded. As stated above, student 17 received a perfect score (12 out of 12) during the interview. Below is an excerpt from her interview for item 19.

(Item 19)

19. According to the theory of natural selection, where did the variation in body size of the three species most likely come from?
    a. The lizards needed to change in order to survive, so beneficial new traits developed.
    b. The lizards wanted to become different in size, so beneficial new traits gradually appeared in the population.
    c. Random genetic changes and new combinations of genes produced through mating both produce variation.
    d. The island environment caused genetic changes in the lizards.

Student 17: I will go with C [*Reading the answer*] so through out the generations like the breeding between like a larger and smaller lizards, those combination of those created varation.

Interviewer: Can you just briefly go over the theory of natural selection?

Student 17: The theory of natural selection…there are all sorts of different factors that play in natural selection. Their (the lizards) ability to survive is[*sic*] like based on certain variations, or um….characteristics that they have.

Their ability to survive is the biggest contributor to um...like whether they can produce the most offspring or not. That is going to allow certain generations with certain triats will thrive in certain evironments, where others wont, which is survival of the fittest.

Interviewer: Would you make any changes?

Student 17: Sexual recomindation, was a little confusing. I feel like it was refered to before, but in different words. But I understood what that was saying.

Student 17 understood how variation can produce new traits, which can have an impact on a species population. She received 2's based on how she verbalized her reasoning. This excerpt was chosen to represent how the student who has an understanding of the concept can verbally articuale how new combinations can form varation. Comparing student 17 with student 18 from before, it can be seen that there is a difference in how one answers the item based on how they understand what the question is asking. Unlike students 1 and 15 who scored a 10, student 17 is very different to those students. The importance of this shows that there are different factors that can allow for a student to answer a question correctly, which is why all of the interviews were scored based on how the student answered the item and stated why. The main point is that understanding the concept helps answer the item, but being able to make the connections and articulate why the answer is correct represents the students actual conception, compared to just being a good test taker.

Based on the interview results, it is apparent that students hold many different conceptions when it comes to the concepts within natural selection. It appears that the linguistic complexity of the 12 original CINS items that were tested hindered student understanding. Furthermore, if the student knew the concept and had a scientific meaning for it, then he or she would have more understanding and would better understand what the items were asking. The students who scored high on the 12 items and could explain themselves are examples of this, compared to, for example, Students 11 and 18, who struggled more with the 12 items, based on their prior knowledge and experiences. The interviews provided data on which words and phrases that gave students trouble and hindered them from understanding what the item was asking. All of the changes that were made to the revised version of the test were done so students can have a better understanding of what the item is asking (Appendices C & D).

Conclusion

The purpose of this study was to use feedback and suggestions that were provided by studwents during interviews to improve six item pairs from the 2004 version of the CINS, since previous research (Anderson et al., 2002) showed that these item pairs did not load together on a factor analysis.

Research Question One: Does the linguistic complexity in the CINS items hinder student understanding of the item by using words and phrases to which they may have been exposed to but do not have tangible sense of what the terms mean?

All of the students who participated in the interviews, had already had the opportunity to learn natural selection in their biology course. Those students who did not perform as well could still be under going the learning process by trying to make sense of what they were being taught and reconciling that with the conceptions they already hold. Thus they would have a harder time answering the items when they were not familiar with content or vocabulary.

Another contributing factor is knowledge that the student already has on the concept. It was noticed that some students who answered some of the 12 items incorrectly, during class, were able to answer the items correctly during the interview. This most likely was due to the student starting to make connections and form scientific conceptions related to natural selection. Another possiblity could be that the student took the interivew more seriously and may have tried harder since it was more of the one-on-one environment. When the students took the CINS in the class, there seemed to be a couple of students who answered a particular item correctly in class, but answered that item incorrectly during the interview. This could be the result of students second-guessing themselves, that they may just be good test takers, that they could have been guessing, or that the students do not have a complete understanding of how natural selection and its mechanisms work.

The amount of knowledge that a student has, as well as how much understanding they have on a concept can impact how one interprets and applies their knowledge to answer a question. For clarification, students can gain and have knowledge but their level of understanding will differ depending on how well the student can apply that knowledge that they have and make the connections to different situations. Certain words did prove to cause some issues with the students and impacted how they answered the item on the CINS (see table in Appendix D). If the student did not know certain words or phrases, they either guessed and/or applied the unscientific knowledge that they already hold to answer the CINS item. Phrases like ‘primary changes’, and ‘sexual recombination’ and words such as ‘proportions’, ‘arise’, and ‘characteristics’, were among those identified by students as not being understood. For the revised version, all of the students’ suggestions were incorporated into improving those 12 items. All of the changes that were done to the 12 CINS items can be found in Appendix E. For example, in the item pairing of 4 and 13, which included the troublesome words ‘primary changes’ and ‘proportions’, “proportions” was changed to “percentages” and the phrase ‘primary changes’ was removed. As a result of changing of the words and phrases completely for the revised CINS version allowed for those imporvements that were seen between the two non-major biology classes. Overall, with the certain words within each item did have an impact on how the student answered each item, and were either removed or changed to a word that would hold more meaning for the student.

Research Question Two: For the revised version of the CINS, are the two items that assess one concept equivalent in terms of how the students answer them? If the student answers only one item of a particular item pair correctly, does this mean that they do not fully understand the concept?

The equivalency of the two items assessing the same concept can be seen in the factor analysis that was conducted for both non-majors' classes. There was a noticeable difference between the original and the revised versions' factor analysis. More concepts loaded together for the revised version, which shows that the improvements made were beneficial and successful in improving the effectiveness of the CINS.

Before the changes were implemented, the concepts of change in population, origin of variation, resources limited, and population stability were the only item pairs loaded together in the factor analysis. After the changes, the revised version had more item pairs load together, thereby indicating that the improved items pairs may be more effective in assessing a concept. Out of the six concepts that were focused on in the interviews, four items parings from the revised version loaded together (*Variation, Change in Population, Origin of Variation, and Differential Survival*), showing that those changes were beneficial in improving how the student understood and interpreted each item and what it was asking. The concepts variation and differential survival loaded in the same component, meaning that those item pairs may be closely related to one another. Four item pairs did not load together: population stability, limited survival, variation inherited, and origin of species. This indicates that there are still some issues and that further revisions will most likely be needed.

How the student answered each item during the interview showed whether or not they essentially understood the concept, which was based on how the student articulated their reasoning why for each of the answer choices. The data does not seem to indicate that if a student answers only one item out of a pair correctly that they do not fully understand the concept. Rather, it appears that many factors may contribute to why the student incorrectly answered the items. For example, lack of understanding and/or knowledge or the item or concept and/or random guessing may contribute.

From observations made during the interviews, it was seen that students would often miss the first item, but when they got to the second item, they were able to either answer that second item correctly or were at least starting to make connections. Item difficulty can also have an impact on how students understand the item, but if they understand the concept, then they will be able to answer the item sufficiently, regardless of the item's difficulty. The students who are lower performers and might not have a complete understanding of a concept tend to guess. No matter how difficult or how easy the item was, they would guess because of their lack of understanding of the concepts within natural selection. This was especially seen with student 18.

Limitations

This study had some limitations. Perhaps most the imporatant limitation was the number of participants. Sometimes small sample sizes do not justify the results, where the bigger sample size would confirm or reject the data. The factor analysis presented in this study indicates that the changes made to the CINS were beneficial. However, if more students had participated, the researcher could be more confident in this clam.

Another limitation may have been the willingness of the students to answer the questions thoughtfully even though they did not get any class credit. It was observed during the non-majors data collection in the spring that some students were finished much more quickly than others, as evidenced by them looking around. The majority of the students, however, seemeed to take the test seriously and fully participate, because only six students opted to not have their data included.

Implications for teachers

Identifying the common alternative conceptions held by students is essential in guiding effective instruction. The CINS provides the teacher with a tool to assess students so that they know the students understanding of the concepts within natural selection. By modification of the CINS items, this study makes progress towards providing teachers with an effective and easy-use tool that teachers can use to accurately assess their students' conceptions on the concepts within natural selection. In the future, teachers can be even more confident in using the CINS as a pre-test (Questions 1–10) and post-test (Questions 11-20) to measure conceptual change before and after instruction. This was Anderson and Fisher's (2002) original goal. Improving the four concepts that did not originally load together within a component makes the CINS one step closer to this goal.

