BOARD ON SCIENCE EDUCATION
CENTER FOR EDUCATION
THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

Grant, B. W. 2008. Practitioner Research as a Way of Knowing: A Case Study of Teacher Learning in Improving Undergraduates' Concept Acquisition of Evolution by Natural Selection. Commissioned Paper from the National Research Council Board On Science Education Workshop "Linking Evidence and Promising Practices in STEM Undergraduate Education," 30 June 2008, at the National Academies, Washington, DC. (http://www7.nationalacademies.org/bose/PP_Commissioned_Papers.html).

Practitioner Research as a Way of Knowing: A Case Study of Teacher Learning in Improving Undergraduates' Concept Acquisition of Evolution by Natural Selection.

Bruce W. Grant, Widener University, Chester, PA, 19013 (bwgrant@widener.edu)

"Teaching requires the recognition that education is ideological."

— Paulo Freire (1998, Pedagogy of Freedom)

Summary:

In the pages that follow, I offer a multi-year case study showing how practitioner research improved my undergraduate freshmen students' concept acquisition of evolution by natural selection (spanning 8 cohorts from 2000-2007). I will first describe (1) the kinds of challenges that I and others face in teaching complex concepts such as evolution to freshmen. Next, I will explain (2) in general, how practitioner research extends many attributes of traditional scientific epistemology, yet jettisons others, and thereby constitutes an unique and complementary "way of knowing" to traditional scientific inquiry. Next, (3) I narrate my practitioner research findings showing evidence-based course revision and evidence of improvements in student performance. Lastly, (4) I conclude with several inductively derived postulates from my "personal theory" about causality in my teaching and my students' learning (Jarvis 1999) to account for my students' conceptual change with suggestions for both qualitative and quantitative deductive inquiries to follow. My overarching message, especially for readers who teach and those who teach future teachers, is one of cautionary advocacy to engage in a learning process to assimilate and wield methods of practitioner research as a "way of knowing" to improve your science teaching and your students' learning.

(1) The Challenges to Teaching Evolution.

Despite that "biological evolution is the central organizing principle of modern biology" (NAS 2008), it is widely known that large numbers of students that we face in our freshman biology classes are woefully unprepared to undertake the study of evolutionary biology (Alters and Nelson 2002, Bowman 2008, Hokayem and Boujaoude 2008). In part, this is the result of an organized resurgence of pre-enlightenment thinking throughout and especially late in the 20th century (Wills 2004) designed to cast doubts upon fundamental constructs of the theory of evolution by natural selection (Brem *et al*. 2003, Miller *et al*. 2006). Whether by design or by neglect (Griffith and Brem 2004), many pre-college school systems nationwide have graduated successive cohorts who harbor, express, and obdurately defend deeply held misconceptions not only about the basic processes of evolution (Munson 1994, Dagher and BouJaoude 1997, Alters and Nelson 2002, Anderson *et al*. 2002, Bowman 2008), but also about the kind of thinking that led to discovering and constructing the evidence in support of evolution to begin with, *i. e.* their view of science emerges from a flawed scientific epistemology (King and Kitchener 1994, 2004, Lederman et al. 2002, Bransford et al. 2000, Sinatra et al. 2003, Smith and Wenk 2006).

One widely used response has been to inventory students' misconceptions in evolutionary biology and design curricular activities that confront these misconceptions directly (*e. g.*, Bishop and Anderson 1990, Rutledge and Warden 1999, Anderson *et al*. 2002, Passmore and Stewart 2002, Hokayem and Boujaoude 2008). However, as Wandersee *et al*. (1994) caution, it has been amply demonstrated that simply telling students that their "mental model" of a phenomenon is incorrect is ineffective at displacing their misconceptions and affecting conceptual change. Following an extensive review and synthesis of evidence about the roles of students' alternative conceptions and misconceptions in their understanding and acceptance of evolution, Alters and Nelson (2002) advocate that a "key improvement" in teaching evolution would be to use more active learning and constructivist approaches in our classroom pedagogy.

The challenge then becomes "how?" Thus, how might one start with somewhat pre-formatted tools (e. g. Anderson *et al*. 2002) or simple writing assignments (*e. g.*, Hein 1999) to identify students' misconceptions not only to the instructor, but also to make these known to the students <u>and engage them</u> in the process of displacing their own misconceptions so that the "expert" constructs may be heard, accommodated, and assimilated (Wandersee *et al*. 1994)? These are classroom research questions, and faculty who will be successful at this must become truly scientific in their approach to teaching. For example, if students perform poorly on a quiz or exam, this should be seen as a "discrepant event" (Liem 1981) eliciting surprise (Alder 2008) and curiosity (Schmitt and Lahroodi 2008), and the challenge is to avoid "deficit explanations" for poor performance (*e.g.*, the students are unintelligent or don't work hard enough, Elmersky and Tobin 2005, or a host of other deeper deficit models stemming from racism, classism, sexism, and other forms of discriminatory oppression that have hugely affected the hegemony of educational cultural design in the US for over two centuries, Valencia 1997), and instead rise to engage in "scientific teaching" (Handelsman *et al*. 2004) to discover reasons for the discrepancy – how did my students arrive at their observed misunderstandings? and how can I design mitigation, or better, pre-emption/ corrective pedagogies to engineer and effect the acquisition of both expert knowledge and expert ways of knowing (Gardner 2004). Knowing the reasons for students' misunderstandings derived from, or despite, our teaching allows us to revise and improve our curriculum, instruction, and assessment and cause improved student learning in science (Bransford *et al*. 2000, D'Avanzo 2003, D'Avanzo *et al*. 2006, Morris *et al*. 2007).

Likely, the reasons teachers generally do not approach these discrepant events in student performance scientifically, and engage in the kinds of thinking necessary to understand and seek solutions scientifically, is because faculty generally lack the combination of scientific thinking skills <u>and</u> evidence-based instructional models in science education needed to work from serendipitous discrepancy to revolutionary discovery/ invention about causality in student learning (Bransford *et al*. 2000, NRC 1996). Faculty who teach undergraduate science courses, of course, have been trained over many years in the methods of positivist scientific research – but the key limiting task in developing classroom research may be to develop their skills in translating and transforming that positivist disciplinary research paradigm into an effective classroom research paradigm (Handelsman *et al*. 2004). Since the principal goal of these research activities is to improve the action of teaching practice, this activity has been referred to as teacher "practitioner research" or "action research" (Stenhouse 1975, Jarvis 1999, Allwright 2005). The same thinking and model of faculty development needs to be applied to undergraduate faculty who teach evolution.

Yet, I enjoin that this is not as easy as I just made it seem — the positivist rational western scientific research paradigm appears to contain substantive epistemological roadblocks of both method and philosophy to this necessary transformation (*e.g.*, Thomas 1998, Carr 2006). In the next section, I will describe the causes and consequences of these roadblocks, and offer navigational suggestions around them.

(2) Practitioner Research as a "Way of Knowing."

> "Action research is a participatory, democratic process concerned with developing practical knowing in the pursuit of worthwhile human purposes, grounded in a participatory worldview which we believe is emerging at this historical moment. It seeks to bring together action and reflection, theory and practice, in participation with others, in the pursuit of practical solutions to issues of pressing concern to people, and more generally the flourishing of individual persons and their communities."
>
> — Reason and Bradbury (2001)

The concept of practitioner research refers to "active professional learning" (Grady 1998) in which practitioners do research to improve the action of their practice (*e. g.*, Schön 1983, Stenhouse 1985, Bassey 1992, Jarvis 1999). Practitioner research is similar to "action research" (*e. g.*, Allwright 1984, 2005, Hollingsworth 1997, Reason and Bradbury 2001, Chandler and Torbert 2003, Carr 2006, Stringer 2008), different definitions for both abound, and the distinctions and nuances of these two classifications are way beyond my goals here (see Chandler and Torbert 2003, Carr 2006).[1] The call for and manifestations of practitioner research have emerged recently in diverse educational disciplines from the teaching of modern languages, to composition, to other areas of the humanities, to the social sciences and social work, and throughout the sciences (K-16 and beyond). As a key example, the NAS National Science Education Standards (NRC 1996) recommended that teacher professional development needs to change teachers from "consumers of knowledge about teaching" to "producers of knowledge about teaching." These ideas are not new, and in fact, in an extensive review of the origins of practitioner research, Carr (2006) citing the work of Elliot (1987) and others, traced the concept to the Aristotelian applied philosophical construct of "praxis" — "it [practitioner research] would be regarded as nothing other than a post-modern manifestation of the pre-modern Aristotelian tradition of practical philosophy." This point is of great significance because according to Carr (2006) the philosophical origin of Aristotle's "praxis" stems from an entirely different construct than the types of human intellectual endeavor from which post-enlightenment western science arose. I will return to this point below after a discussion of how the methods of practitioner research align or do not align with the steps of the scientific method in a positivist research tradition and the present day view of science as a "way of knowing."

To illustrate the challenge to science faculty who generally are trained in a positivist disciplinary research paradigm to retrain themselves to conduct practitioner research to improve their teaching and their students' learning, one need only disarticulate the steps of the scientific method and align these with the kinds of things that effective practitioner researches will need to know how to do. Many versions of this list are possible, and the one I present below should suffice for my purposes.

Step 1 is the observation of something intriguing or paradoxical that elicits curiosity, wonder, surprise, and/or excitement and leads to questions and hypotheses

[1] For my purposes, I ask readers to overlook my blurring of these terms and my use of the term "practitioner research" for the generic set of methods I present in this section.

derived inductively about causality. Science faculty are generally expert in making the kinds of keen observations about scientific phenomena in their discipline that drill down to the specifics of what really is apparent and quickly lead to specific and hopefully answerable questions leading to hypotheses about causality. These are learned skills and require expert filters and interpreters of basic sensory information, *i. e.* highly developed cognitive skills in observation and information processing. In addition, and more importantly, expert observers know the existing research in their areas and their minds are trained to access this base of prior knowledge in accommodating new phenomena. Expert observers are made most curious when new things fail to fit neatly into previously structured schema. What training do faculty need to accomplish expert cognitive processing of observations of teaching and student learning? What observations would an expert focus upon that even constitute evidence of the effects of our teaching on students' learning? What are the core hypotheses and theories about how people learn (*e. g.*, Bransford *et al*. 2000) and how do these schema define our internal rules of cognitive linguistics to even put words to what we see and thereby begin a path to knowing? In the absence of expert understanding of the learning processes in class and needs of our students, it is a daunting challenge to articulate what we see in a way that leads to testable questions.

