Curricular reform and inquiry teaching in biology: where are our efforts most fruitfully invested?‡

Briana E. Timmerman,1,* Denise C. Strickland*† and Susan M. Carstensen*

†Department of Biological Science, University of South Carolina, Columbia, SC 29208, USA; †Department of Instruction and Teacher Education, University of South Carolina, Columbia, SC 29208, USA

Synopsis University faculty often express frustration with the accuracy of students’ understanding of science in general and of evolution in particular. A rich research literature suggests that inquiry-based pedagogies are more effective in producing meaningful learning than are traditional, didactic approaches. A pragmatic investigation into the efficacy of inquiry-based curricular reforms compared to traditional laboratory activities was undertaken in the introductory biology course for majors at a large state university in the southeastern United States. The topics of the course focused on biodiversity, evolution, and plant and animal anatomy and physiology. Students’ learning in the inquiry versus traditional units was compared using both a test of pre–post content knowledge as well as open-ended written responses in which students described events in which there was meaningful learning and conceptual changes. The pre–post tests were replicated over five semesters of the same course (n=1493 students). Students’ misconceptions as well as examples of meaningful learning were gathered for two semesters in the same course (n=518 students). Results consistently revealed that descriptive, concrete topics such as anatomy can be taught effectively using traditional didactic methods; average effect sizes (a measure of the difference between pretest scores and posttest scores) range from 1.8 to 2.1. The inquiry units also increased knowledge of content on the topics of evolution and biodiversity by a significant degree (average effect sizes range from 1.0 to 1.1), despite the fact that students spent less than half the instructional time on these units compared to the didactic units. In addition, a literature review indicated that highly abstract or mathematical concepts such as evolution or geologic time require greater formal reasoning ability and that students often show lesser gains in these areas compared to more concrete topics. It was therefore especially notable that the frequency of meaningful learning events was significantly higher in the units on evolution compared to the traditional units (χ² P<0.5 to 0.001). A catalog of students’ misconceptions (some of which were quite unexpected) was also generated and found useful for future teaching. Therefore, we feel that when time and resources for curricular reform are limited, those efforts should prioritize abstract and foundational topics such as evolution. Didactic teaching appears sufficient for more concrete topics such as anatomy.

Introduction

University science faculty express dissatisfaction with the level of understanding and skills achieved by students in traditional course settings (Bowen et al. 1999; Alters and Nelson 2002) and many have proposed inquiry-based pedagogies to address these needs across a wide variety of fields (Hake 1998; Birk and Kurtz 1999; Bell 2001; Alters and Nelson 2002; Powell 2003; Prince 2004; Kalinowski et al. 2006). For biology in particular, the recent report by the National Research Council (NRC), Bio 2010 (Committee on Undergraduate Biology Education 2003) emphasizes the importance of how students are taught, as well as the content addressed in courses. Bio 2010 strongly advocates the use of an inquiry-based approach and recommends reforming laboratory experiences from traditional, confirmational and “cookbook” experiments to experiences that “cultivate the ability of students to think independently” (p. 75). Specifically the report describes the use of “project-based laboratories with discovery components” thereby stimulating students’ motivation and participation as well as providing students with opportunities for “developing their scientific writing, speaking and presentation skills” (p. 75).

From the symposium “Evolution vs. Creationism in the Classroom: Evolving Student Attitudes” presented at the annual meeting of the Society for Integrative and Comparative Biology, January 2–6, 2008, at San Antonio, Texas.

E-mail: timmerman@schc.sc.edu


Advanced Access publication June 24, 2008

© The Author 2008. Published by Oxford University Press on behalf of the Society for Integrative and Comparative Biology. All rights reserved. For permissions please email: journals.permissions@oxfordjournals.org.
These emphases are backed by rich literature of research demonstrating that inquiry-based pedagogies promote deeper understanding of content and/or greater skills in scientific reasoning (Shymansky et al. 1983; Jensen and Finley 1996; Johnson and Lawson 1998; Christianson and Fisher 1999; Musheno and Lawson 1999; Abd-El-Khalick and Lederman 2000; Cabrera et al. 2001; Pelaez 2002; Schneider et al. 2002; Marx et al. 2004). A recent meta-analysis indicates collaborative, inquiry-based pedagogies that provide a personally relevant context for learning are consistently effective at generating learning (Schroeder et al. 2007). The power of inquiry-based learning appears to reside in the opportunities it provides for students to confront their prior ideas and misconceptions (Guzzetti et al. 1993) as well as to investigate personally relevant questions in a meaningful way, thereby increasing motivation for learning (Mistler-Jackson and Songer 2000; van Berkel and Schmidt 2000).