Future Research

The differences between the scores of the two non-majors' classes were significant suggesting that the improved, 5th version of the CINS is effective in assessing students understanding on natural selection, even for students who are not majoring in the biological sciences. More work will need to be done in terms of wording adjustments to the concept of variation inherited (7 and 17) and origin of species (8 and 20) so that all of the CINS item pairs load together and accurately measure the same concept.

Furthermore, future studies should consist of a larger student sample size, and should take a more in-depth look at how the student answers each item. If the student answers one item correctly, that should be an indicator that they will successfully answer the paired item correctly since it is designed to assess the same concept. Once this is done, the CINS can be looked at to accurately measure college students' understanding of natural selection and can be an even more useful tool that teachers can continue to use in the classroom.

Appendix A: Original 20 CINS items

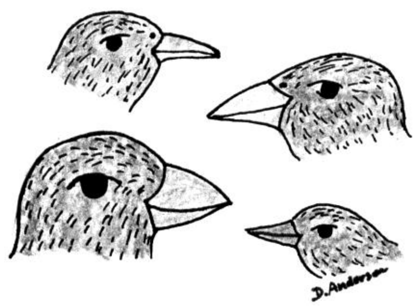
Conceptual Inventory of Natural Selection

4th edition - 2003

Your answers will assess your understanding of the Theory of Natural Selection. Please choose the answer that best reflects how a biologist would think about each question.

Galapagos finches

Scientists have long believed that the 14 species of finches on the Galapagos Islands evolved from a single species of finch that migrated to the islands one to five million years ago (Lack, 1940). Recent DNA analyses support the conclusion that all of the Galapagos finches evolved from the warbler finch (Grant, Grant & Petren, 2001; Petren, Grant & Grant, 1999). Different species live on different islands. For example, the medium ground finch and the cactus finch live on one island. The large cactus finch occupies another island. One of the major changes in the finches is in their beak sizes and shapes as shown in this figure.



Choose the one answer that best reflects how an evolutionary biologist would answer.

1. What would happen if a breeding pair of finches was placed on an island under ideal conditions with no predators and unlimited food so that all individuals survived?
   a. The finch population would stay small because birds only have enough babies to replace themselves.
   b. The finch population would double and then stay relatively stable.
   c. The finch population would increase dramatically.
   d. The finch population would grow slowly and then level off.

2. Finches on the Galapagos Islands require food to eat and water to drink. How does this fact impact the population?
   a. When food and water are scarce, some birds may be unable to obtain what they need to survive.
   b. When food and water are limited, the finches will find other food sources, so there is always enough.
   c. When food and water are scarce, the finches all eat and drink less so that all birds survive.
   d. There is always plenty of food and water on the Galapagos Islands to meet the finches' needs.

3. Once a population of finches has lived on a particular island for many years, what will most likely happen to the population?
   a. The population continues to grow rapidly.
   b. The population remains relatively stable, with some fluctuations.
   c. The population dramatically increases and decreases each year.
   d. The population will decrease steadily, than increase.

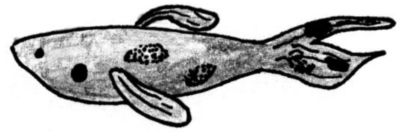
4. In the finch population, what are the primary changes that occur gradually over time?
   a. The traits of each finch within a population gradually change.
   b. The proportions of finches having different traits within a population change.
   c. Successful behaviors learned by finches are passed on to offspring.
   d. Mutations occur to meet the needs of the finches as the environment changes.

5. Depending on their beak size and shape, some finches get nectar from flowers, some eat grubs from bark, some eat small seeds, and some eat large nuts. Which statement best describes the interactions among the finches and the food supply?
   a. Most of the finches on an island cooperate to find food and share what they find.
   b. Many of the finches on an island fight with one another and the physically strongest ones win.
   c. There is more than enough food to meet all the finches' needs so they don't need to compete for food.
   d. Finches compete primarily with closely related finches that eat the same kinds of food**,** and some may die from lack of food.

6. How did the different beak types first arise in the Galapagos finches?
   a. The changes in the finches' beak size and shape occurred because of their need to be able to eat different kinds of food to survive.
   b. Changes in the finches' beaks occurred by chance, and when there was a good match between beak structure and available food, those birds had more offspring.
   c. The changes in the finches' beaks occurred because the environment induced the desired genetic changes.
   d. The finches' beaks changed a little bit in size and shape with each successive generation, some getting larger and some getting smaller.

7. What type of variation in finches is passed to the offspring?
   a. Any behaviors that were learned during a finch's lifetime
   b. Only characteristics that were beneficial during a finch's lifetime
   c. All characteristics that were genetically determined
   d. Any characteristics that were positively influenced by the environment during a finch's lifetime.

8. What caused populations of birds having different beak shapes and sizes to become distinct species distributed on the various islands?
   a. The finches were quite variable, and those whose features were best suited to the available food supply on each island reproduced most successfully.
   b. All finches are essentially alike and there are not really fourteen different species.
   c. Different foods are available on different islands and for that reason, individual finches on each island gradually developed the beaks they needed.
   d. Different lines of finches developed different beak types because they needed them in order to obtain the available food.



Venezuelan guppies

Guppies are small fish found in streams in Venezuela. Male guppies are brightly colored, with black, red, blue and iridescent (reflective) spots. Males cannot be too brightly colored or they will be seen and consumed by predators, but if they are too plain, females will choose other males. Natural selection and sexual selection push in opposite directions. When a guppy population lives in a stream in the absence of predators, the proportion of males that are bright and flashy increases in the population. If a few aggressive predators are added to the same stream, the proportion of bright-colored males decreases within about five months (3-4 generations). The effects of predators on guppy coloration have been studied in artificial ponds with mild, aggressive, and no predators, and by similar manipulations of natural stream environments (Endler, 1980).

Choose the one answer that best reflects how an evolutionary biologist would answer.

9. A typical natural population of guppies consists of hundreds of guppies. Which statement best describes the guppies of a single species in an isolated population?
   a. The guppies share all of the same characteristics and are identical to each other.
   b. The guppies share all of the essential characteristics of the species; the minor variations they display don't affect survival.
   c. The guppies are all identical on the inside, but have many differences in appearance.
   d. The guppies share many essential characteristics, but also vary in many features.

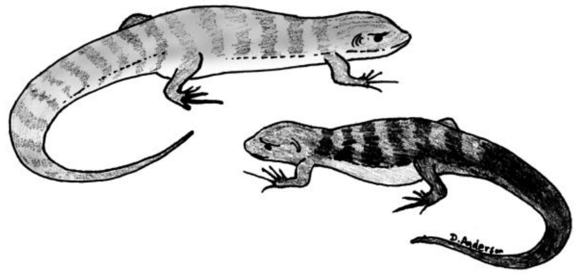
10. Fitness is a term often used by biologists to explain the evolutionary success of certain organisms. Which feature would a biologist consider to be most important in determining which guppies were the "most fit"?
    a. large body size and ability to swim quickly away from predators
    b. excellent ability to compete for food
    c. high number of offspring that survived to reproductive age
    d. high number of matings with many different females.

11. Assuming ideal conditions with abundant food and space, and no predators, what would happen if a mating pair of guppies was placed in a large pond?
    a. The guppy population would grow slowly, as guppies would have only the number of offspring that are needed to replenish the population.
    b. The guppy population would grow slowly at first, then would grow rapidly, and thousands of guppies would fill the pond.
    c. The guppy population would never become very large, because only organisms such as insects and bacteria reproduce in that manner.
    d. The guppy population would continue to grow slowly over time.

12. Once a population of guppies has been established for a number of years in a pond with other organisms including predators, what will likely happen to the population if conditions remain constant?
    a. The guppy population will stay about the same size.
    b. The guppy population will continue to rapidly grow in size.
    c. The guppy population will gradually decrease until no more guppies are left.
    d. It is impossible to tell because populations do not follow patterns.

13. In guppy populations, what are the primary changes that occur gradually over time?
    a. The traits of each individual guppy within a population gradually change.
    b. The proportions of guppies having different traits within a population change.
    c. Successful behaviors learned by certain guppies are passed on to offspring.
    d. Mutations occur to meet the needs of the guppies as the environment changes.

Canary Island Lizards



The Canary Islands are seven islands just west of the African continent. The islands gradually became colonized with life: plants, lizards, birds, etc. Three different species of lizards found on the islands are similar to one species found on the African continent (Thorpe & Brown, 1989). Because of this, scientists assume that the lizards traveled from Africa to the Canary Islands by floating on tree trunks washed out to sea.