For example, those of us who teach freshmen and/or non-science majors are keenly aware of the chaos and noise of novices' observational skills that fill the air during lab inductive inquiry activities such as observing termites racing along curvy ink lines and circles made by some pens but not others (the pens that use ink with a chemical additive that by coincidence mimics the termites' trail pheromone elicits this response, but termites ignore other ink, pencils, etc.). How different are we from these freshmen when we witness student responses on tests and quizzes indicating that only 30% of the class got anywhere near the correct answer on a question but "I am sure I covered it at least 5 times!" And at a larger scale, how many of us have noticed huge dropout rates of potentially successful students as they "fail to progress" in our programs and paused to wonder why?

As a last example, using real data from one of my own classes (freshman fall semester introductory ecology and evolution course, Bio161), consider Figure 2.1 (below) which shows the individual students' quiz/exam scores on a single question worded identically (Q — please offer a brief and concise definition of "Evolution") at four different times during the same semester. Note that individual scores on this basic definition are bouncing all over the place and, at any given assessment event following the pre-test, many students scored worse than they did on the previous test. Note also, that the average class performance (red line) increased, however, this trajectory bears little resemblance to the vast majority of any of the students. What chance does a disciplinary faculty (but a novice teacher at practitioner research) have at posing testable hypotheses about causality amidst this kind of mess? What cognitive skills in educational research inquiry into what's going on in these students' minds and lives are necessary and sufficient to buffer one against the temptation, if not reflex, to pursue deficit explanations (Valencia 1997, Elmersky and Tobin 2005)? What does one need to know to react with curiosity and ask questions versus experiencing revulsion or malaise and in self-defense turn away? Moreover, who would think or take the time to archive and plot data such as in Figure 2.1 and look for this kind of pattern? What kinds of drivers in academia affect science faculty priorities to pursue these inquiries or not?

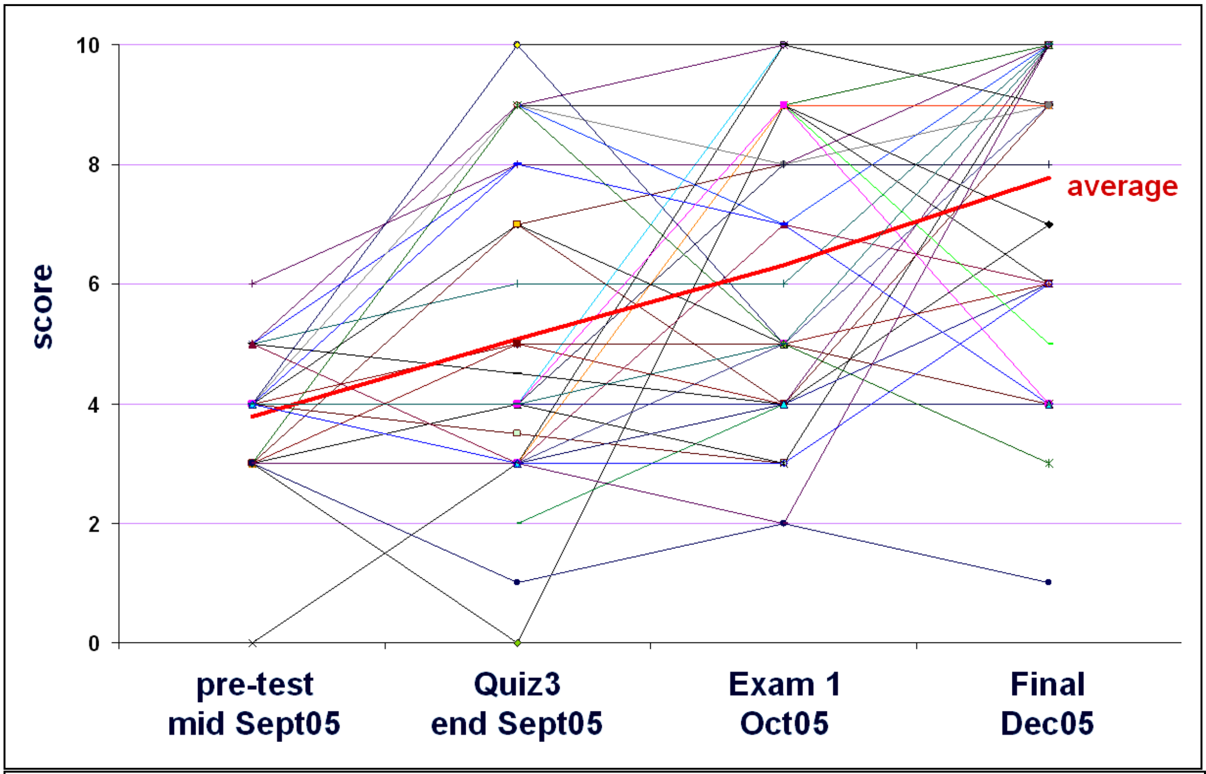


Figure 2.1 — Individual variation in student scores on the “≡Evolution” question in Bio161 during the Fall 2005 semester on a 10 point scale (red line is the average). The points labeled “pre-test” were prior to any instruction in my course. The scoring rubric appears below.

Step 2 is the experimental design and execution stage to test hypotheses derived inductively about causality in Step 1. For practitioner research, this step marks a radical departure from the positivist scientific paradigm principally in that the frame of the “experiment” is no longer defined by a dipole of researcher and researched. The unit of the experiment is not simply its object of study — the students. Barton (2001) enjoined one should do “research with rather than research on” one’s students. Through practitioner research one develops (whether by invention or discovery) a “personal theory” (Jarvis 1999) to improve student learning through the action of one’s own practice. Key facets of this personal theory are as much a result of investigations into understanding and re-engineering one’s own capacities to teach as they are about how the ecology of the classroom that one elicits through one’s curriculum, instruction, and assessment causes personal and social engagement to affect (or obstruct) student learning. The interaction of teacher and taught is the domain of inquiry.

Importantly, personal reflection forms a core method of one’s research design. Reflection, when properly undertaken, is to being awake what REM dreaming is to being asleep. In a reflection, one stops time and gazes past the surface of life events (especially those for which we may have developed strong emotions, such as student failure) into that virtual mental space to make sense of and seek meaning to targeted events in the time and space through which we have just passed in our teaching. Not only is it a mental winnowing of the events of our experiences to parse signal from noise, but we gather those pieces of our experiences and kernels of interpretation that we can use to explain these events and make things “fit together.”

As a result, the methodology of practitioner research necessitates a critical blurring between Steps 1 and 2 above, because it is through reflection that new observations and new hypotheses for new teaching methods and interventions emerge as fundamentally integral outcomes of one's research design. Thus, there is a non-linear fluid mosaic to the practitioner research process as one develops a sense of cognition while engaged in "organizational improvisation" *as a research paradigm* (Weick 1998). Lévi-Strauss used the term "bricolage" to capture this "chaotic advancement… disjointed incrementalism… [and] a kind of anarchy" of knowledge construction in this research methodology (Thomas 1998) — and with this, pre-scripted linear positivism quickly looses instrumental relevance.

And if that were not enough, the types of analytic methods to distinguish among hypotheses that require methods from qualitative and ethnographic research and hybrid qualitative*quantitative "mixed methods designs" (e.g. Merriam 1998, Reason and Bradbury 2001, Stringer 2008) may appear to be as daunting and impenetrable as the gates of Mordor to novice classroom researchers. Nonetheless, in my own experience about which I will share in the next section, there really is not that much to learn in order to get started. I enjoin that among the challenges I raise to developing capacity and agency in practitioner research, these types of methods acquisition issues are the least problematic, yet not without concerns.

Lastly with this step, I raise the question of "how do you know when you know?" which is the challenge of metacognitive function (Flavell 1976). Similar to issues raised in Step 1, this skill lies at the core of what constitutes expert capacity and agency. Expert conceptual understanding goes hand in hand with capacity for metacognitive function (Schraw 1998, Sternberg 1998, Isaacson and Fujita 2006). If results are reducible to interpreting quantitative test statistics, most disciplinary science faculty would be in their element — but, for reasons given above, this may be rare in one's practitioner research project. One critical application of metacognitive capacity in practitioner research is in the languaging of reflections from which qualitative and quantitative analyses and synthesis emerge: it is a creative process (Thomas 1998), the skills are learned, and unskilled reflection can be deeply misleading (more on this below).

Step 3 closes the loop on scholarship and constitutes the publication of one's practitioner research findings. As with Steps 1 and 2, disciplinary science faculty generally have little to no experience publishing in journals appropriate for practitioner research let alone in any science education journals. The best options to seek are in venues owned and published by one's own scientific society, of which the number increases every year as more disciplinary societies recognize value in the scholarship of science education in their respective disciplines. I am very much involved with one of these efforts and over the past 8 years as Co-PI with Charlene D'Avanzo on three NSF grants, we and others created a peer-reviewed journal for ecological educational scholarship, specifically including practitioner research, published by the Ecological Society of America entitled *Teaching Issues and Experiments in Ecology, TIEE,* http://tiee.ecoed.net (D'Avanzo *et al*. 2006, Morris et al. 2007). At present, *TIEE* contains 6 volumes of about 50 peer reviewed submissions, and we hope to see *TIEE* continue as an important member of ESA's family of journals.

The previous discussion hopefully clarifies critical epistemological roadblocks of method facing traditional disciplinary scientists who embark on a practitioner research project. Next, I explore a major epistemological roadblock of fundamental philosophy, i.e. "way of knowing", to this journey that stems quite simply from the question "what is the goal of the activity?" An excellent summary of the idea of science as a "way of knowing" is by Francisco Ayala (see also Moore *et al*. 1984):

> "Science is a wondrously successful way of knowing. Science seeks explanations of the natural world by formulating hypotheses that are subject to the possibility of empirical falsification or corroboration. A scientific hypothesis is tested by ascertaining whether or not predictions about the world of experience derived as logical consequences from the hypothesis agree with what is actually observed. Science as a mode of inquiry into the nature of the universe has been successful and of great consequence." (Ayala 2008)

Practitioner research shares with traditional scientific research those core post-enlightenment tenets that distinguish science from theology, from much of metaphysics, and from all propaganda and "bullshit" (Postman 1969, 1979, Frankfurt 2005, see also Mason 2005). Moreover, although I have never seen the data, I posit that the vast majority of disciplinary scientists and practitioner researchers would concur with Feynman (1999) that there is a certain "pleasure in finding things out."

However, a divide becomes apparent when examining the issues of "empirical falsification or corroboration" both of which are problematic since practitioner research findings, *i. e.* meaningful personal devices to improve the action of one's practice, often emerge through qualitative processes of personal reflection and amidst smaller sample sizes inherent in case study and much of ethnographic research. A great deal has been written on the limited resolution and reliability of self assessment (*e. g.*, Claxton 1997, Kruger and Dunning 1999, Dunning *et al.* 2003), and wrote Thomas (1998), "Unless our personal musings are subjected to rigorous criticism — and survive — they may constitute no more than prejudice, generalisation and dogma." The extent of this divide and the metacognitive rules for its resolution are socially negotiated differently among different disciplines. Much has been written on this problematic concern of empirical sufficiency, but this concern is minor by comparison with the issue I raise next.