Despite the general consensus that inquiry-based pedagogies improve students’ learning, many university faculty and departments are still constrained by the multiple demands of teaching, research, service, and graduate student mentoring. Reform of curricula requires significant time and effort, particularly when it alters how a course is taught rather than just what is taught. In situations in which faculty do not have unlimited time to apply to curricular reform, it would be beneficial for them to have a sense of which areas of biological curricula are most amenable to reform and would provide the greatest benefit for the time and effort invested.

The goal of this article is to describe the lessons learned during efforts to reform an introductory biology laboratory course at a large research-oriented university in the United States. In particular, we discuss the gains in students’ learning that occurred with didactic versus inquiry-based teaching styles and suggest how efforts can be most fruitfully prioritized. Within this article, we provide a chronological description of the process of reform including: (1) a brief description of new curricula, (2) methods and results of quantitative assessment of gains in knowledge of the course’s content, (3) a description of how course context and assumptions can affect interpretation of data on student learning, (4) methods and insights provided by students’ descriptions of meaningful learning events, and (5) implications and recommendations for others about revising biology courses.

**Context of the study**

The curricular reforms described here occurred in the laboratory portion of the second semester of an introductory biology sequence for majors. Enrollment in the course was large (average $n = 329 \pm 46$ students per semester over the course of the study), predominantly female (66%), Caucasian (66%), and composed mainly of freshmen (63%) and sophomores (23%). The largest proportion of students was biology (44%) or other health-related majors (e.g. pharmacy, exercise science) who were required to take the course (an additional 39%). One third to one half of the students were also declared pre-meds. Students were concurrently enrolled in one of 15–18 laboratory sections with no more than 24 students per section. Graduate teaching assistants (TAs) typically taught two laboratory sections each (48 students total). TAs were predominantly first-year or second-year doctoral candidates in biology and had varying levels of teaching experience. Weekly preparatory meetings (3 h each) were led by a full-time faculty laboratory coordinator (the third author) who provided logistical and pedagogical support to the TAs and was responsible for the overall academic integrity of the course.

**Design of the new curricula**

The original laboratory curriculum focused on three major topics: (1) survey of biological diversity, (2) plant anatomy and physiology, and (3) animal anatomy and physiology. These were covered using a traditional didactic approach focused on memorization of factual content. For example, the biodiversity unit was a “march through the phyla” taught using plasticized, dried, or otherwise preserved specimens from each of the major phyla of plants and animals. Plant physiology was largely taught via demonstrations by the instructor or by dissection of flowers and fruits. Animal anatomy and physiology were taught using a guided six-week dissection of chemically preserved, latex-injected white rats. Evolution was not part of the laboratory curriculum.

The curricular reforms reported here focused on two areas: (1) re-shaping the unit on biodiversity to focus on students’ understanding of the ecological consequences of biodiversity and (2) constructing a unit on evolution. In addition, reformed curricula required students to find and use appropriate primary literature and engage in formal, anonymous peer review. The units on plant and animal anatomy and physiology remained unaltered except that the time spent on mammalian dissection was reduced from six to three weeks and the use of a supplemental digital dissection was added (*The Rat: A Functional Anatomy*, Films for the Humanities and Sciences, Princeton NJ, www.films.com). The units...
on plants and animals comprise what is hereafter referred to as the “traditional, didactic” units.

The new inquiry-based laboratory units (biodiversity and evolution) were designed to be more project-based and to conform to the National Research Council (1996) criteria for inquiry. For example, new curricular materials were student-centered and engaged students in authentic scientific practices such as collection and analysis of data while emphasizing social discourse (including peer review) as a mechanism for meaningful learning. The new units therefore differ from the original laboratories in multiple ways: (1) context, (2) level of inquiry, (3) degree of collaboration, and (4) type of assessment (Table 1).

In the unit on evolution, students first engaged in a hands-on simulation of the effect of phenotypic variation on organisms’ foraging success and resulting natural selection. They then generated and tested their own hypotheses concerning the cause and consequence of a large mortality event in a population of finches. This investigation used a multi-year dataset of phenotypic measurements, data on mortality, and field notes gathered by Rosemary and Peter Grant that have been integrated into the freeware (Who will survive? Galápagos Finches. Brian Reiser, Northwestern University, www.letus.org/bguile). This software program includes the first five years of a 30-year dataset by Grant and Grant (2002). Student investigations were then written in the format of a scientific report which was formally peer reviewed using the Calibrated Peer Review system (http://cpr.molsci.ucla.edu) prior to being submitted for a grade.

In the unit on biodiversity, students selected an organism from a list compiled by the instructor of ecologically compatible species and researched the organism’s basic ecology. They then generated a research question featuring their organism and attempted to answer it using the available primary literature. Through oral presentations and subsequent discussion, students constructed an understanding of basic ecological connections among organisms in different kingdoms and phyla. Students received peer feedback on their oral presentations prior to writing their final reports. This unit also contains a specific activity in which students identified primary literature and distinguished it from other sources of information.