Choose the one answer that best reflects how an evolutionary biologist would answer.

14. Lizards eat a variety of insects and plants. Which statement describes the availability of food for lizards on the Canary Islands?
    a. Finding food is not a problem since food is always in abundant supply.
    b. Since lizards can eat a variety of foods, there is likely to be enough food for all of the lizards at all times.
    c. Lizards can get by on very little food, so the food supply does not matter.
    d. It is likely that sometimes there is enough food, but at other times there is not enough food for all of the lizards.

15. What do you think happens among the lizards of a certain species when the food supply is limited?
    a. The lizards will cooperate to find food and share what they find.
    b. The lizards fight for the available food and the stronger lizards kill the weaker ones.
    c. Genetic changes that would allow lizards to eat new food sources are more likely to occur..
    d. The lizards least successful in the competition for food are likely to die of starvation and malnutrition.

16. A well-established population of lizards is made up of hundreds of individual lizards. On an island, all lizards in a lizard population are likely to . . .
    a. be indistinguishable, since there is a lot of interbreeding in isolated populations.
    b. be the same on the inside but display differences in their external features.
    c. be similar, yet have some significant differences in their internal and external features.
    d. be the same on the outside but display differences in their internal features.

17. Which statement best describes how traits in lizards will be inherited by offspring?
    a. When parent lizards learn to catch particular insects, their offspring can inherit their specific insect-catching-skills.
    b. When parent lizards develop stronger claws through repeated use in catching prey, their offspring can inherit their stronger-claw trait.
    c. When parent lizards’ claws are underdeveloped because easy food sources are available, their offspring can inherit their weakened claws.
    d. When a parent lizard is born with an extra finger on its claws, its offspring can inherit six-fingered claws.

18. Fitness is a term often used by biologists to explain the evolutionary success of certain organisms. Below are descriptions of four fictional female lizards. Which lizard might a biologist consider to be the "most fit"?

	Lizard A	Lizard B	Lizard C	Lizard D
Body length	20 cm	12 cm	10 cm	15 cm
Offspring surviving to adulthood	19	28	22	26
Age at death	4 years	5 years	4 years	6 years
Comments	Lizard A is very healthy, strong, and clever	Lizard B has mated with many lizards	Lizard C is dark-colored and very qui	Lizard D has the largest territory of all the lizards.

a. Lizard A
b. Lizard B
c. Lizard C
d. Lizard D

19. According to the theory of natural selection, where did the variations in body size in the three species of lizards most likely come from?
   a. The lizards needed to change in order to survive, so beneficial new traits developed.
   b. The lizards wanted to become different in size, so beneficial new traits gradually appeared in the population.
   c. Random genetic changes and sexual recombination both created new variations.
   d. The island environment caused genetic changes in the lizards.

20. What could cause one species to change into three species over time?
   a. Groups of lizards encountered different island environments so the lizards needed to become new species with different traits in order to survive.
   b. Groups of lizards must have been geographically isolated from other groups and random genetic changes must have accumulated in these lizard populations over time.
   c. There may be minor variations, but all lizards are essentially alike and all are members of a single species.
   d. In order to survive, different groups of lizards needed to adapt to the different islands, and so all organisms in each group gradually evolved to become a new lizard species.

Appendix B: IRB approval for study

PLNU IRB
Exempt Review
794

Date: Thursday, November 4, 2010
PI: Danielle Dwyer
Additional Investigators: N/A
Faculty Advisor: Dianne Anderson, Ph.D.
Title: Using data from semi-structured interviews with non-major biology students to improve the conceptual inventory of natural selection.

The research proposal was reviewed and verified as Exempt from further review under category 2 and has been approved in accordance with PLNU's IRB and federal requirements pertaining to human subjects protections within the **Federal Law 45 CFR 46.101 b**. Your project will be subject to approval for one year from the November 1, 2010 date of approval. After completion of your study or by November 1, 2011, you must submit a summary of your project or a request for continuation to the IRB. If any changes to your study are planned or you require additional time to complete your project, please notify the IRB chair.

For questions related to this correspondence, please contact the IRB Chair, Ross A. Oakes Mueller, Ph.D., at the contact information below. To access the IRB to request a review for a modification or renewal of your protocol, or to access relevant policies and guidelines related to the involvement of human subjects in research, please visit the PLNU IRB web site.

Best wishes on your study,

Ross A. Oakes Mueller, Ph.D.
Associate Professor
Department of Psychology
IRB Chair
Point Loma Nazarene University
3900 Lomaland Dr. San Diego, CA 92106
619.849.2905
rossoakesmueller@pointloma.edu

Appendix C: Student consent form given to all of the classes that participated in the study

<u>Given to the Fall Non-Majors Class:</u>

I, Danielle Dwyer, am a biology graduate student at PLNU. I am studying how to improve science education by improving assessment questions on the CINS test for my thesis. You, the participant, will have the opportunity to have your data included in the research study, along with the opportunity to volunteer for the interviews that I will be conducting in which you will be asked about your understanding of some of the CINS questions. For publication purposes, your name and any other identifying information will be removed. Your participation is strictly voluntary, and has no direct effect on your grade in this course. Please read the options that are available to you. Check all that apply.

If you have questions about this study, please contact one of the following:

Graduate student Danielle Dwyer
dmdwyer1986@pointloma.edu
PLNU Advisor Dr. Dianne Anderson
dianneanderson@pointloma.edu
IRB committee chair Dr. Ross Oakes-Mueller
rossoakesmueller@pointloma.edu

Tear here if you want to keep the top portion for later reference.

Check all that apply:

☐ I do wish to have my data included in the current research.

☐ I would like to be selected to participate in the approximately 30-minute interviews for this research and to earn a $10 gift card for my participation.

☐ I would like to be emailed the results when this research has been concluded.

Participant's Name: ____
Participant's Signature: ____
Email Address: ____

Given to the Spring Classes (non-majors and majors)

I, Danielle Dwyer, am a biology graduate student at PLNU. I am studying how to improve science education by improving assessment questions on the CINS test for my thesis. You, the participant, will have the opportunity to have your data included in the research study. For publication purposes, your name and any other identifying information will be removed. Your participation is strictly voluntary, and has no direct effect on your grade in this course. Please read the options that are available to you. Check all that apply.

If you have questions about this study, please contact one of the following:

Graduate student	Danielle Dwyer	dmdwyer1986@pointloma.edu
PLNU Advisor	Dr. Dianne Anderson	dianneanderson@pointloma.edu
IRB committee chair	Dr. Ross Oakes-Mueller	rossoakesmueller@pointloma.edu

Tear here if you want to keep the top portion for later reference.

Check all that apply:

☐ I do wish to have my data included in the current research.

☐ I would like to be emailed the results when this research has been concluded.

Participant's Name: ____

Participant's Signature: ____

Email Address: ____

Appendix D: Interview Data

Feedback and suggestions for CINS items

Item	Things heard/ noticed/ stated by students	Possible Solutions stated by students
4	-Primary Changes. -Proportions: this word seemed to be new to some of the interviewees.	-what type of primary changes are you referring to? Morphological, ability, behavioral, etc. Define/state what you mean by that specifically, explain, and expand. Also is it each finch or each species of finch? -maybe say 'Main changes' -maybe use another word. The interviewees who had concern could not think of another word to replace it.
6	-Mixed feelings with the word <u>first</u> and how it was underlined. Some interviewees like how the word was underlined, which created something to focus on, while others thought it was confusing. Meaning what is it asking. -what is first referring to? -arise	-Solution: clarify the meaning of first for the sentence. Is it referring to right when they moved or have they been there for a while? Specify the timeframe. -one interviewee believed that first did not need to be underlined. -Change the question to: "How did the different beak types arise?" had an issue with the word first, thought of "In the beginning" -one interviewee would like to see clarification on this word. Was it how they first appeared?
7	-Change "is" in the question. -Primary changes. -for the second question with primary changes, some of the interviewees did not have an issue with the word. It actually made more sense. -what are characteristics? -The answer choices were distinct from each other, which help them answer the question. -variation	-change it to, "What type of variation in finches *are* passed to the offspring?" -what are the primary changes? -One interviewee wasn't sure if it was referring to physical qualities and would think the question would make more sense if that word was defined. -clarification on variation and is it being linked to characteristics.