Practitioner research has a very different primary goal than does traditional scientific research. As summarized by Stringer (2008), practitioner researchers "engage in careful, diligent inquiry, not for purposes of discovering new facts or revising accepted laws or theories, but to acquire information having practical application to the solution of specific problems related to their work."

Jarvis (1999) states that practitioner research "reconceptualizes theory, arguing that all practitioners generate their own personal theories, and that apparently objective knowledge of traditional theory is no more than information to be learned and experimented with in practice… [and as such] the traditional relationship between theory and practice no longer holds." And later, Jarvis relays "it is no longer possible to treat theory as a coherent entity that can be generalized to all practice situations. Indeed, it is questionable whether the high status of theory should be retained in its present form in light of the current emphasis on practice, reflective practice, practical knowledge, and practitioner-researchers' research into practice. Practical knowledge, however, might be regarded as a new aspect of theory, a personal theory of practice." (p. 144)

Thomas (1998) goes even further by asserting that for practitioner researchers the goal of seeking theoretical generalities is not only irrelevant but misguided: "the problem is… with the belief that one's own observations and reflections can be corralled, cleansed and transformed to provide an improved explanatory structure and practical guide for one's future professional life." According to Thomas,

> "The key question is whether we can deliberately theorise in such a way that our practice is affected; whether there are two distinguishable and separable processes, or whether 'tacit knowing' (Polanyi 1958) is a kind of knowing out of which theory cannot be drawn… the skills of the teacher or physician can be shared only by a process similar to apprenticeship - in the same way that knowledge about language use is shared with the developing infant. They are reducible neither to technique nor to theory… The analogous distinction… is between 'knowing how and knowing that' [Ryle 1949]… Knowing how to do something, in other words, is not predicated on knowing principles for doing it or the possession of articulated knowledge. As Ryle put it, 'Intelligent practice is not a step-child of theory' (p. 26). … When someone does or thinks something in the practical world, asking them to reflect or theorise on it merely produces what Ryle calls a ghostly double, 'a soliloquised or muttered rehearsal' "…

Thomas continues with,

> "By calling the collection of vernacular constructs, reflections and ideas 'theory' we claim some epistemological legitimacy and explanatory currency for them. But personal views culled from everyday experience are not theory. To elevate ideas, impressions and transient hypotheses to 'theory' dilutes the word theory to such an extent that it ceases to be of any descriptive value. Merely to call these cognitive phenomena 'theory' blesses them with neither validity nor utility. Musings are not theory. Unless our personal musings are subjected to rigorous criticism - and survive - they may constitute no more than prejudice, generalisation and dogma."

In sum, practitioner research jettisons the core western scientific construct of generalizability, or the discovery or invention of generality as the primary goal of the activity. As shown in Figure 2.2 (below), the Research Cycle of Practice (which is based upon the Aristotelian concept of *Praxis*) begins with observations of one's classroom practice from which one resolves patterns of relationships between one's teaching and evidence of students' learning (spanning qualitative and quantitative data types). Then, one reflects upon possible relationships and inductively arrives at a "Personal Theory of Practice" according to which students are hypothesized to be learning (or not) due to one's teaching (curriculum and instruction). Following this, one then can make revisions to one's teaching, return to the classroom, and repeat the cycle again — hopefully this time with improved learning outcomes. Importantly, each step I have outlined is deeply informed by pre-existing knowledge and theories of learning generated elsewhere by other researchers and publishing practitioners. One's research cycle of practice must occur within a learning community of practitioners; however, the

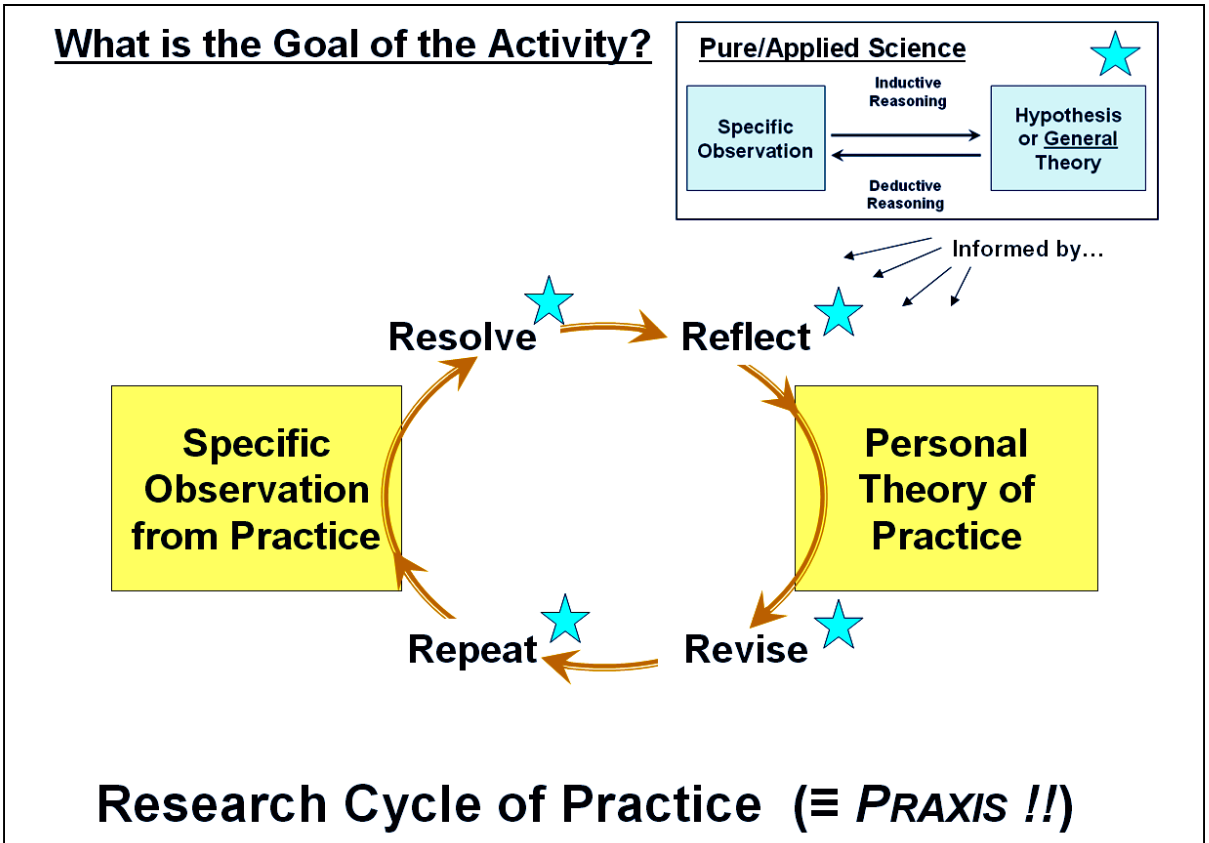


Figure 2.2 shows the Research Cycle of Practice (based upon the Aristotelian concept of *Praxis*). This cycle begins with observations of one's classroom practice from which one resolves the effects of one's teaching on students' learning, and one reflects upon and inductively hypothesizes causal relationships toward a "Personal Theory of Practice", according to which students are hypothesized to be learning (or not) due to one's teaching (curriculum and instruction). Lastly, one then can make revisions to one's teaching, return to the classroom, and repeat the cycle again — hopefully this time with improved learning outcomes.

primary goal of the activity is NOT the generation of new generalizable knowledge, which IS the primary goal of traditional scientific research — the primary goal of practitioner research is to teach better so as to improve one's students' learning. I next turn to the underlying issue why is this goal important — why do we want to improve practice? I offer next that the answer to this question lies in the history of the construct of practitioner research and is at the heart of the matter.

Wilfred Carr (2006) summarizes the origins of practitioner research (he uses the term "action research") as descending from the Aristotelian tradition of "practical philosophy" and the specific construct of "praxis", which is

> "a form of action directed towards the achievement of some end… [and] the 'end' of praxis is not to make or produce some object or artefact, but progressively to realise the idea of the 'good' constitutive of a morally worthwhile form of human life. But praxis is not ethically neutral action by means of which the good life can be achieved. …praxis is a form of 'doing' action precisely because its 'end'—to promote the good life—only exists, and can only be realised, in and through praxis itself. … Praxis is thus nothing other than a practical manifestation of how the idea of the good is being understood, just as

knowledge of the good is nothing other than an abstract way of specifying the mode of human conduct through which this idea is given practical expression. In praxis, acquiring knowledge of what the good is and knowing how to apply it in particular situations are thus not two separate processes but two mutually supportive constitutive elements within a single dialectical process of practical reasoning."

According to Carr, Aristotle's construct of praxis founded the tradition of 'practical philosophy' from which practitioner research originated. However, practical philosophy was deliberately marginalized in modern times because "the indeterminate and imprecise nature of praxis unavoidably entails that practical philosophy is an 'inexact' science which yields a form of knowledge that cannot be applied universally and unconditionally." Thus,

> "practical philosophy… is nothing other than a pre-modern version of twentieth century action research. Like action research, it takes ethically informed human practice as its unique object domain. Like action research, it can be defined as 'a form of reflective enquiry undertaken by practitioners in order to improve their own practices, their understanding of these practices and the situation in which these practices are carried out' (Kemmis 1988, p. 42). And, like action research, it accepts that the knowledge that informs and guides practice is 'contextualised knowledge that cannot be separated from the practical context in which it is embedded' (Somekh 2006, p. 28)."

This analysis clearly demonstrates that the core philosophy of practitioner research diverged from that of western science over 2000 years ago and now embodies a "contemporary rehabilitation of practical philosophy" (Carr 2006) — the goal is not generalizable theory, but rather the construction of contextualized knowledge that improves the action of one's practice *because* the outcomes produce good ends. The message for educators is that practitioner research is value driven, goal-oriented, and transformative advocacy for both teacher and student for the good of both. It is a form of social engineering to improve classroom praxis because <u>that</u> is the right thing to do. Education is liberating, the capacity to teach and learn is emancipatory, our students deserve it, our culture demands it, and these outcomes clearly embody not only a core western ideological position (Freire 1998) but moreover, it is another "way of knowing" – and one that I enjoin is an essential complement to the way of knowing of western traditional science (see Ayala's quote above).

Continues Carr (2006): "although it was readily conceded that practical philosophy does not provide a body of knowledge that practitioners can simply apply, this did not undermine its claim to be the 'science' that enables practitioners progressively to improve their practical knowledge and develop their understanding of how the good [that is] internal to their practice may, in their own particular situation, be more appropriately pursued."

In conclusion, my overarching message for readers who teach (and especially those who teach future teachers) is one of cautionary advocacy to engage in a learning process to assimilate and wield methods of practitioner research as a "way of knowing" to improve your science teaching and your students' learning.