### Methods

#### General

This project underwent Institutional Review Board approval as part of a larger assessment project for the Biological Sciences Department and was ruled to be exempt. Students were still informed of the nature, goals, and intent of the project, however, by a description in the course syllabus.

#### Test of students’ knowledge of content

Changes in students’ knowledge of course topics were measured by a test of our own design.
Collection of meaningful learning events through open-ended response items

In addition to the content-knowledge test, in two semesters students were asked to identify a particularly meaningful learning moment. Specifically, they were asked to describe a misconception they held prior to the class that they had subsequently corrected as a result of the course. They were also asked to define their new understanding to ensure the new ideas were scientifically accurate, and to explain what aspect or specific event of the course stimulated the learning or conceptual change. These questions were posed as open-ended written items worth a total of approximately 5% of the final exam grade. Students were informed that points would be awarded for any appropriately detailed response. More than 99% of the students provided appropriately detailed responses (receiving full credit). No topic was suggested for this question; students were free to describe any learning event from the course that leapt to mind.

Four aspects of these open-ended responses were evaluated: (1) content (topic) of the response, (2) nature of the original concept or misconception, (3) nature of the new concept, and (4) degree of awareness of what caused the conceptual change (often referred to as “metacognition,” e.g. Baird and White 1996; Keys et al. 1999). As the authors reviewed the responses, a coding scheme emerged based on the degree of sophistication and specificity of the students’ ideas (Tables 2, 3, and 4). Data for each semester were coded by three independent raters selected from a pool of five. Three of the raters were the authors and two were graduate assistants hired and trained for the specific purpose of coding. Whenever any two of the three raters were in exact agreement, the mutually agreed score was assigned for that response (Johnson et al. 2000). Exact agreement occurred 72–88% of the time depending on the aspect of response being evaluated. Disputed responses were subsequently scored by open discussion, leading to consensus among coders (Johnson et al. 2005) or dropped from the analysis if consensus could not be reached (<3%).

Statistical analysis

As the coding of students’ meaningful learning events generated frequencies within discrete categories, Chi-square tests were used in all cases for determining statistical significance. Expected values were determined by assuming that the types of responses would be equally distributed among curricular units. Distribution of responses among topics was recalculated for each Chi-square test to account for minor differences in the sample size due to the exclusion of disputed scores (responses on which raters did not agree). In all cases, an initial Chi-square test was run to test for bias in that distribution of disputed responses among topics, units, and categories. Such bias was never found (max \( \chi^2 = 1.42 \)); in all cases the distribution of disputed scores among topics and categories was indistinguishable from chance. Thus, the exclusion of disputed responses should not impact the interpretation of results.
Knowledge of content

Notable gains were made in all topics (overall average effect sizes per unit ranged from 1.0 to 2.1, Table 5) demonstrating that the average student improved by one to two standard deviations from pretest to posttest. The inquiry-based segment on the primary literature, when separated from the overall unit on biodiversity, was particularly successful and produces an effect size of 3.7 for the five semesters. Contrary to our expectations, however, the didactic units (plant and animal anatomy and physiology) showed greater gains (average effect size of 1.8–2.1) than did the inquiry-based units of biodiversity and evolution (average effect sizes of 1.0–1.1). The greater gain in the didactic units occurred consistently across all five semesters (Table 5). The experimental design planned for this study assumed that time-on-task and conceptual difficulty of the four topics were equivalent. Post hoc discussion with faculty involved in the course and a review of the literature indicated that neither of these assumptions was valid. The subsequent impact of these differences among curricular units is discussed.

Unanticipated differences among curricular units relevant to the interpretation of results

Factor 1: “Time-on-task” (amount of time spent on a task)

The broad-scale patterns reconstructed from syllabi and discussions with faculty indicated that the time allocated to each curricular unit was not equivalent and that more time was spent on the topics in the didactic units (Fig. 1). The total amount of instructional time (excluding time spent administering tests and handling other administrative tasks)