8	-liked how the question gave the size of the population, and clarifies the difference between population and species. -"what caused populations of birds to become distinct species." -too many lizards in the question. -liked how the question especially talked about the lizards. -clarification on the lizards that the question is addressing. -quite variable -the options are wordy	-majority liked the question; some had issues with the words "quite variable". Mention the traits instead. -state whether they are different beak shapes and sizes to get the main point across. -throw in a pronoun to make it sound/read more smoothly. -allowed them to focus on the lizard as an individual. -would the population of lizards be 1 or multiple species? -confusing, maybe put diverse in variability. -condense down so the options can flow.
9	-One interviewee thought the first sentence, "A typical natural population of guppies consists of hundreds of guppies." should not be in the question. -isolated population -single species -Actual question	-take it out completely. -clarify: is the isolated population a lot of individuals, or types of species? -clarify what this means in the question. -combine the sentence and question together will make more sense.
10	-the definition of fitness did help majority of the interviewees answer the question in the way they understand the word fitness. -'most fit' -Darwin's idea -evolutionary success	-Maybe order the lizards based on size to show that there are differences in lengths. -is it referring to the individual or the population? -insert survival of the fittest within the question. How the guppies was able to survive and pass on their traits. -one interviewee thought it could be interpreted in two ways: 1. a species evolving over time 2. things just evolving period. -maybe add a sentence to clarify the meaning you are wanting to answer the question.
13	-primary changes.	-make the options more specific with example of primary changes. Give actual features of the primary changes (i.e. color). The options need to address what primary changes. -'what are the changes that happen over time?'
16	No change was stated.	

17	No change was stated.	
18	-One interviewee stated that they focused on the comments to answer the question. -chart was very helpful. -comments section -charge organization	-it swayed the interviewee when answering the question. -allowed for them to see which lizard was the 'most fit' by organizing the information. -change to notes on adaptations. -maybe go by body length to show an order.
19	-change part of the question. -natural selection -sexual recombination	-"where did the variation in body size *for* the 3 species of lizards." -put the definition of Natural Selection within the question. -these two words seemed to be new for a couple of interviewee's.
20	No change was stated.	

Paragraphs:
The majority of student participants believed that the paragraphs were very helpful, gave knowledge, and provided good background information. Most said that they could answer the questions without the paragraphs, but some also noted that it would be a little challenging to answer the questions. For the guppies and lizards paragraphs, most wanted them in, because the examples that they had in class were not on these species. It was also stated that the paragraphs were an effective way of separating the questions from each other in sections. A couple of interviewee's stated that the paragraphs helped with answering the questions better and helped give them some understanding on the species that the paragraph is focusing on. They also felt that, with the guppies paragraph, discussing sexual selection helped them to answer the questions.

Appendix E: Suggestions and feedback from the interviews, using 12 out of the 20 items from the 2004 CINS

Original CINS Question	**Revised CINS Question**
4. In the finch population, what are the primary changes that occur gradually over time? a. The traits of each finch within a population gradually change. b. The proportions of finches having different traits within a population change. c. Successful behaviors learned by finches are passed on to offspring. d. Mutations occur to meet the needs of the finches as the environment changes.\	4. In the finch population, what are the changes that occur gradually over time? a. The traits of each finch within a population gradually change. b. The percentages of finches having different traits within a population change. c. Successful behaviors learned by finches are passed on to offspring. d. Mutations occur to meet the needs of the finches as the environment changes.
6. How did the different beak types first arise in the Galapagos finches? a. The changes in the finches' beak size and shape occurred because of their need to be able to eat different kinds of food to survive. b. Changes in the finches' beaks occurred by chance, and when there was a good match between beak structure and available food, those birds had more offspring. c. The changes in the finches' beaks occurred because the environment induced the desired genetic changes. d. The finches' beaks changed a little bit in size and shape with each successive generation, some getting larger and some getting smaller.	6. How did the different beak types first appear in the Galapagos finches? a. The changes in the finches' beak size and shape occurred because of their need to be able to eat different kinds of food to survive. b. Changes in the finches' beaks occurred randomly, and when there was a good match between beak structure and available food, those birds had more offspring. c. The changes in the finches' beaks occurred because the environment induced the desired genetic changes. d. The finches' beaks changed in size and shape with each successive generation, some getting larger and some getting smaller because of environmental changes.
7. What type of variation in finches is passed to the offspring? a. Any behaviors that were learned during a finch's lifetime b. Only characteristics that were beneficial during a finch's lifetime c. All characteristics that were genetically determined d. Any characteristics that were positively influenced by the environment during a finch's lifetime	7. What type of variation in the finches' traits is passed to the offspring? a. Only behaviors that were learned during a finch's lifetime b. Only traits that were beneficial during a finch's lifetime c. Only traits that were genetically determined d. Only traits that were positively influenced by the environment during a finch's lifetime

8. What caused populations of birds having different beak shapes and sizes to become distinct species distributed on the various islands? a. The finches were quite variable, and those whose features were best suited to the available food supply on each island reproduced most successfully. b. All finches are essentially alike and there are <u>not</u> really fourteen different species. c. Different foods are available on different islands and for that reason, individual finches on each island gradually developed the beaks they needed. d. Different lines of finches developed different beak types because they needed them in order to obtain the available food.	8. What caused populations of birds having different beak shapes and sizes to become distinct species distributed on the various islands? a. The finches were quite varied, and those whose features were best suited to the available food supply on each island reproduced most successfully. b. All finches are essentially alike and there are <u>not</u> really fourteen different species. c. Different foods are available on different islands and for that reason, individual finches on each island gradually developed the beaks they needed. d. Different lines of finches developed different beak types because they needed them in order to obtain the available food.
9. A typical natural population of guppies consists of hundreds of guppies. Which statement best describes the guppies of a single species in an isolated population? a. The guppies share all of the same characteristics and are identical to each other. b. The guppies share all of the essential characteristics of the species; the minor variations they display don't affect survival. c. The guppies are all identical on the inside, but have many differences in appearance. d. The guppies share many essential characteristics, but also vary in many features.	9. A natural population of guppies consists of hundreds of guppies of a single species. Which statement best describes the population of guppies? a. The guppies share all of the same characteristics and are identical to each other. b. The guppies share all of the essential characteristics of the species; the minor variations they display don't affect survival. c. The guppies are all identical on the inside, but have many differences in appearance. d. The guppies share many essential characteristics, but also vary in many features.
10. Fitness is a term often used by biologists to explain the evolutionary success of certain organisms. Which feature would a biologist consider to be most important in determining which guppies were the "most fit"? a. large body size and ability to swim quickly away from predators b. excellent ability to compete for food c. high number of offspring that survived to reproductive age d. high number of matings with many different females.	10. Fitness is a term often used by biologists to explain the evolutionary success of certain organisms. Which feature would a biologist consider to be most important in determining which guppies were the "most fit" according to Darwin's theory? a. large body size and ability to swim quickly away from predators b. excellent ability to compete for food c. high number of offspring that survived to reproductive age d. high number of matings with many different females.

13. In guppy populations, what are the primary changes that occur gradually over time? a. The traits of each individual guppy within a population gradually change. b. The proportions of guppies having different traits within a population change. c. Successful behaviors learned by certain guppies are passed on to offspring. d. Mutations occur to meet the needs of the guppies as the environment changes.	13. In guppy populations, what are the changes that occur gradually over time? a. The traits of each individual guppy within a population gradually change. b. The percentage of guppies having different traits within a population change. c. Successful behaviors learned by certain guppies are passed on to offspring. d. Mutations occur to meet the needs of the guppies as the environment changes.
16. A well-established population of lizards is made up of hundreds of individual lizards. On an island, all lizards in a lizard population are likely to . . . a. be indistinguishable, since there is a lot of interbreeding in isolated populations. b. be the same on the inside but display differences in their external features. c. be similar, yet have some significant differences in their internal and external features. d. be the same on the outside but display differences in their internal features.	16. A population of lizards is made up of hundreds of individuals. Which statement describes how <u>similar</u> they are likely to be to other lizards in the population. a. All lizards are likely to be almost exactly the same. b. All lizards are exactly the same on the inside but display differences in their external features. c. All lizards share many similarities, yet likely have some significant differences in their features. d. All lizards are likely to be the same on the outside but display differences in their internal features.
17. Which statement best describes how traits in lizards will be inherited by offspring? a. When parent lizards learn to catch particular insects, their offspring can inherit their specific insect-catching-skills. b. When parent lizards develop stronger claws through repeated use in catching prey, their offspring can inherit their stronger-claw trait. c. When parent lizards' claws are underdeveloped because easy food sources are available, their offspring can inherit their weakened claws. d. When a parent lizard is born with an extra finger on its claws, its offspring can inherit six-fingered claws.	17. Which statement best describes how traits in lizards will be inherited by offspring? a. When parent lizards learn to catch particular insects, their offspring can inherit their specific insect-catching-skills. b. When parent lizards develop stronger claws through repeated use in catching prey, their offspring can inherit their stronger-claw trait. c. When parent lizards' claws are underdeveloped because easy food sources are available, their offspring can inherit their weakened claws. d. When a parent lizard is born with an extra claw on each limb, its offspring can inherit the extra claw.