(3) A Case Study of Teacher Learning in Improving Undergraduates' Concept Acquisition of Evolution by Natural Selection.

Since 2000, I have been engaged in a long-term practitioner research project in my fall freshman introductory course in evolutionary biology (Bio161) at Widener University. I present evidence that my methods of practitioner research over the past 8 years, and in particular a major course revision including a novel method of confronting students' misconceptions that I developed in 2006, have contributed to significant reductions in my students' misconceptions and significant increases in their demonstrated understanding of basic concepts of evolutionary biology.

In the fall of 2000, following a complete revision of the freshman and sophomore majors' curriculum in Biology at Widener University, where I have taught since 1993, a colleague and I created the fall freshman course, "Biology 161: Evolutionary Ecology." Bio161 is divided into thirds: (1) Origins (basic review of the origin of the Universe, our solar system, pre-biotic chemistry, and life on Earth), (2) Genetics (micro- to macroevolution, and the evolutionary diversification and classification of life), and (3) Evolutionary Ecology (individuals, populations, communities, and ecosystems) and Conservation Biology. I have taught the first and last third of this course each fall from 2000-2005 with 45-60 students each year. Evolution and coevolution are central concepts to this course and these terms appear prominently in eleven and five out of fifteen weeks or the course, respectively (see Supplemental Figure S1 for the basic lecture outline for this course as well as the numerous places where the term evolution are coevolution are featured). During 2000 to 2005, in my lecture classes, I frequently made use of active-learning pedagogical techniques including Turn-To-Your Neighbor, Minute Papers, and Guided Discussions (see *TIEE Teaching Resources* {http://tiee.ecoed.net}).

Measuring Students' Understanding of Evolution by Natural Selection

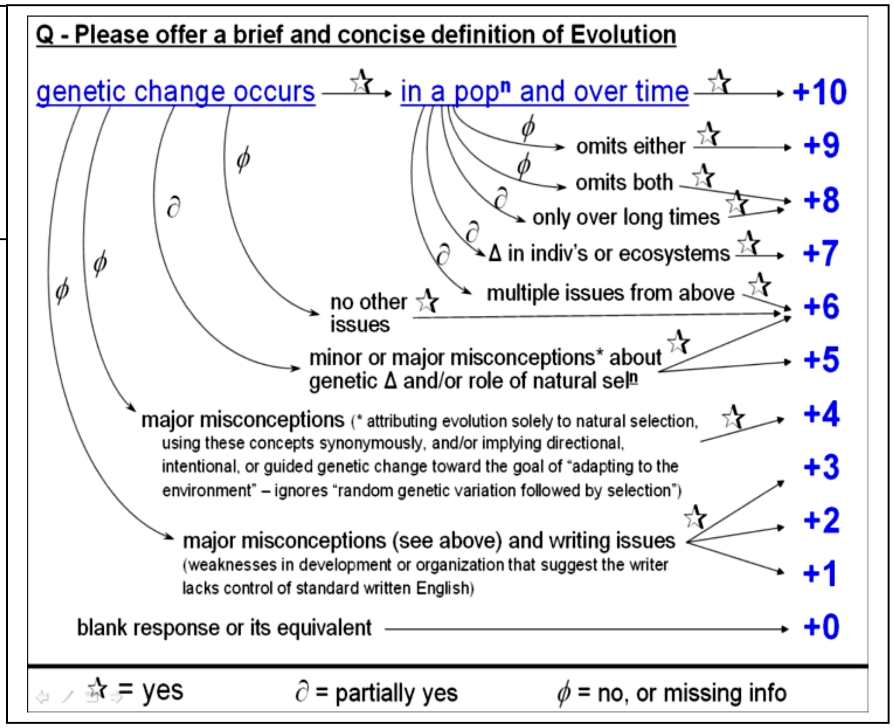
This paper reports multiyear results (2000-2007) from two sets of quiz/exam questions that serve to elicit students' understanding of (1) the definition of evolution and (2) the process of evolution by natural selection.

(1) Definition of Evolution. I put the question "Please offer a brief and concise definition of Evolution" on the Bio161 Final Exam each December from 2000 to the present. I re-read and re-scored all Final Exam responses (n=379) to this question using the rubric in Figure 1, which I also used to compile frequencies of students' misconceptions.

According to the rubric (Figure 3.1), students must have noted the role of "genetic change" in evolution to excel in this question (the emphasis is on microevolution). Next, failing but score-able responses exhibited major misconceptions such as equating evolution with natural selection, or implying teleological directed evolution based on vaguely adaptive need. Also, a student's performance at simply writing clear sentences affected their score (max of minus one point).

Since the beginning of 2005, in addition to final exams, I administered this question as a pre-test and on several quizzes, and I also saved copies of all pre-tests,

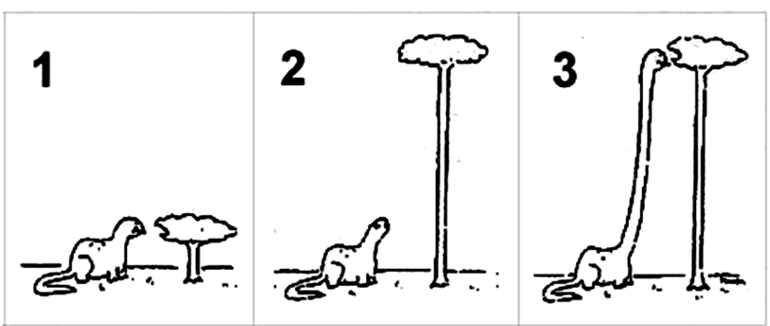
Figure 3.1. Scoring Rubric (max +10 pts) for the question “Please offer a brief and concise definition of evolution.”



quizzes, and mid-term tests. Thus for 2005 through 2007, I have an additional data set of responses (n = 647) to indicate changes in students' performance on this question pre-, during, and post-instruction. I re-read and re-scored all of these responses.

(2) Evolution by Natural Selection. One of the most useful instruments I have to measure students' conceptions about evolution by natural selection is the “Dino Neck” question (Figure 3.2). I administered this survey to all of my freshman Bio161 students during the first and last weeks of class from 2000 to 2007. I re-read and re-scored all responses using the rubric in Table 3.1 (below). I also used this rubric to compile frequencies of students' misconceptions.

Below is a cartoon entitled “Evolution Made Simple” from a popular magazine:



Task: Create a biological scenario that explains the phenotypic changes in the tree and the animal. Use your understanding of evolution by natural selection.

(Ebert-May 2000)

Figure 3.2. “Evolution Made Simple” assessment question from Diane Ebert-May (2000).

To excel, students must have noted for both the tree and then for the animal that (a) coexisting individuals in the tree/animal populations differ from each other (in tree height or neck length) due to random genetic variation, (b) stated that due to these differences trees/animals differ in ecological performance (survival and reproduction, or fitness) since taller trees are eaten less/ longer necked animals get more food, which is natural selection, and (c) since these taller/ longer necked individuals leave more offspring that would inherit these traits, the populations of trees/ animals subsequently increase in their proportions of taller/ longer necked individuals – which is evolution by natural selection. Lower scores stem from major misconceptions such as stating that variation among individuals and/or over time is directed by teleological need, or that variation is mere growth. Also, a student's writing performance slightly affected the score (max -1 point).

Table 3.1. My scoring rubric for the "Dino Neck" assessment question from Diane Ebert-May (2000).

pts	summary	key features
=10	expert answer	explains role of random mutation in creating genetic variation and phenotypic change in trees w/ some trees taller & some shorter; explains role of natural selection by dinosaurs on tree height causing the shorter trees to become rare and the taller trees to become common, which results in evolution of taller trees since previously height was attributed to genetic variation; explains role of random mutation in creating genetic variation and phenotypic change in dinosaur neck length w/ some necks longer & some necks shorter; explains role of natural selection imposed by the tall trees on dinosaur neck length causing the shorter necked dinosaurs to become rare and the longer necks to become common, which results in evolution of longer necks since previously neck length was attributed to genetic variation; in addition, may note this is an example of coevolution (if so, give ☆)
9-10)	near-expert answer	explains role of natural selection in causing the evolution of tall trees and then longer necks in two distinct stages – uses "natural selection" and/or "fitness" correctly in both cases; may be slightly vague about the genetic basis, but must use genetic variation correctly;
8-9)	good answer	explains role of natural selection in causing the evolution of tall trees and then longer necks in two distinct stages – may fail to name the process "natural selection" in one or both cases; vague about but still must mention the genetic basis to phenotypic variation in height or necks among **coexisting** trees and dino's;
7-8)	developing answer	explains role of natural selection in causing the evolution of tall trees and then longer necks in two distinct stages, but may omit one of these if other is well-explained; shows NO evidence of arguments about either teleological directed mutation/ evolution or about Lamarckian thinking; genetic basis to phenotypic variation in tree and neck length may be missing, and may discuss phenotypic var^n among **coexisting** ☆'s;
6-7)	minimally passing	vague about the role of natural selection in causing the evolution of either tall trees or longer necks, (may omit one of these if other is better explained) ambiguous evidence of arguments about either teleological directed mutation/ evolution and/or about Lamarckian thinking genetic basis to phenotypic variation in tree and neck length IS missing
5-6)	failed but nearly passing	poor understanding of the role of natural selection in causing the evolution of either tall trees or longer necks, arguments are made about either teleological directed mutation/ evolution and/or about Lamarckian thinking to explain the pattern causes of phenotypic variation in tree and neck length missing based on arguments of NEED sentences are clear and well-sequenced, however the concepts are wrong
4-5)	content and concept failure	NO understanding of the role of natural selection in causing the evolution of either tall trees or longer necks, arguments are made about either teleological directed mutation/ evolution and/or about Lamarckian thinking (or utter nonsense) to explain the pattern causes of phenotypic variation in tree and neck length missing based on arguments of NEED sentences are poorly organized and arguments ill-conceived
3-4)	content and concept failure	arguments are alluded to about either directed mutation/ evolution and/or about Lamarckian thinking, however, the main points are off track, unclear, and incomplete otherwise maybe a score of 4, but with three or fewer lines
2-3)	major failure	a few key words are present, and they are embedded in incomplete and erroneous phrases; or, the writing might be adequate, but the pattern is entirely attributed to mere growth, ontogeny, or development;
1-2)	total failure	a few key words are present, but are ideas are skimpy and incoherent
0-1)	total failure	irrelevant information or blank

Importantly, and in contrast to the "definition of evolution" assessment for which I used the same exact question repeatedly on tests and quizzes, my students only saw the "Dino Neck" question for the pre-course and final exam post-tests. Instead, in class and on quizzes I made up new questions modeled after the "Dino Neck" question that I scored using the same rubric but with different content. These new questions assessed students' understanding of evolution and coevolution for predators and prey, competitive character displacement, and mutualism.