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>No prior idea on this topic; no conceptual change required for learning (“blank slate”)</td>
<td>“I didn’t know anything about plants.” (Plants 236171)</td>
</tr>
<tr>
<td>Vague</td>
<td>Prior idea vague or source undefined, but an incorrect conception is identified.</td>
<td>“Dinosaurs and humans existed on the planet at the same time.” (Biodiversity 632528)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“I thought that blood just kind of moved around the body without any structure or path.” (Animals 820228)</td>
</tr>
<tr>
<td>Well developed</td>
<td>Prior idea includes a rationale or the source of the prior idea is identified. Student may indicate emotional investment in prior idea.</td>
<td>“I thought individuals could evolve through mutation. I have seen movies and TV where people change because of mutations.” (Evolution 558528)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“[I thought] that rats did not have a similar uterus to humans because they are lower animals.” (Animals 652959)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vague</td>
<td>New idea is vague or source is explicitly defined as coming from the instructor</td>
<td>“I learned that not all flowers are the same.” (Plants 236171)</td>
</tr>
<tr>
<td>Concrete</td>
<td>New idea is specific and examples remain within the context of the course</td>
<td>“Humans and dinosaurs existences were separated by millions of years.” (Biodiversity 632528)</td>
</tr>
<tr>
<td>Generalized</td>
<td>New idea is a generalization. Example(s) given of new idea extend the concept beyond the context of the course.</td>
<td>“Evolution is an ongoing process which is driven by selective mating, genetic drift, mutation, migration and environmental changes. It is a constant process which never ceases and it doesn’t simply result in new species like a train arriving at a final destination. Evolution is never satisfied.” (Evolution 615686)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“I now know that flowers are categorized into two distinct groups, monocots and dicots. Differences between the two include the leaf arrangement and number, the arrangement of the vascular bundles in the stem and roots of each and the seed characteristics of both... These differences may seem minor, but Biology lab has taught me that flowers are very different.” (Plants 786873)</td>
</tr>
</tbody>
</table>

Table 2 Coding rubric for sophistication of students’ prior concepts

Table 3 Coding rubric for sophistication of students’ new concepts

Results

Knowledge of content

Table 5
in laboratory and lecture was calculated for each semester and divided by the number of major topics to provide the expected time-on-task if all topics were covered equivalently. The time actually spent was then subtracted from the expected. Results were averaged for multiple semesters. Statistical tests were not performed as these were posthoc estimates rather than intentional measurements.

The discrepancies in time spent on different topics in the laboratory portion of the class were largely due to the time lost in teaching the students unanticipated skills (how to use computer software, how to logon and navigate the websites, safety procedures, etc.). Given the chronology of the class, these lost class periods detracted primarily from the laboratory periods allocated to the inquiry-based topics. The lecture portion of the course has traditionally been relatively independent of the laboratory and the allocation of time to topics in the lecture is predominantly a product of the lecturer’s sense of priority. Lecture time-on-task was heavily skewed toward plant and animal
Therefore, it is clear that our first assumption was invalid; time-on-task was not equivalent among topics nor among curricular units. Globally, more time was spent on the topics taught in a didactic fashion than on the topics taught in an inquiry-based manner across both lecture and laboratory. Thus, the larger gains in the plant and animal units are to a certain degree expected as students had greater opportunities to learn that material.

**Factor 2: Relative conceptual difficulty of topics**

Review of the literature on science education research illustrated that some scientific concepts are harder for students to learn than are others. For example, students’ prior conceptions of evolution and natural selection appear to be particularly difficult to alter (Greene 1990; Lord and Marino 1993; Demastes et al. 1995; Jensen and Finley 1996; Dagher and BouJaoude 1997; Downie and Barron 2000; Brem et al. 2003; Stern 2004). Obstacles to a sophisticated understanding of evolution appear to fall into three broad categories: (1) difficulty with the abstract and mathematical nature of the topic (Lawson 1983; Lawson et al. 2000a, 2000b; Zimmerman 2000), (2) the wealth of alternative conceptions students bring into the classroom gathered from a lifetime of movies, media, and other informal learning (Deadman and Kelly 1978; Bishop and Anderson 1990; Greene 1990; Lord and Marino 1993; Jensen and Finley 1996; Anderson et al. 2002; Moore et al. 2002; Blackwell et al. 2003; Hokayem and BouJaoude 2008), and (3) the tenacity of those alternative conceptions due to the perceived conflict between evolutionary theory and social or religious values (Demastes et al. 1995; Dagher and BouJaoude 1997; Downie and Barron 2000; Newport 2007; Brem et al. 2003; Ingram and Nelson 2006; Deniz et al. 2008). The greater difficulty experienced by students in learning abstract concepts is also seen in other subjects such as geologic time (Dodick and Orion 2003), chemical bonding (Birk and Kurtz 1999, p. 28), and meiosis (Kindfield 1994; Wynne 2001).

Lawson et al. (2000a, 2000b) provided a tested framework for evaluating the conceptual difficulty of topics in introductory biology. Their work demonstrated that abstract, theoretical topics were more difficult for students to learn than were concrete, descriptive topics. In the study by Lawson et al. (2000a), an entire semester course was taught using inquiry-based pedagogies for all topics (e.g. Lawson 2008) and they found that the type of topic (theoretical, hypothetical, or descriptive) significantly affected students’ achievement; theoretical topics were the most challenging. When the topics tested by our pre/post tests of knowledge of content were categorized using Lawson et al.’s framework, a similar pattern emerged (Table 6). The topics taught by the didactic instructional approach also tended to be more descriptive and concrete (and therefore were likely to be more easily learned) than those taught by the inquiry-based approach which included most of the theoretical or highly abstract concepts in the course.