18. Fitness is a term often used by biologists to explain the evolutionary success of certain organisms. Below are descriptions of four fictional female lizards. Which lizard might a biologist consider to be the "most fit"?

	Lizard A	Lizard B	Lizard C	Lizard D
Body length	20 cm	12 cm	10 cm	15 cm
Offspring surviving to adulthood	19	28	22	26
Age at death	4 years	5 years	4 years	6 years
Comments	Lizard A is very healthy, strong, and clever	Lizard B has mated with many lizards	Lizard C is dark-colored and very quick.	Lizard D has the largest territory of all the lizards.

a. Lizard A
b. Lizard B
c. Lizard C
d. Lizard D

18. Fitness is a term often used by biologists to explain the evolutionary success of certain organisms. Below are descriptions of four fictional female lizards. Which lizard might a biologist consider to be the "most fit" according to Darwin's theory?

	Lizard A	Lizard B	Lizard C	Lizard D
Body length	20 cm	12 cm	10 cm	15 cm
Offspring surviving to adulthood	19	28	22	26
Age at death	4 years	5 years	4 years	6 years
Additional Facts	Lizard A is very healthy, strong, and clever	Lizard B has mated with many lizards	Lizard C is dark-colored and very quick.	Lizard D has the largest territory of all the lizards.

a. Lizard A
b. Lizard B
c. Lizard C
d. Lizard D

19. According to the theory of natural selection, where did the variations in body size in the three species of lizards most likely come from?
a. The lizards needed to change in order to survive, so beneficial new traits developed.
b. The lizards wanted to become different in size, so beneficial new traits gradually appeared in the population.
c. Random genetic changes and sexual recombination both created new variations.
d. The island environment caused genetic changes in the lizards.

19. According to the theory of natural selection, where did the variation in body size of the three species most likely come from?
a. The lizards needed to change in order to survive, so beneficial new traits developed.
b. The lizards wanted to become different in size, so beneficial new traits gradually appeared in the population.
c. Random genetic changes and new combinations of genes produced through mating both produce variation.
d. The island environment caused genetic changes in the lizards.

20. What could cause one species to change into three species over time? a. Groups of lizards encountered different island environments so the lizards needed to become new species with different traits in order to survive. b. Groups of lizards must have been geographically isolated from other groups and random genetic changes must have accumulated in these lizard populations over time**.** c. There may be minor variations, but all lizards are essentially alike and all are members of a single species. d. In order to survive, different groups of lizards needed to adapt to the different islands, and so all organisms in each group gradually evolved to become a new lizard species.	20. What could cause one species to change into three species over time? a. Groups of lizards encountered different island environments so the lizards needed to become new species with different traits in order to survive. b. Groups of lizards must have been geographically isolated from other groups and random genetic changes must have accumulated in these lizard populations over time**.** c. There may be minor variations, but all lizards are essentially alike and all are members of a single species. d. In order to survive, different groups of lizards needed to adapt to the different islands, and so all organisms in each group gradually evolved to become a new lizard species.

Appendix F: Improvements made by original CINS authors on items that were not the focus of the study

Original CINS items – Not included in interviews	**Revised CINS Items from Anderson, Dwyer & Fisher**
1. What would happen if a breeding pair of finches was placed on an island under ideal conditions with no predators and unlimited food so that all individuals survived? Given enough time a. the finch population would stay small because birds only have enough babies to replace themselves. b. the finch population would double and then stay relatively stable. c. the finch population would increase dramatically. d. the finch population would grow slowly and then level off.	1. What would happen if a breeding pair of finches was placed on an island under ideal conditions with no predators and unlimited food so that all individuals survived? a. The finch population would stay small because birds only have enough babies to replace themselves. b. The finch population would double and then stay relatively stable. c. The finch population would increase dramatically. d. The finch population would grow slowly and then level off.
2. Finches on the Galapagos Islands require food to eat and water to drink. a. When food and water are scarce, some birds may be unable to obtain what they need to survive. b. When food and water are limited, the finches will find other food sources, so there is always enough. c. When food and water are scarce, the finches all eat and drink less so that all birds survive. d. There is always plenty of food and water on the Galapagos Islands to meet the finches' needs.	2. Finches on the Galapagos Islands require food to eat and water to drink. How does this fact impact the population? a. When food and water are scarce, some birds may be unable to obtain what they need to survive. b. When food and water are limited, the finches will find other food sources, so there is always enough. c. When food and water are scarce, the finches all eat and drink less so that all birds survive. d. There is always plenty of food and water on the Galapagos Islands to meet the finches' needs.
3. Once a population of finches has lived on a particular island for many years, a. the population continues to grow rapidly. b. the population remains relatively stable, with some fluctuations. c. the population dramatically increases and decreases each year. d. the population will decrease steadily.	3. Once a population of finches has lived on a particular island for many years, what will most likely happen to the population? a. The population continues to grow rapidly. b. The population remains relatively stable, with some fluctuations. c. The population dramatically increases and decreases each year. d. The population will decrease steadily, then increase.

5. Depending on their beak size and shape, some finches get nectar from flowers, some eat grubs from bark, some eat small seeds, and some eat large nuts. Which statement best describes the interactions among the finches and the food supply? a. Most of the finches on an island cooperate to find food and share what they find. b. Many of the finches on an island fight with one another and the physically strongest ones win. c. There is more than enough food to meet all the finches' needs so they don't need to compete for food. d. Finches compete primarily with closely related finches that eat the same kinds of food**,** and some may die from lack of food.	5. Depending on their beak size and shape, some finches get nectar from flowers, some eat grubs from bark, some eat small seeds, and some eat large nuts. Which statement best describes the interactions among the finches and the food supply? a. Most of the finches on an island cooperate to find food and share what they find. b. Many of the finches on an island fight with one another and the physically strongest ones win. c. There is more than enough food to meet all the finches' needs so they don't need to compete for food. d. Finches compete primarily with closely related finches that eat the same kinds of food**,** and some may die from lack of food.
11. Assuming ideal conditions with abundant food and space and no predators, what would happen if a pair of guppies were placed in a large pond? a. The guppy population would grow slowly, as guppies would have only the number of babies that are needed to replenish the population. b. The guppy population would grow slowly at first, then would grow rapidly, and thousands of guppies would fill the pond. c. The guppy population would never become very large, because only organisms such as insects and bacteria reproduce in that manner. d. The guppy population would continue to grow slowly over time.	11. Assuming ideal conditions with abundant food and space, and no predators, what would happen if a mating pair of guppies was placed in a large pond? a. The guppy population would grow slowly, as guppies would have only the number of offspring that are needed to replenish the population. b. The guppy population would grow slowly at first, then would grow rapidly, and thousands of guppies would fill the pond. c. The guppy population would stay small, because the larger populations would be organisms such as insects and bacteria, since those conditions are more ideal for them. d. The guppy population would continue to grow slowly over time.
12. Once a population of guppies has been established for a number of years in a real (not ideal) pond with other organisms including predators, what will likely happen to the population? a. The guppy population will stay about the same size. b. The guppy population will continue to rapidly grow in size. c. The guppy population will gradually decrease until no more guppies are left. d. It is impossible to tell because populations do not follow patterns.	12. Once a population of guppies has been established for a number of years in a pond with other organisms including predators, what will likely happen to the population if conditions remain constant? a. The guppy population will stay about the same size. b. The guppy population will continue to rapidly grow in size. c. The guppy population will gradually decrease until no more guppies are left. d. It is impossible to tell because populations do not follow patterns.