New Practitioner Research Methods and Course Revisions in 2005 and 2006

Beginning in fall 2005, I included almost daily 5 minute short-answer surveys of students' pre-conceptions about key areas of content before these were discussed in class, although the syllabus remained the same as in previous years. However, it was apparent after the fall 2005 semester that student learning of the basic concepts of evolutionary biology was somewhat improved, but still inadequate. Many students who presented evidence on pre-tests that they harbored substantial misconceptions in fact remained highly resistant to instruction, and often defended their misconceptions using course appropriate terminology, but incorrectly, on the course final exam. In other words, many had hijacked course content in service of their misconceptions.

Beginning in fall 2006, I radically altered what I did with the pre-test data. Instead of my just reading and synthesizing this information and using it to measure the summative effects of my teaching in a positivist paradigm, I designed a plan to directly confront my students with their misconceptions and engage them in the shared struggle to dislodge these misconceptions so that expert learning could occur. Specifically, for each pre-unit survey, I presented histograms of their survey responses (prior knowledge and misconceptions) at the beginning of the next class. During this time, I specifically addressed the major categories of students' prior correct knowledge AND major misconceptions (why such and such an idea was "wrong" and why a different concept was "right"), I re-framed the outline of the class conspicuously around and sequentially addressing these major misconceptions, and I re-projected the misconceptions slide multiple times on multiple days.

For example, I showed the histogram of students' pre-test responses on the "definition of evolution" question multiple times in class, and these misconception categories formed the bases to the lecture content on evolution throughout the semester. For example, I pointed to their misconception that evolution and natural selection are synonymous (see Results), at the beginning and end of the discussion about how other factors can cause evolution such as mutation, gene flow (migration), genetic drift in small populations (founder effects, genetic bottlenecks, etc). Next, returning to natural selection I presented teleological views, such as by Lamarck, and showed why these might seem reasonable, but because they lack mechanisms these are all unfounded. Lastly, regarding the misconception that evolution by natural selection only occurs over long times, I presented evidence from the famous study by Peter and Rosemary Grant (chronicled in Jacob Weiner's [1994] *The Beak of the Finch*, the first three chapters of which are required reading) showing that populations of Darwin's finches can evolve rapidly from one generation to the next under selection. As I mentioned before, these are the same topics that I "covered" in previous years, but the difference in 2006 and 2007 was that these topics were presented in direct confrontation with students' misconceptions.

In addition, I asked them in guided discussions and turn-to-your neighbor activities to visualize and reflect upon the kinds of evidence and arguments I needed to present that would help them to understand the expert knowledge and ways of knowing I wanted them to attain versus those that many exhibited in the pre-tests. Thus, my methodological research paradigm for 2006 and 2007 moved away from viewing my students as research objects and toward a model of their engagement in their learning in a participatory action research, or post-positivist, mode (McTaggart 1997, Hollingsworth 1997, McNiff 2002, Reason and Bradbury 2006). In many instances, this had little effect on the span of content I had already slated for that class, but simply involved rearranging the sequence and changing the emphasis of this or that idea.

A second major change in the course in fall 2006 was a rather substantial content reduction and shifting of emphasis, which was mandated in part by the extra class time that needed to be devoted to increased writing and discourse-based instruction. I removed the “origin of life” story from the first third, and I moved the first half of the last third on “Individual and Population Ecology” into the beginning third. Figure 3.3 illustrates these changes in detail. The major content areas were largely unaffected. For example, basic topics in individual design (energy processing – basic biochemistry of metabolism and photosynthesis) were simply presented in the context of environmental adaptation and physiological ecology.

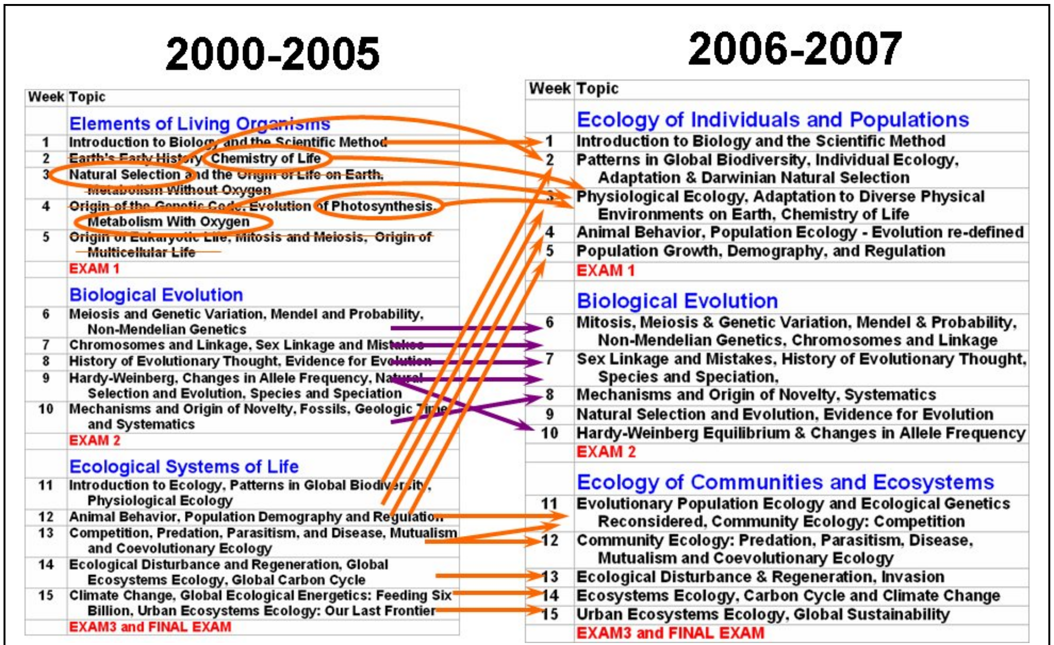


Figure 3.3 illustrates the changes in the Bio161 syllabus in 2006.

To reiterate, in my view the key innovation in 2006-2007 was to directly confront and make use of the students’ misconceptions as the bases to new outlines of topics that I then covered in class. In other words, beginning in 2006, I re-cast my whole in-class lecture pedagogy to teach directly to my students’ misconceptions about the content of my course. Nonetheless, it also was clearly a good idea to reduce the sheer volume of material I asked my students to learn and be able to manipulate.

Results

(1) Definition of Evolution.

Figure 3.4 shows the trajectories of student responses to the question “Please offer a brief and concise definition of Evolution” which appeared on the Bio161 Final Exam each December from 2000 to 2007 (n = 379). Note that the fraction of students correctly answering this question (on a 10 point scale, see rubric in Figure 3.1) between 2000 and 2005 was about 50%, which was way too low and was one of the main factors motivating my course revisions for fall 2006 (see above). Note also that there was a dramatic increase in this fraction (exceeding 90%) beginning the year of my course revisions in fall 2006, which of course was also associated with dramatic decreases in

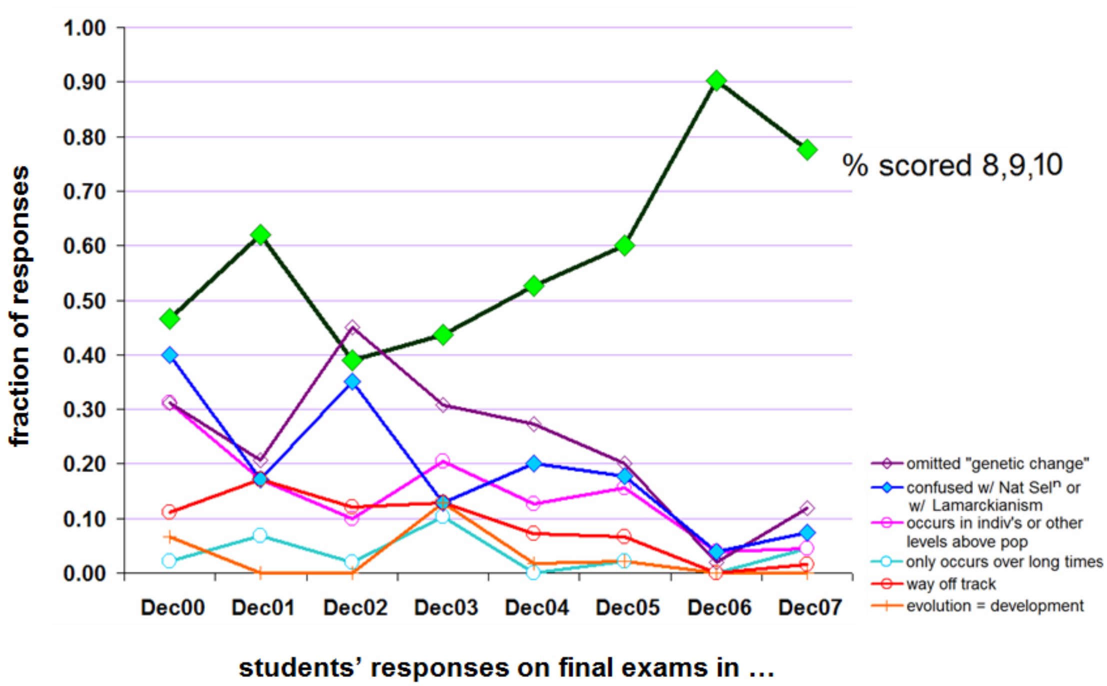


Figure 3.4. Students' scores on the final exams (2000-2007) on the "definition of evolution" assessment question. Also included are the major categories of misconceptions exhibited (see Figure 3.2).

frequencies of students' misconceptions for 2006 (all of which dropped below 5%). According to an analysis of variance on the numerical scores for this question there was a significant increase in students' scores for 2006 and 2007 relative to other years (Table 3.2, SAS Proc GLM, Duncan's Multiple Range Test).

Table 3.2. Comparison of students' scores on the final exams (2000-2007) on the "definition of evolution" assessment question (SAS Proc GLM, means with the same letter do not differ, Duncan Multiple Range Test).

	mean	1SD	N	Duncan MRT (P<0.05)			
Dec 2006	9.51	1.24	51	A			
Dec 2007	8.79	1.97	67	A	B		
Dec 2005	7.80	2.51	44		B	C	
Dec 2001	7.59	2.86	29			C	D
Dec 2004	7.11	3.23	53			C	D
Dec 2000	6.61	2.91	46				D
Dec 2003	6.51	3.28	39				D
Dec 2002	6.44	2.67	50				D

Next, I will present in more detail what happened in 2006/ 2007 that could have produced these differences among years. Figure 3.5 shows the histogram of students' responses on the "definition of evolution" question from the pre-test assessment in September 2006. I projected these results multiple times in class in fall 2006 (and a similar graph from fall 2007 [not shown]), and these misconception categories formed the bases to the lecture content on evolution throughout the semester (see Methods).