Additionally, note that the single largest effect size occurred in the primary literature portion of the biodiversity unit, a concrete topic taught using an inquiry-based pedagogy. This subtopic had an effect size that was two to three times larger than that of any other topic (Table 5) over the five replicated semesters. These results concur with those found by Lawson et al. (2000a, 2000b), thereby suggesting that the conceptual difficulty of a topic significantly affects students’ learning of content. Therefore, our second assumption was determined to be false; the topics varied in their conceptual difficulty and the topics selected for the inquiry-based units were more abstract and therefore were more challenging. The larger effect sizes for the traditional, didactic units were therefore at least partially due to the fact that when taught by us they were more descriptive, concrete topics that were easier for students to learn. The fact that the largest gains of any topic occurred in the primary literature section of the test also suggests that when inquiry-based approaches were applied to descriptive or concrete topics, gains may be larger than with didactic approaches.
The content knowledge test is an objective measure of students' gains in knowledge. The second half of this study investigated student-identified meaningful learning events. The frequency of topics volunteered by students roughly mirrored the time spent on the various topics; the didactically taught topics (plant and animal anatomy and physiology) comprised slightly more than half of the responses (31 and 27% of the total responses, respectively and 58% combined; Table 7). The remaining responses were split between evolution (21%), biodiversity content (17%), and primary literature/peer review (4%). Of note, the relative percentage of topics reported did not mirror the sequence of topics over the course of the semester (e.g. animals was the last topic taught and therefore the freshest in students’ minds, but did not comprise the largest percentage of responses).

Student-identified meaningful learning events

The content knowledge test is an objective measure of students’ gains in knowledge. The second half of this study investigated student-identified meaningful learning events. The frequency of topics volunteered by students roughly mirrored the time spent on the various topics; the didactically taught topics (plant and animal anatomy and physiology) comprised slightly more than half of the responses (31 and 27% of the total responses, respectively and 58% combined; Table 7). The remaining responses were split between evolution (21%), biodiversity content (17%), and primary literature/peer review (4%). Of note, the relative percentage of topics reported did not mirror the sequence of topics over the course of the semester (e.g. animals was the last topic taught and therefore the freshest in students’ minds, but did not comprise the largest percentage of responses).

Did these misconceptions vary in sophistication among topics?

As the resistance of old ideas to change is often related to the extent and development of those alternative conceptual frameworks, students’ understandings of these basic topics were characterized by their degree of sophistication. Most (85%) were categorized as being Vague (n = 488 responses) with the remaining 13.5% being coded as Well-developed (see Table 2 for definition of categories). Well-developed concepts did not appear significantly more often within any given topic at the α = 0.05 level, although this result is borderline (χ² = 8.81, critical value is 9.488, Fig. 2). The topic of evolution was the most common well-developed prior idea (Fig. 2) despite being a minority of the topics mentioned (21% of total responses).

It would therefore be expected that students’ learning in the area of evolution would encounter greater challenges than would be the case for other topics. Specifically, the expectation based on past research would be for the evolution unit to produce smaller effect sizes and yield fewer meaningful learning events than would the other topics tested.

To address the issue of which instructional strategy (inquiry-based or traditional pedagogies) was more likely to cause meaningful learning, we turned to the examples of learning reported by the students at the end of each semester.

Is one instructional approach more likely to produce generalizable knowledge?

Students’ new concepts were coded using a schema similar to the codes for their prior ideas; new ideas could be Vague, simple restatements of what the instructor had said (14% of total responses), Concrete, new ideas were specific but remained within the context of the course (55%), or Generalized, new ideas indicated some explicit transfer of a scientific concept to novel material or to situations outside of the context of the course, (31%). These generalized new ideas are examples of meaningful learning because students are not just reiterating information provided to them in the course, but have taken that knowledge and intentionally applied it to a novel situation. If traditional didactic teaching were as effective as the inquiry-based teaching, the distribution of generalized ideas
would be expected to be the same regardless of the type of unit or the topic. Given the more challenging nature of abstract concepts such as evolution, lesser frequencies of meaningful learning events might also be expected for the topic of evolution.