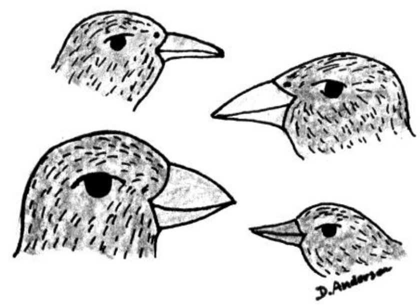
14. Lizards eat a variety of insects and plants. Which statement describes the availability of food for lizards on the Canary Islands? a. Finding food is not a problem since food is always in abundant supply. b. Since lizards can eat a variety of foods, there is likely to be enough food for all of the lizards at all times. c. Lizards can get by on very little food, so the food supply does not matter. d. It is likely that sometimes there is enough food, but at other times there is not enough food for all of the lizards.	14. Lizards eat a variety of insects and plants. Which statement describes the availability of food for lizards on the Canary Islands? a. Finding food is not a problem, because the food most often eaten by the lizards is always in abundant supply. b. Since lizards can eat a variety of foods, there is likely to be enough food for all of the lizards at all times. c. Lizards can get by on very little food, so the amount available on the island does not matter. d. It is likely that sometimes there is enough food, but at other times there is not enough food for all of the lizards.
15. What do you think happens among the lizards of a certain species when the food supply is limited? a. The lizards cooperate to find food and share what they find. b. The lizards fight for the available food and the strongest lizards kill the weaker ones. c. Genetic changes that would allow lizards to eat new food sources are likely to be induced. d. The lizards least successful in the competition for food are likely to die of starvation and malnutrition.	15. What happens with the lizard population when the food supply is limited? a. The lizards will cooperate to find food and share what they find. b. The lizards fight for the available food and the stronger lizards kill the weaker ones. c. Genetic changes that would allow lizards to eat new food sources are more likely to occur. d. The lizards least successful in the competition for food are likely to die of starvation and malnutrition.

Appendix G: Final Draft of the Revised version of the CINS 2011, 5th version

Conceptual Inventory of Natural Selection
5th edition - 2011

Your answers will assess your understanding of the Theory of Natural Selection. Please choose the answer that best reflects how a biologist would think about each question.

Galapagos finches



Scientists have long believed that the 14 species of finches on the Galapagos Islands evolved from a single species of finch that migrated to the islands one to five million years ago (Lack, 1940). Recent DNA analyses support the conclusion that all of the Galapagos finches evolved from the warbler finch (Grant, Grant & Petren, 2001; Petren, Grant & Grant, 1999). Different species live on different islands. For example, the medium ground finch and the cactus finch live on one island. The large cactus finch occupies another island. One of the major changes in the finches is in their beak sizes and shapes as shown in this figure.

Choose the one answer that best reflects how an evolutionary biologist would answer.

1. What would happen if a breeding pair of finches was placed on an island under ideal conditions with no predators and unlimited food so that all individuals survived?
   a. The finch population would stay small because birds only have enough babies to replace themselves.
   b. The finch population would double and then stay relatively stable.
   c. The finch population would increase dramatically.
   d. The finch population would grow slowly and then level off.

2. Finches on the Galapagos Islands require food to eat and water to drink. How does this fact impact the population?
   a. When food and water are scarce, some birds may be unable to obtain what they need to survive.
   b. When food and water are limited, the finches will find other food sources, so there is always enough.
   c. When food and water are scarce, the finches all eat and drink less so that all birds survive.
   d. There is always plenty of food and water on the Galapagos Islands to meet the finches' needs.

3. Once a population of finches has lived on a particular island for many years, what will most likely happen to the population?
   a. The population continues to grow rapidly.
   b. The population remains relatively stable, with some fluctuations.
   c. The population dramatically increases and decreases each year.
   d. The population will decrease steadily, than increase.

4. What is the best way to describe the evolutionary changes that occur in a finch population over time?
   a. The traits of each finch within a population gradually change.
   b. The percentages of finches having different traits within a population change.
   c. Successful behaviors learned by finches are passed on to offspring.
   d. Mutations occur to meet the needs of the finches as the environment changes.

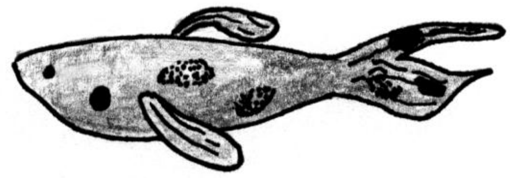
5. Depending on their beak size and shape, some finches get nectar from flowers, some eat grubs from bark, some eat small seeds, and some eat large nuts. Which statement best describes the interactions among the finches and the food supply?
   a. Most of the finches on an island cooperate to find food and share what they find.
   b. Many of the finches on an island fight with one another and the physically strongest ones win.
   c. There is more than enough food to meet all the finches' needs so they don't need to compete for food.
   d. Finches compete primarily with closely related finches that eat the same kinds of food**,** and some may die from lack of food.

6. How did the different beak types <u>first</u> appear in the Galapagos finches?
   a. The changes in the finches' beak size and shape occurred because of their need to be able to eat different kinds of food to survive.
   b. Changes in the finches' beaks occurred randomly, and when there was a good match between beak structure and available food, those birds usually had more offspring.
   c. The changes in the finches' beaks occurred because the environment caused the desired genetic changes.
   d. The finches' beaks changed a little bit in size and shape with each successive generation, some getting larger and some getting smaller because of environmental changes.

7. What type of variation in the finches' traits is passed to the offspring?
   a. Only behaviors that were learned during a finch's lifetime
   b. Only traits that were beneficial during a finch's lifetime
   c. Only traits that were genetically determined
   d. Only traits that were positively influenced by the environment during a finch's lifetime.

8. What caused populations of birds having different beak shapes and sizes to become distinct species distributed on the various islands?
   a. The finches were quite varied, and those whose features were best suited to the available food supply on each island reproduced most successfully.
   b. All finches are essentially alike and there are <u>not</u> really fourteen different species.
   c. Different foods are available on different islands and for that reason, individual finches on each island gradually developed the beaks they needed.
   d. Different lines of finches developed different beak types because they needed them in order to obtain the available food.

Venezuelan Guppies

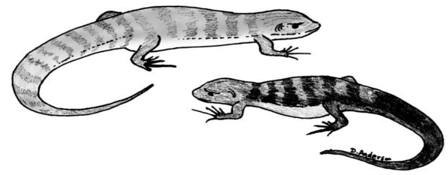


Guppies are small fish found in streams in Venezuela. Male guppies are brightly colored, with black, red, blue and iridescent (reflective) spots. Males cannot be too brightly colored or they will be seen and consumed by predators, but if they are too plain, females will choose other males. Natural selection and sexual selection push in opposite directions. When a guppy population lives in a stream in the absence of predators, the proportion of males that are bright and flashy increases in the population. If a few aggressive predators are added to the same stream, the proportion of bright-colored males decreases within about five months (3-4 generations). The effects of predators on guppy coloration have been studied in artificial ponds with mild, aggressive, and no predators, and by similar manipulations of natural stream environments (Endler, 1980).

Choose the one answer that best reflects how an evolutionary biologist would answer.

9. A natural population of guppies consists of hundreds of fish of a single species. Which statement best describes the population of guppies?
   a. The guppies share all of the same characteristics and are identical to each other.
   b. The guppies share all of the most important characteristics of the species; the small differences between them don't affect survival.
   c. The guppies are all identical on the inside, but have many differences in appearance.
   d. The guppies share most important characteristics, but also have differences that may affect survival.
10. Fitness is a term often used by biologists to explain the evolutionary success of certain organisms. Which characteristics would a biologist consider to be most important in determining which guppies were the "most fit" according to Darwin's theory?
   a. large body size and ability to swim quickly away from predators
   b. excellent ability to compete for food
   c. high number of offspring that survived to reproductive age
   d. high number of matings with many different females.
11. Assuming ideal conditions with abundant food and space, and no predators, what would happen if a mating pair of guppies was placed in a large pond?
   a. The guppy population would grow slowly, as guppies would have only the number of offspring that are needed to replenish the population.
   b. The guppy population would grow slowly at first, then would grow rapidly, and thousands of guppies would fill the pond.
   c. The guppy population would never become very large, because only organisms such as insects and bacteria reproduce in that manner.
   d. The guppy population would continue to grow slowly over time.
12. Once a population of guppies has been established for a number of years in a pond with other organisms including predators, what will likely happen to the population if conditions remain constant?
   a. The guppy population will stay about the same size.
   b. The guppy population will continue to rapidly grow in size.
   c. The guppy population will gradually decrease until no more guppies are left.
   d. It is impossible to tell because populations do not follow patterns.
13. What is the best way to describe the evolutionary changes that occur in a guppy population over time?
   a. The traits of each individual guppy within a population gradually change.
   b. The percentage of guppies having different traits within a population change.
   c. Successful behaviors learned by certain guppies are passed on to offspring.
   d. Mutations occur to meet the needs of the guppies as the environment changes.