Figure 3.5. Histogram of the frequencies of students' responses on the "definition of evolution" question from the pre-test assessment in September 2006. This is the histogram I projected multiple times in class and these misconception categories formed the bases to the lecture content on evolution.

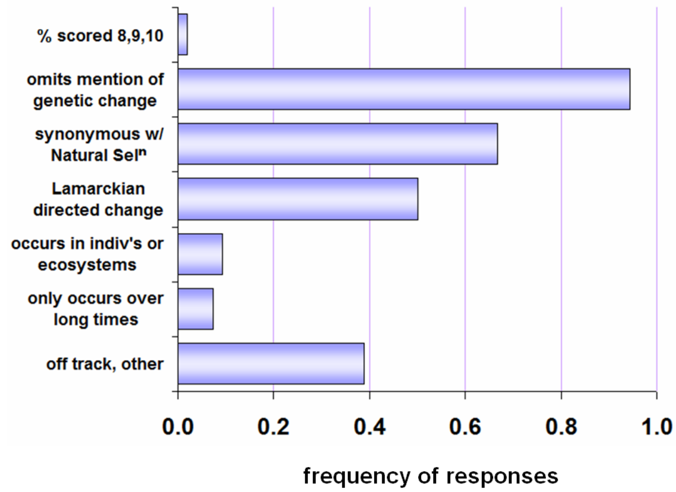


Table 3.3 shows students' scores on the pre-test, quizzes, and exams for (A) fall 2005, (B) fall 2006, and (C) fall 2007 on the "definition of evolution" assessment question. Note in all three years for which I have pre-post data, there were huge increases in the fraction of students exhibiting an understanding of the meaning of the word "evolution" as a result of taking my course.

Table 3.3. Pair-wise t-tests on students' scores on the pre-test vs. final exam for fall 2005, 2006, and 2007 on the "definition of evolution" assessment question.

	pre-test	**final exam**	**t-stat**	**d.f.**	**P**
fall 2005	**3.78**	**7.84**	**10.20**	**44**	**<0.001**
fall 2006	**4.04**	**9.51**	**21.02**	**50**	**<0.001**
fall 2007	**3.99**	**8.79**	**18.56**	**66**	**<0.001**

Figure 3.6 shows the trajectories of students' scores on the pre-test, quizzes, and exams for (A) fall 2005, (B) fall 2006, and (C) fall 2007 on the "definition of evolution" assessment question over the course of the semester. Note the huge increases in the fraction of students exhibiting an understanding of the meaning of the word "evolution" and decreases by the end of the semester in the major categories of misconceptions noted in Figure 3.5. In particular note that for 2006 and 2007 students improved much more rapidly earlier in the semester than they did in 2005 with sharp declines in frequencies of all misconceptions by the end of the first unit.

Interestingly, note also that in 2006, a sharp drop occurred between “Exam1 Oct06” and “Quiz7 Nov06” between which many students apparently slipped out of calibration during the 4 week unit on genetics. For fall 2007, with this knowledge from fall 2006, we made adjustments in the course (more reviewing) to enhance student retention. Note that in fall 2007, there was a much smaller dip in performance at this same stage. In addition, individual student’s trajectories reveal tremendous variation in performance from one assessment event to the next (see Supplemental Figure S2).

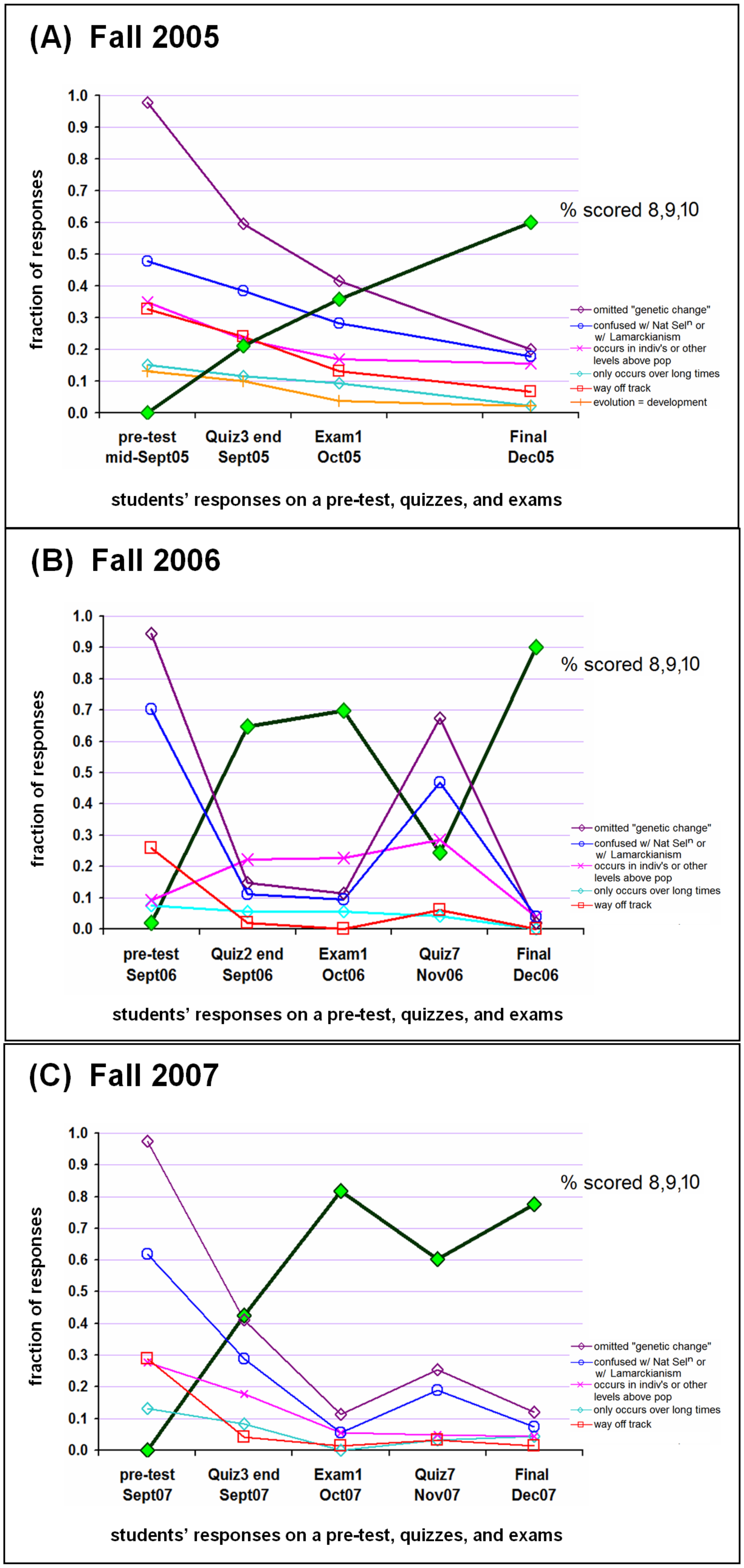


Figure 3.6. Students’ scores on the pre-test, quizzes, and exams for (A) fall 2005, (B) fall 2006, and (C) fall 2007 on the “definition of evolution” assessment question. Also included are the major categories of misconceptions exhibited (see Figure 3.4).

(2) Evolution by Natural Selection: the “Dino Neck” question.

Figure 3.7 shows a summary of the “Dino Neck” question scoring rubric (see Table 3.1), and the frequencies of students’ scores on the final exams. Note that for 2000-2005 (at right), very few of the students scored 8, 9, or 10 (about 3%), thus few understood how to answer this question even after considerable instruction on numerous occasions about the process of evolution by natural selection. In fact, most students (about 85%) exhibit major misconceptions, such as that phenotypic variation in the tree and animal stem from directed evolution based on Lamarckian or teleological need. In contrast, the results for 2006-2007 (Figure 3.7, right) show huge improvement in the fraction of students who scored 8, 9, or 10 (about 54%). Thus, more than half of my students correctly answered this question following course revision, and the average scores were significantly higher as well (for 2000-2005, average = 4.38, n = 265, and for 2006-2007, average = 7.36, n = 118; t-test (≠ var) = 12.70, $\underline{P} < 0.001$).

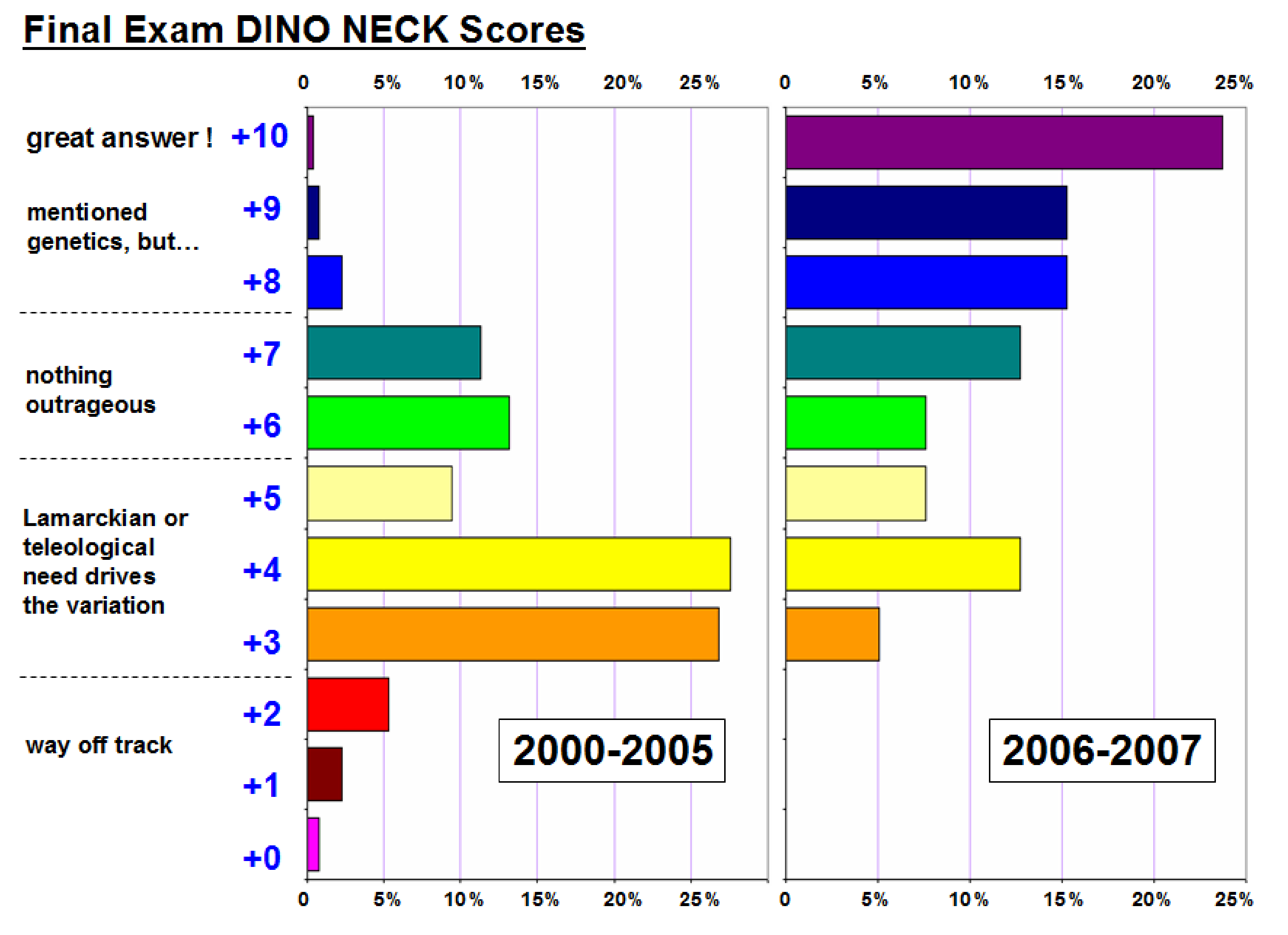


Figure 3.7. Simplified “Dino Neck” question scoring rubric on the left (see Table 1 for complete rubric), and frequencies of students’ scores on the final exams for 2000-2005 and 2006-2007.