When new ideas evidencing a transfer to novel situations were categorized by instructional strategy, however, inquiry-based units produced generalized new ideas more often than expected compared to didactic units ($\chi^2 = 6.53$, critical value at $\alpha = 0.01$, df 1 is 3.841, Fig. 3). The new ideas produced by the didactic units were predominantly concrete (meaning that the new ideas were indistinguishable from rote learning of the course material) or vague. Thus, the inquiry-based units performed better than would be expected by

---

### Table 7 Summary of student open-ended responses from Spring and Fall semesters of 2003 by topic (n = 518 student responses)

<table>
<thead>
<tr>
<th>Topic (Frequency)</th>
<th>Examples of the most common misconceptions for each topic</th>
<th>Literature source(s) reporting similar alternate conceptions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Plants (31%)</strong></td>
<td>• Did not realize that many vegetables (e.g. tomatoes, cucumbers, etc.) are botanically defined as fruit  &lt;br&gt; • Thought water is pushed up from the roots to the rest of the plant  &lt;br&gt; • Plants can absorb water through their leaves  &lt;br&gt; • Ignorant of the role of axillary bud in defining a leaf (thought leaves are any green blade)  &lt;br&gt; • Plants are simple (lacking complex structures) or do not need to overcome the same challenges as animals.</td>
<td>(Wandersee 1985) purpose of leaf is to catch water; major source of plants “food” comes from the soil.  &lt;br&gt; (Haslam and Treagust 1987) capture of water; photosynthesis</td>
</tr>
<tr>
<td><strong>Animals (27%)</strong></td>
<td>• Compared to humans, thought that other animals would be very different and simpler (ignorance of homology)  &lt;br&gt; • Oversimplified circulatory system (ignorance of systemic vs. pulmonary circulatory systems)  &lt;br&gt; • Thought heart and lungs were separate (heart pumps blood, and gas diffusion just happens or gas circulation is an entirely separate system)  &lt;br&gt; • Thought most of digestion happens in the stomach (as opposed to small intestine).</td>
<td>(Bell 1981) ‘Animals’ used synonymously with ‘mammal’  &lt;br&gt; (Arnaudin and Mintzes 1985) ignorance of path of circulation and/or double circulation, and/or pulmonary system</td>
</tr>
<tr>
<td><strong>Evolution (21%)</strong></td>
<td>• Did not believe evolution occurred (usually due to religious convictions)  &lt;br&gt; • Thought individuals could evolve.</td>
<td>(Demastes et al. 1995; Dagher and BouJaoude 1997; Downie and Barron 2000; Brem et al. 2003; Deniz et al. 2008; Hokayem and BouJaoude 2008) disbelief in scientific view of evolution  &lt;br&gt; (Brumby 1984) evolution occurs within lifespan of single organism  &lt;br&gt; (Greene 1990; Jensen and Finley 1996) evolution/adaptation occurs as a response to need</td>
</tr>
<tr>
<td><strong>Biodiversity (17%)</strong></td>
<td>• Thought that humans (in some form) co-existed with dinosaurs  &lt;br&gt; • Thought humans/plants/animals were around for a significant portion of the earth’s existence.  &lt;br&gt; • Thought earth was much younger.  &lt;br&gt; • Thought fungi most closely related to plants or plants most closely related to animals</td>
<td>(Schoon and Boone 1998) cavemen coexisted with dinosaurs</td>
</tr>
</tbody>
</table>

Frequency is the percentage of the total responses that focused on that topic. Plant and Animal were didactically taught topics; Evolution, Biodiversity and Primary Literature were part of the inquiry-based units. A complete, detailed list of the prior concepts volunteered by students is available upon request.
producing not just an equivalent number of meaningful learning events, but a disproportionately greater number.

When broken out by topic, an analysis of the standardized residuals indicated that the source of the significantly different result for the inquiry units versus the didactic ones was the greater than expected number of generalized responses for evolution and primary literature and the fewer than expected responses on evolution in the concrete category (Table 8).

This result is especially noteworthy because while the topic of evolution and the primary literature were mentioned less frequently than were other topics, a greater proportion of those responses demonstrated that students had constructed transferable knowledge.

**What learning events did the students think were most meaningful?**

To gauge which elements of the curriculum were having the greatest impact, we asked students to identify the cause of the meaningful learning event they described in their earlier responses. We categorized the students’ responses by their specificity and self-awareness of active learning (Table 4) because awareness of one’s own learning, often termed “metacognition” (e.g. Keys et al. 1999), is necessary for truly meaningful learning or conceptual change to occur (Baird and White 1996; Treagust et al. 1996). Responses that exhibited little specificity were categorized as Vague (9%) and those that indicated rote-learning were coded as Unit-based (32%) (e.g. “The biodiversity unit taught me my new idea”). Responses that identified a particular portion of the unit or Specific activity were so designated (41%). Those that also described precisely how and when the activity caused conceptual change were coded as Metacognitive (19%). For example, “I became dissatisfied [with my old idea] when we performed the evolution lab with allele frequencies.” (Evolution, 535114) is an example of a specific response. In contrast, the following is an example of a metacognitive response.