Canary Island Lizards



The Canary Islands are seven islands just west of the African continent. The islands gradually became colonized with life: plants, lizards, birds, etc. Three different species of lizards found on the islands are similar to one species found on the African continent (Thorpe & Brown, 1989). Because of this, scientists assume that the lizards traveled from Africa to the Canary Islands by floating on tree trunks washed out to sea.

Choose the one answer that best reflects how an evolutionary biologist would answer.

14. Lizards eat a variety of insects and plants. Which statement describes the availability of food for lizards on the Canary Islands?
    a. Finding food is not a problem since food is always in abundant supply.
    b. Since lizards can eat a variety of foods, there is likely to be enough food for all of the lizards at all times.
    c. Lizards can get by on very little food, so the food supply does not matter.
    d. It is likely that sometimes there is enough food, but at other times there is not enough food for all of the lizards.

15. What do you think happens among the lizards of a certain species when the food supply is limited?
    a. The lizards will cooperate to find food and share what they find.
    b. The lizards fight for the available food and the stronger lizards kill the weaker ones.
    c. Genetic changes that would allow lizards to eat new food sources are more likely to occur..
    d. The lizards least successful in the competition for food are likely to die of starvation and malnutrition.

16. A population of lizards is made up of hundreds of individuals. Which statement describes how similar they are likely to be to other lizards in the population?
    a. All lizards are likely to be almost exactly the same.
    b. All lizards are exactly the same on the inside but display differences in their external features.
    c. All lizards share many similarities, yet likely have some significant differences in their features.
    d. All lizards are likely to be the same on the outside but display differences in their internal features.

17. Which statement best describes how traits in lizards will be inherited by offspring?
    a. When parent lizards learn to catch particular insects, their offspring can inherit their specific insect-catching-skills.
    b. When parent lizards develop stronger claws through repeated use in catching prey, their offspring can inherit their stronger-claw trait.
    c. When parent lizards' claws are underdeveloped because the available prey is easy to catch, their offspring can inherit their weakened claws.
    d. When a parent lizard is born with an extra claw on each limb, its offspring can inherit the extra claw.

18. Fitness is a term often used by biologists to explain the evolutionary success of certain organisms. Below are descriptions of four fictional female lizards. Which lizard might a biologist consider to be the "most fit" according to Darwin's Theory?

	Lizard A	Lizard B	Lizard C	Lizard D
Body length	20 cm	12 cm	10 cm	15 cm
Offspring surviving to adulthood	19	28	22	26
Age at death	4 years	5 years	4 years	6 years
Additional Facts	Lizard A is very healthy, strong, and clever	Lizard B has mated with many lizards	Lizard C is dark-colored and very quick.	Lizard D has the largest territory of all the lizards.

a. Lizard A
b. Lizard B
c. Lizard C
d. Lizard D

19. According to the theory of natural selection, where did the variation in body size of the three species most likely come from?
a. The lizards needed to change in order to survive, so beneficial new traits developed.
b. The lizards wanted to become different in size, so beneficial new traits gradually appeared in the population.
c. Random genetic changes and new combinations of genes produced through mating both produce variation.
d. The island environment caused genetic changes in the lizards.

20. What could cause one species to change into three species over time?
a. Groups of lizards encountered different island environments so the lizards needed to become new species with different traits in order to survive.
b. Groups of lizards may have been geographically isolated from other groups and random genetic changes may have accumulated in these lizard populations over time.
c. There may be minor variations, but all lizards are essentially alike and all are members of a single species.
d. In order to survive, different groups of lizards needed to adapt to the different islands, and so all organisms in each group gradually evolved to become a new lizard species.

Appendix H: Factor Analysis-Matrices

Rotated Component Matrix for the original version, non-majors only.

	Component							
	1	2	3	4	5	6	7	8
1	.086	.318	.009	.077	.155	.203	-.025	.656
11	-.112	.124	-.023	-.039	-.121	.840	.241	.042
2	.756	-.136	.160	-.286	.075	-.102	.122	-.030
14	.431	.104	-.226	.356	-.028	.134	.448	-.240
3	.856	-.016	-.113	.088	.036	-.081	.072	.199
12	.405	.282	.322	-.187	-.047	.291	-.312	.339
5	.252	.029	.335	-.674	-.007	.120	.065	.095
15	.481	.234	.429	.009	-.416	.219	-.122	-.131
9	.069	.011	.102	-.157	-.007	.164	.828	.116
16	-.050	-.035	.911	-.007	.152	-.025	.046	-.027
7	.742	.166	.006	-.005	-.065	-.038	-.010	-.192
17	.196	.511	.419	.032	-.352	-.366	.188	-.111
10	.497	-.042	.078	.173	.342	.429	-.360	.059
18	.228	.288	.112	.076	.269	.207	-.191	-.676
4	.084	.198	.091	-.012	.874	-.112	-.054	-.022
13	-.186	.531	.228	-.013	.491	.318	.200	-.080
6	.212	.717	.202	.029	.136	-.194	.190	.039
19	-.027	.694	-.031	.311	.058	.130	-.130	.252
8	.088	.075	.207	.869	-.007	.039	-.074	.081
20	-.044	.775	-.188	-.137	.079	.208	-.084	-.070

Rotated Component Matrix for the original version, non-majors and freshmen majors.

	Component						
	1	2	3	4	5	6	7
1	.028	.098	.079	.189	-.007	-.015	.877
11	-.115	-.058	-.069	.734	-.015	.331	.133
2	.760	.021	.182	-.006	-.313	.085	-.119
14	.389	-.009	.083	.068	.408	.578	-.162
3	.835	-.042	.052	.040	.121	.139	.057
12	.253	.063	.290	.673	-.053	-.275	.069
5	.233	.158	.329	.274	-.640	.170	-.153
15	.369	-.073	.664	.233	.076	-.038	.094
9	.069	.164	.140	.063	-.160	.801	.032
16	-.055	.515	.554	-.068	-.197	.008	-.180
7	.606	.106	.186	-.167	.027	-.007	.460
17	.058	.028	.760	-.008	.149	.289	.121
10	.474	.384	-.130	.401	.171	-.165	-.189
18	.211	.345	.238	.283	.308	-.033	-.320
4	.170	.843	-.119	-.084	.034	-.011	.194
13	-.135	.699	.162	.255	.088	.351	-.007
6	.097	.453	.421	.178	.246	.437	.048
19	-.114	.370	.168	.372	.444	.071	.137
8	.086	.158	.221	.111	.724	.054	-.114
20	-.096	.428	.076	.528	.229	.270	-.021

Rotated Component Matrix, for the revised version of the CINS, taken by freshman, junior, and senior majors.

	Component							
	1	2	3	4	5	6	7	8
1	.428	.212	.002	.258	-.336	.019	.608	-.031
11	.469	.600	.281	.053	-.148	.235	-.228	-.091
2	-.184	-.184	.281	.561	.375	.410	.047	-.054
14	.189	.023	-.029	.274	-.032	.776	.051	-.050
3	-.147	.185	.107	-.084	.035	-.104	.796	.070
12	.568	.475	.279	-.143	.254	.004	-.174	-.118
5	.847	.232	.087	-.057	.236	.027	-.039	-.043
15	.859	.015	-.116	-.007	.103	.181	.063	-.043
9	.113	-.055	-.059	.755	-.010	.157	.070	.325
16	-.019	-.094	.064	.254	-.031	-.437	-.303	.683
7	.075	.271	.083	-.031	.201	.808	-.226	.102
17	.236	-.173	.565	.403	-.094	.150	.201	.174
10	.318	.244	-.064	-.107	.739	.017	-.043	-.011
18	.105	.030	.149	.143	.878	.119	-.024	-.131
4	-.007	.119	.900	.054	.027	-.044	.031	-.028
13	-.021	.113	.877	-.132	.117	.037	.005	-.012
6	.090	.807	-.023	-.164	.201	.085	.306	.143
19	.145	.880	.090	.156	.072	.094	.207	-.011
8	-.140	.114	-.017	-.106	-.127	.166	.206	.813
20	-.178	.149	-.016	.805	-.003	-.015	-.121	-.265