Figure 3.8 shows students' pre-post scores on the "Dino Neck" question for 2000 to 2007. Note that for 2000 to 2005 pre- and post- scores were all very similar and abysmally low (averaging between 4 and 5), which stemmed from major misconceptions that my course did not displace (see Table 3.1). According to a pairwise t-test for all pre-post data for 2000-2005, the post-test scores were significantly but only slightly higher than the pre-tests (4.01 vs. 4.40, pair wise t-stat = 3.74, P <0.001, n = 255). However, note that both averages are between 4 and 4.5, which means that many of my students entered and exited my course maintaining deeply non-Darwinian misconceptions that were inert to instruction in these years. In contrast, in 2006 and 2007 post-test scores on the final exam were significantly and hugely higher than pre-test scores (3.56 vs. 7.34, pair wise t-stat = 16.71, P<0.0001, n = 115).

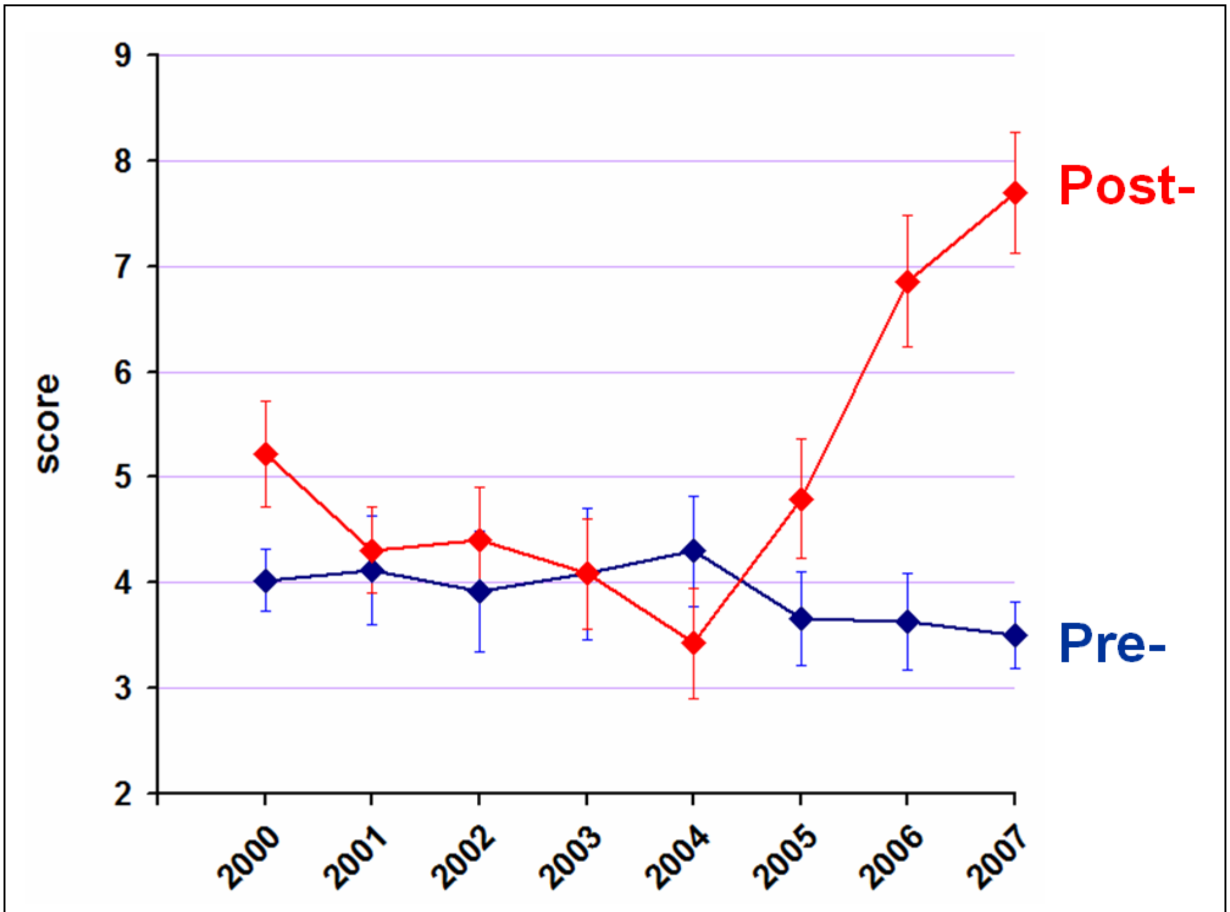


Figure 3.8. Averages of students' scores on the pre- post-tests on the final exam "Dino Neck" question for 2000 to 2007, n = 370).

When one compares students' "Dino Neck" pre-post scores, one can see the dramatic effect of course revision on most students' understanding of evolution at the individual level. Figure 3.9 shows the counts of the number of students with a given "Dino Neck" score at the beginning of the course (columns) versus the same student's score at the end of the course (rows) for the period 2000-2005 (left) and 2006-2007 (right). In other words, each cell shows the count of students by score pre-course in September versus their post-course score in December.

For 2000-2005 (Figure 3.9, left), although there are about twice as many students above left (n=118) than below right (n=64) of the diagonal (i. e., more improved than not), only 15% of post-tests scored above 7 at the end of the course, which is really the minimum for Darwinian thinking (see Table 3.1). In contrast, for 2006-2007 (Figure 3.9, right), 90% of the students scored above the diagonal post-test (% = 105/115) whereas only 3% scored worse on the post-test. Thus, under the revised course design, 90% of my students improved in their ability to demonstrate an understanding of evolution by natural selection.

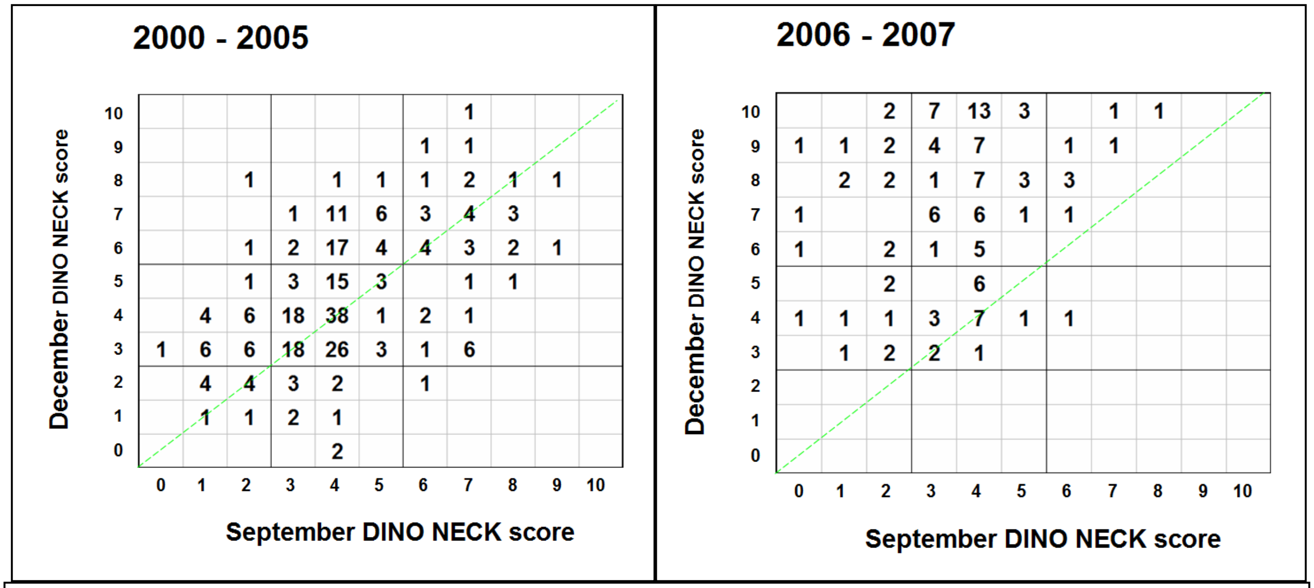


Figure 3.9. Counts for students' September Dino Neck score versus the same student's December Dino Neck score for 2000 – 2005 (left) and 2006 – 2007 (right).

Table 3.4 shows detailed results for students' misconceptions as evidenced by the text in their extended essay responses to the pre- vs. post DINO NECK assessment for years 2000 – 2005 combined. Very few students either pre- or post course showed evidence that they understood that variation among coexisting individuals (Trees or Dinos) was relevant to the explanation (rows H and I). Most only realized that phenotypes differed in time (rows F and G). Related to this, nearly 90% of the students recognized that short necked Dinos were hungry or starving (row E), but more than 60% attributed phenotypic change to teleologically based need, rather than to differential survival in the population (row K). Most had no explanation for variation in trees.

Very few students implicated heritable variation or genetics in any context and even fewer articulated any role played by random mutation in creating variation (rows L to U, all < 20%). Interestingly, there was a significant increase in those mentioning mutation for the Dino's (row O), but much of this increase was due to the misconception (which increased from 1% to 9%) that invoked "directed mutation" to justify their belief that teleologically based need drives variation to elongate the Dino necks (row Q). Thus, for these students, because my course failed to displace their prior non-Darwinian misconceptions, the content from the course's genetics unit was merely lifted and inappropriately applied to reinforce these prior misconceptions.

Table 3.4 compares the aggregate responses for the pre- vs. post course “Dino Neck” assessment tests for 2000 — 2005 years combined (n = 579). Statistical comparisons used Fisher Exact tests on all surveys. One advantage of the Fisher test is its simplicity, although data on individual variation is not considered.