**According to factual finch data from a primary literature source [Galapagos Finches], finch beaks evolved into longer beaks over a period of time with environmental pressure. Th[ese] data revealed to me that evolution can occur according to the criteria (environment can’t support all, variation of phenotypes, variation is heritable, differential reproduction and survival).** (Evolution, 98648)

In this instance the student has applied the criteria that he was given in class to the dataset and concluded...
that evolution has occurred. Metacognition, in other words, occurs when a student intentionally takes the knowledge given in class to make connections and conclusions for themselves. Thus, while all the metacognitive responses were attributable to a specific activity, they went further and defined an “aha” moment during the activity when the student made a mental connection between data or experiences and concepts.

When the number of metacognitive events was compared by instructional strategy, the inquiry-based units produced significantly more metacognitive events ($\chi^2 = 2737$, critical value at $\alpha = 0.001$ is 16.266). Explicitly, responses from inquiry-based topics comprised only 44% of the total responses, but 69% of the “metacognitive moments” (Fig. 4).

A Chi-square Test of Homogeneity determined that the source of this statistical significance between the didactic and inquiry-based units was largely due to lower than expected instances of metacognition for the didactic units and greater than expected frequency of metacognition in the Evolution unit (Table 9).

Thus, while the gains in knowledge of content were largest in the didactic units, at least a portion of those gains were due to greater time-on-task and lesser conceptual difficulty of content for beginning students. In contrast, while the topic of evolution showed lesser gains in knowledge of content due to its greater abstractness and informally derived preconceptions, the Evolution unit outperformed the other units in terms of the frequency of meaningful learning events.

**Discussion and recommendations**

This study provides support for the conclusion that traditional teaching methods appear satisfactory for gains in knowledge of content in concrete topics while inquiry-based pedagogies are effective in stimulating respectable gains in knowledge of content in more abstract topics. As this study does not comprise a clean comparison of pedagogical strategy, we cannot state that inquiry-based learning is superior to traditional teaching methods. We can report, however, that our Evolution unit produced gains in learning comparable to or greater than those reported in a variety of other studies. Specifically, when the results of a meta-analysis of student learning by topic (Guzzetti et al. 1993) were reanalyzed to focus only on student learning in the area of evolution, the average effect size found across seven studies was 0.29 (± standard error of 0.20) which is much smaller than our average effect size of 1.1 ± 0.3. Additionally, Schroeder et al. (2007) compared the effectiveness of different teaching strategies in 62 studies of locally designed tests of students’ learning and the average effect sizes reported there were similar to, or smaller than, our effect sizes for our inquiry-based unit. Specifically, Enhanced context strategies (making material personally relevant to students) had the largest average effect size at 1.48 and was the only effect size reported that was larger than ours. The 12 studies on Inquiry-based strategies (student-centered investigations involving data) had an average effect size of 0.65 and the Collaborative Learning and Formative Assessment strategy studies had average effect sizes of
0.95 and 0.51, respectively. No studies of the effectiveness of didactic teaching alone were found by Schroeder et al. (2007). Thus, we are confident that our inquiry-based units produce notable gains in knowledge compared to others reported in the literature. Additionally, our study revealed a decided impact of the inquiry-based pedagogy on students’ acceptance of the theory of evolution. Students responses which reported a meaningful learning event for the topic of evolution, displayed a disproportionate tendency then to generalize knowledge to novel situations and to have meta-cognitive awareness of the learning process. In contrast, the learning events reported for the other topics tended to be more concrete and students less frequently applied that knowledge to life outside of the course. Based on these experiences, we make the following recommendations to others considering investing in curricular reform.

**Use more than one method of assessment**

The greater gains in the didactic units must be considered within the contexts of the study and of the type of assessment. While the multiple-choice questions in the pre- and post-tests on knowledge of content were designed to test a range of understanding from factual recall to application, there are limitations in the depth of understanding that can be probed in such a format. We assert that examples of conceptual change offered by students in the open-ended responses provide an equally valid measure of the impact of the two approaches to curricula. “Meaningful learning” is defined by Novak (2002, p. 549) as “the intentional, substantive, non-arbitrary integration of concepts and propositions into existing knowledge and cognitive structure”. Whenever students have prior experience with a topic, meaningful learning must take the form of “conceptual change.” We are defining “conceptual change” here as the process of supplanting or reducing in status scientifically incorrect ideas in favor of more explanatory, fruitful, and scientifically accurate concepts (Posner et al. 1982; Hewson and Hewson 1983; Duit and Treagust 2003). Therefore, metacognitive responses regarding the causal mechanisms for conceptual change indicate that meaningful learning has occurred. Analysis of the examples of conceptual change provided by our students demonstrate that when meaningful learning did occur, it was significantly more often stimulated by the inquiry-based curriculum than by the didactic approach (Fig. 4, \(P<0.001\)). The inquiry-based curricula also produced significantly more generalized ideas (students provided evidence that they had transferred the concept to a novel situation) than did the didactic curricula (Fig. 3, \(P<0.01\)).