Rotated component matrix for the revised version, taken by non-majors

	Component					
	1	2	3	4	5	6
1	.132	.039	-.038	.868	.179	-.021
11	.337	-.227	.022	.485	-.077	.618
2	.057	.159	.330	.709	.071	.020
14	-.111	.052	.269	-.118	.158	.804
3	.070	.137	.837	.060	.010	.078
12	.189	.377	.447	.416	.154	-.103
5	.307	-.020	.832	.044	-.056	.172
15	.624	.400	-.207	.169	-.361	.195
9	.610	-.272	.076	.257	.258	-.240
16	.800	-.224	.183	-.024	.265	-.179
7	.101	-.076	.627	.478	.212	.097
17	.435	.432	.123	.175	.162	.148
10	.751	.038	.456	-.002	-.080	.245
18	.727	.236	.176	.171	.012	.050
4	-.220	.845	-.069	-.111	-.071	-.089
13	.129	.848	.029	.190	.276	-.068
6	.494	.589	.067	.046	.510	.113
19	.284	.262	.054	.089	.727	.336
8	.051	.800	.178	.058	.188	.079
20	-.044	.221	.003	.272	.839	-.064

References

Abedi, J., Courtney, M., Mirocha, J., Leon, S., & Goldberg, J. (2005). Language accommodations for English language learners in large-scale assessments: Bilingual dictionaries and linguistic modification. *National Center for Research on Evaluation, Standards, and Student Testing (CRESST).* University of California, Los Angeles.

Alters, B., & Nelson, C. (2002). Perspective: Teaching evolution in higher education. *Evolution International Journal of Organic Evolution*, 56 (10), 1891-1901.

Anderson, A., Fisher, K., & Norman, G. (2002). Development and evaluation of the Conceptual Inventory of Natural Selection. *Journal of Research in Science Teaching*, 39 (10), 952-978.

Anderson, D., Fisher, K., & Smith, M. (2010). Support for the CINS as a diagnostic conceptual inventory: Response to Nehm and Schonfeld (2008). *Journal of Research in Science Teaching*, 47 (3), 354-357.

Battisti, B., Hanegan, N., Sudweeks, R., & Cates, R. (2010). *International Journals of Science and Mathematics Education*, 8 (5), 845-868.

Bishop, B., & Anderson, C. (1990). Student conceptions of natural selection and its role in evolution. *Journal of Research in Science Teaching*, 27 (5), 415-427.

Blanchard, M., Southerland, S., Osborne, J., Sampson, V., Annetta, L., & Granger, E. (2010). Is inquiry possible in light of accountability?: A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction. *Science Education*, 94 (4), 577-816.

Bock, R. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometricka*, 37 (1), 29-51.

Bock, R. (1997). The nominal categories model. In W.J. van der Linden and R.K. Hambleton (Eds), *Handbook of modern item response theory* (pp. 23-49). New York: Springer.

Catley, K. (2006). Darwin's missing link – A novel paradigm for evolution education. Science Education, 90 (5), 767-783.

Chinsamy, A., & Plagányi, E., (2008). Accepting Evolution. *Evolution*, 62 (1), 248-254.

Crawford, B., Zembal-Saul, C., Munford, D., & Friedrichsen, P. (2010.) Confronting prospective teachers' ideas of evolution and scientific inquiry using technology and inquiry-based tasks. *Journal of Research in Science Teaching,* 42 (6), 613-637.

Dagher, Z., & BouJaoude, S. (2005). Students' perceptions of the nature of evolutionary theory. *Science Education,* 89 (3), 378-391.

Ereshefsky, M. (2010). Mystery of mysteries: Darwin and the species problem. *Cladistics*, 26 (Early View), 1-13.

Ercikan, K., Arim, R., Law, D., Domene, J., Gagnon, F., & Lacroix, S. (2010). Application of think-aloud protocols for examining and confirming sources of differential item functioning identified by expert reviews. *Educational Measurement: Issues and Practice*, 29 (2), 24-35.

Gould, S. J. (2007)$_a$. *The Richness of Life: The essential Stephen Jay Gould*. New York: W.W. Norton and Company.

Gould, S. J. (2007)$_b$. *Punctuated Equilibrium*. Massachusetts: The Belknap Press of Harvard University Press.

Grehan, J. (2001). Biogeography and evolution of the Galapagos: Integration of the biological and geological evidence. *Biological Journal of the Linnaean Society*, 74 (3), 267-287.

Hewson, P., Tabachnick, R., Zeichner, K., Blomker, K., Meyer, H., Lemberger, J., Marion, R., Park, H. J., & Toolin, R. (1999). Educating prospective teachers of biology: Introduction and research methods. *Science Education*, 83 (3), 247-273.

Hokayem, H., & BouJaoude, S. (2008). College students' perceptions of the theory of evolution. *Journal of Research in Science Teaching*, 45 (4), 395-419.

Jensen, M., & Finley, F. (1996). Changes in students' understanding of evolution resulting from different curricular and instructional strategies. *Journal of Research in Science Teaching*, 33 (8), 879-900.

Larsson, A., & Hallden, O. (2010). A structural view on the emergence of a conception: Conceptual change as radical reconstruction of contexts. *Science Education*, 94 (4), 640-664.

Lewens, T. (2010). Natural selection then and now. *Biological Reviews*, 85 (4), 829-835.

Mayr, E. (1982). The growth of biological thought: Diversity, evolution and inheritance. Cambridge, MA: Harvard University Press.

Meece, J., & Daniels, D. (2008). Child & adolescent development for educators. (pgs. 129-164) New York, NY: McGraw Hill

Nehm, R., Kim, S., & Sheppard, K. (2009) Academic preparations in biology and advocacy for teaching Evolution: Biology versus non-biology teachers. *Science Education*, 93 (6), 1122-1146.

Nehm, R., & Schonfeld, I.S. (2008). Measuring knowledge of natural selection: A comparison of the CINS, and open-response instrument, and an oral interview. *Journal of Research in Science Teaching*, 45 (10), 1131-1160.

Nehm, R., & Schonfeld, I. (2010). The future of natural selection knowledge measurement: A reply to Anderson et al. (2010). *Journal of Research in Science Teaching*, 47 (3), 358-362.

O'Donnell, A., Reeve, J., & Smith, J. (2007) Educational psychology: Reflection for action. Custom edition for Arizona State University. (pgs 37-40) Hoboken, NJ: John Wiley & Sons, Inc.

Piaget, J. (1963). Origins of intelligence in children. New York: Norton.

Sadler, P. (1998). Psychometric models of student conceptions in science: Reconciling qualitative studies and distractor-driven assessment instruments. *Journal of Research in Science Teaching*, 35 (3), 265-296.

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores [psychometric monograph, 17]*. Iowa City: Psychometric Society

Southerland, S., Smith, M., & Cummins, C. (1999). "What do you mean by that?: Using structured interviews to assess science understanding". In: *Assessing Science Understanding: A Human Constructivist View*, Chapter 4, pgs. 122-133.

Taber, K., (2001). Shifting sands: a case study of conceptual development as competition between alternative conceptions. *International Journal of Science Education*. 23 (7), 731-753.

Tanner, K., & Allen, D. (2005). Approaches to biology teaching and learning: Understanding the wrong answers-teaching toward conceptual change. *Cell Biology Education*. 4 (2), 112-117.

Thissen, D., & Steinberg, L. (1984). A response model for multiple-choice items. *Psychometrika*, 49 (4), 501-519.

Thissen, D., & Steinberg, L. (1997). A response model for multiple-choice items. *Handbook of modern item response theory* (pp. 51-65). New York: Springer.

Tsui, C. & Treagust, D. (2007). Understanding genetics: Analysis of secondary students' conceptual status. *Journal of Research in Science Teaching*, 44 (2), 747-785.

Weiner, J. (1995). *The Beak of the Finch*. New York: Vintage Books.

Wesson, R. (1991). *Beyond Natural Selection*. Massachusetts: MIT

Witting, L. (2008). Inevitable evolution: Back to *The Origin* and beyond the 20th century paradigm of contingent evolution by historical natural selection. *Biological Reviews*, 83 (3), 259-294.

Vosniadou, S. (2007). Conceptual Change and Education. *Human Development*, 50 (1), 47-54.

Zimmer, C. (2001). *Evolution: The Triumph of an Idea*. New York: HarperCollins Books.

Zirbel, E. (2004) Framework for conceptual change. *Astronomy Education Review*, 3 (1), 62-76.