The use of a mixed method for assessing students’ learning provides a more holistic and informative comparison of the two curricular approaches. Mathison (1988) asserted that “the value of triangulation is not as a technological solution to a data collection and analysis problem, [but] as a technique which provides more and better evidence from which researchers can construct meaningful propositions [emphasis original] about the social world” (p. 15). We further assert that it is important that course assessments mirror the types of understanding (recall, application, open-ended communication, etc.) valued by the designers of the curriculum. Namely, a single type of assessment demonstrates learning for only a single type of understanding.

**Consider that all concepts are equally easy to learn**

As evidenced by our *post hoc* analysis, our assumptions significantly affected our interpretation of gains in students’ learning and others are cautioned to monitor time-on-task and differences in conceptual difficulty when comparing pedagogical strategies. We further wish to point out that our data on the sophistication of students’ prior knowledge of evolution should be interpreted with caution because a *single additional* well-developed prior idea on the topic of evolution would have resulted in statistical significance at the traditionally accepted \(\alpha=0.05\) level. Thus, the topic of evolution likely suffers from a greater propensity of well-developed misconceptions than do other topics. Previous studies (Dagher and BouJaoude 1997; Brem et al. 2003) described the tenacity of students’ prior ideas on evolution and attribute this tenacity to the long-standing nature of the prior conceptions (people learn about evolution from school, mass media, and informal communities such as organized religion for years before attending university) and the social and emotional implications of conceptual change in this area due to individuals’ religious beliefs. Therefore, while the sophistication of students’ prior ideas initially appears to be equivalent, suggesting that there should be no notable difference in the potential for conceptual change among topics, we would caution future investigators from ruling out the possibility that there may be more resistance to conceptual change in one topic than in another, and particularly so in the case of evolution.
Prioritize efforts in curricular reform based on the conceptual difficulty (abstractness) of topics and on the sophistication of students’ prior ideas about those topics

Our data indicate that didactic pedagogies were effective for concrete, descriptive topics with which students can have direct experience, such as plant and animal anatomy. Our inquiry-based units allowed students to more often generalize beyond the context of the course and created more transferable conceptual frameworks than did the didactic units. Most likely this is because inquiry-based teaching strategies allowed students greater opportunities to confront their prior ideas (Duit and Confrey 1996). Recent literature reported that abstract concepts are more difficult for students to learn than are concrete topics, particularly for students who lack strong, formal reasoning skills. For example, Lawson et al. (2000b) found that students’ achievement was approximately 30% less for more abstract (theoretical) topics and only students who also demonstrated formal reasoning ability were able to see large gains in knowledge for such abstract concepts.

Thus, abstract topics or topics for which students tend to have well-developed prior ideas may require inquiry-based approaches in order for meaningful learning to occur. We therefore recommend that efforts to reform curricula prioritize more abstract concepts in general and evolution in particular.

If time and resources allow for curricular reform to extend to all topics, our data suggest that when inquiry-based approaches are applied to concrete, descriptive topics, gains may surpass those of the didactic units. For example, our segment on primary literature produced the largest effect sizes in our study. We attribute this success to the multiple, direct experiences of the students with primary literature and the process of peer review, making this topic extremely concrete and personal. Furthermore, these notable effect sizes occurred despite a minimum of time-on-task, suggesting that under certain circumstances, inquiry-based curricula may be more efficient than is didactic teaching. Thus, when faced with limited time or resources for curricular reform, our data suggest that more ‘high-powered’ methods, such as inquiry-based curricular reform should focus on more abstract topics or those known to be resistant to conceptual change. Indeed, we would hypothesize that it is nearly impossible to stimulate meaningful learning or alter students’ conceptions of abstract topics such as evolution using only didactic methods. Inquiry-based methods which allow students to confront their prior conceptions are crucially important for meaningful learning to occur in these areas.

Acknowledgments

We thank Dr Sarah (Sally) Woodin, Chair of the Department of Biological Sciences for her much needed support and encouragement, especially in the early stages of the project. Drs Laurel Hester and Charles Duggins provided valuable comments on early stages of the article. Dr Jon Singer suggested the Galapagos Finches software and provided laptop computers for the pilot study (data not reported here) as well as provided editorial comments at early stages of the article. The authors would also like to express their thanks and gratitude to all the BIOL 102 teaching assistants who implemented the new curriculum and facilitated the subsequent collection of data and to Drs Nelson and Duggins, the BIOL 102 lecturing faculty, for their cooperation and assistance. We are also grateful to David Degenhardt and Cecilia Ma for assistance with the qualitative coding. Funding for this work was provided by National Science Foundation (0410992 to B.E.T.).

References