	All responses 2000 to 2005	SEPT n=305	DEC n=274	Δ	P value
A	scored 8, 9, or 10	0.03	0.03	0.00	N.S.
	➢ few students (3%) excelled on this question (≥8) and there was no difference pre- vs. post				
B	TREEs just grew (ontogeny)	0.61	0.43	-0.18	P<0.001
C	DINOs just grew (ontogeny)	0.11	0.02	-0.09	P<0.001
	➢ significantly fewer students maintained post course that the trees and Dino’s merely grew; thus; more students articulated a mechanism for change other than ontogeny				
D	smaller TREEs eaten more	0.27	0.41	+0.14	P<0.001
	➢ significantly more students recognized the key point that tree height is under selection since short necked Dino’s eat short trees, but the number was only 41% at the end of the course				
E	shorter DINOs starved	0.84	0.88	0.04	N.S.
	➢ there was no difference in the % of students who recognized that Dino neck length was under selection pre- vs. post; however, the % was well into the 80’s				
F	TREE pheno's vary in time	0.91	0.89	-0.02	N.S.
G	DINO pheno's vary in time	0.90	0.91	0.00	N.S.
	➢ nearly all students noted that there were changes in the phenotypes of the trees and Dino’s in time (about 90%)				
H	coexisting TREEs vary	0.09	0.18	+0.08	P<0.01
I	coexisting DINOs vary	0.22	0.32	+0.10	P<0.01
	➢ although significantly more students recognized post-test the key concept that coexisting trees and Dino’s must exhibit variation in order for natural selection to distinguish among them – these %’s were very low for both				
J	TREE NEEDs caused the Δ	0.26	0.26	0.01	N.S.
	➢ about ¼ of the students maintained that teleologically based needs caused the evolution of taller trees for both pre- and post				
K	DINO NEEDs caused the Δ	0.70	0.62	-0.07	P<0.05
	➢ more than ⅔ of the students maintained that teleologically based needs caused the evolution of longer Dino necks pre-course, and although this number declined significantly, the post course value of 62% is way too high,				
L	progeny of tall TREEs are tall	0.03	0.03	0.00	N.S.
M	progeny of long neck DINOs are long	0.10	0.12	0.01	N.S.
	➢ very few students noted that offspring characteristics are related to those of their parents, and my course had no effect on the frequencies of this oversight for the trees and Dino’s,				
N	TREE w/ mutation or any genetic change	0.04	0.07	0.03	N.S.
O	DINO w/ mutation or any genetic change	0.08	0.17	0.09	P<0.01
P	TREE w/ directed mutation or genetic Δ	0.01	0.02	0.02	N.S.
Q	DINO w/ directed mutation or genetic Δ	0.01	0.09	0.08	P<0.001
	➢ very few students implicated any role of mutation in this scenario, and although there was a significant increase in those mentioning mutation for the Dino’s, much of this increase was				

	due to a misconception that invoked "directed mutation" to justify their belief that teleologically based need drives variation to elongate the Dino necks				
R	**TREE w/ random mutation or genetic Δ**	**0.00**	**0.00**	**0.00**	**N.S.**
S	**DINO w/ random mutation or genetic Δ**	**0.00**	**0.02**	**0.01**	**N.S.**
T	**TREE w/ mutation followed by selection**	**0.02**	**0.02**	**0.00**	**N.S.**
U	**DINO w/ mutation followed by selection**	**0.03**	**0.04**	**0.01**	**N.S.**
	➢ trivial percentages (all < 5%) implicated the expert understanding that random mutation was the initial source of variation in tree height and neck length, and few also correctly sequenced events of selection as following and acting upon these mutant varieties				
V	**explanation way off**	**0.23**	**0.13**	**-0.10**	**P<0.01**
W	**poor writing skills**	**0.14**	**0.09**	**-0.04**	**N.S.**
X	**less than 5 line answer**	**0.15**	**0.32**	**0.17**	**P<0.01**
	➢ at the very least, significantly fewer students were classified as "way off", incidence of "poor" writing was unchanged at around 10%, and but significantly more students offered fewer than 5 lines for their answer, suggesting they simply did not understand how to respond				

Table 3.5 compares the aggregate responses for the pre- vs. post course tests for years 2006 and 2007 combined (n = 247). About 2/3 of my students post course showed evidence that they understood that variation among coexisting individuals (trees or Dinos) was relevant to the explanation (rows H and I) and nearly all realized that phenotypes differed in time (rows F and G). Related to this, over 95% of the students recognized that short necked Dinos were hungry or starving (row E), only 1/4 attributed phenotypic change to teleologically based need, rather than differential survival in the population, which was significantly lower than for the pro-test (row K). Post-test results were similar for trees.

Significantly more students implicated heritable variation, genetics, and the important role played by random mutation in creating variation (rows L to O, and R to U). Interestingly, there was still significant increase in those defending the misconception (which increased from 2% to 10%) that invoked "directed mutation" to justify their belief that teleologically based need drives variation to elongate the Dino necks (row Q). Thus, for these students, my course still failed to displace their prior non-Darwinian misconceptions (see Table 3.4, row Q, as well).

Table 3.5 compares the aggregate responses for the pre- vs. post course "Dino Neck" assessment tests for 2006 and 2007 combined (n = 247). Comparisons used Fisher Exact tests.

	All responses 2006 to 2007	**SEPT n=129**	**DEC n=118**	**Δ**	**P value**
A	**scored 8, 9, or 10**	**0.02**	**0.54**	**+0.53**	**P<0.001**
	➢ significantly more students excelled on this question post-course				
B	**TREE just grew (ontogeny)**	**0.61**	**0.23**	**-0.38**	**P<0.001**
C	**DINO just grew (ontogeny)**	**0.06**	**0.01**	**-0.05**	**P<0.03**
	➢ significantly fewer students maintained post course that the trees and Dino's merely grew; thus, more students articulated a mechanism for change other than ontogeny				
D	**smaller TREEs eaten more**	**0.22**	**0.74**	**+0.52**	**P<0.001**

E	**shorter DINOs starved**	**0.79**	**0.96**	**+0.17**	**P<0.001**
	➢ significantly more students recognized the key point that tree height and Dino neck length were under selection				
F	**TREE pheno's vary in time**	**0.87**	**0.91**	**0.04**	**N.S.**
G	**DINO pheno's vary in time**	**0.85**	**0.92**	**0.06**	**N.S.**
	➢ nearly all students noted that there were changes in the phenotypes of the trees and Dino's in time (> 90%)				
H	**coexisting TREEs vary**	**0.07**	**0.68**	**+0.61**	**P<0.001**
I	**coexisting DINOs vary**	**0.16**	**0.86**	**+0.70**	**P<0.001**
	➢ there were huge and significant increases in students' recognition of the key concept that coexisting trees and Dino's must exhibit variation in order for natural selection to distinguish among them – the ending percentages on the post-tests were 68% and 86% for trees and Dino's, respectively, which shows huge gains				
J	**TREE NEEDs caused the Δ**	**0.25**	**0.16**	**-0.09**	**N.S.**
	➢ about 16% of the students maintained that teleologically based needs caused the evolution of taller trees for both pre- and post which is not significantly lower than the pre-course value of 25%, thus there is room for improvement				
K	**DINO NEEDs caused the Δ**	**0.75**	**0.24**	**-0.51**	**P<0.001**
	➢ significantly fewer students maintained that teleologically based needs caused the evolution of longer Dino necks post-course				
L	**progeny of tall TREEs are tall**	**0.03**	**0.40**	**+0.37**	**P<0.001**
M	**progeny of long neck DINOs are long**	**0.09**	**0.56**	**+0.47**	**P<0.001**
	➢ significantly more students specifically noted that offspring characteristics are related to those of their parents post-course, however, these numbers are below 50% on the key aspect of the mechanisms of evolution by natural selection, so there is much room for improvement				
N	**TREE w/ mutation or any genetic change**	**0.05**	**0.73**	**+0.68**	**P<0.001**
O	**DINO w/ mutation or any genetic change**	**0.07**	**0.84**	**+0.77**	**P<0.001**
	➢ there were huge and significant increases in the numbers of students who implicated any role of mutation or genetic change in their narrative				
P	**TREE w/ directed mutation or genetic Δ**	**0.02**	**0.06**	**0.04**	**N.S.**
Q	**DINO w/ directed mutation or genetic Δ**	**0.02**	**0.10**	**+0.08**	**P<0.01**
	➢ still present was an increase in the misconception that invoked "directed mutation" to justify some students' belief that teleologically based need drives variation to elongate the Dino necks				
R	**TREE w/ random mutation or genetic Δ**	**0.00**	**0.45**	**+0.45**	**P<0.001**
S	**DINO w/ random mutation or genetic Δ**	**0.00**	**0.47**	**+0.47**	**P<0.001**
T	**TREE w/ mutation followed by selection**	**0.01**	**0.56**	**+0.55**	**P<0.001**
U	**DINO w/ mutation followed by selection**	**0.01**	**0.63**	**+0.62**	**P<0.001**
	➢ there were significant increases in the numbers of students who implicated the expert understanding that random mutation was the initial source of variation in tree height and neck length, and also who correctly sequenced events of selection as following and acting upon these mutant varieties, however, the final rates ranged from 45% to 63%, which indicated there is still much room for improvement,				
V	**explanation way off**	**0.23**	**0.07**	**-0.16**	**P<0.001**
W	**poor writing skills**	**0.15**	**0.07**	**-0.08**	**P<0.05**
X	**less than 5 line answer**	**0.37**	**0.03**	**-0.34**	**P<0.001**
	➢ significantly fewer students were classified as "way off", the incidence of "poor" writing declined, and more than 95% of students offered at least 5 lines for their answer				

Table 3.6 compares the aggregate post-test responses before versus after course (2000-2005, n = 274 vs. 2006-2007, n = 118). Post-test scores for students prior to the course revision are so similar to pre-test scores in 2006-2007, that the results for Table 6 below are very similar to those just described in Table 3.5. Following course revision significantly more students no longer harbored major misconceptions impeding their understanding of evolution by natural selection. For example, more of the 2006-2007 students recognized the importance of variation among coexisting individuals to the explanation (rows H and I). More also recognized that shorter trees and shorter necked Dinos were selected against (rows D and E), and fewer attributed phenotypic change in trees and Dino's to teleologically based need, rather than differential survival in the population (rows J and K). Significantly more students implicated heritable variation, genetics, and the important role played by random mutation in creating variation (rows L to O, and R to U). Lastly, following course redesign, there were significant increases in the numbers of students who correctly sequenced the events of selection as following and acting upon these mutant varieties through differential reproductive success.

Table 3.6 compares the aggregate responses for the post course "Dino Neck" assessment tests for 2000 to 2005 combined (n = 274) versus 2006 to 2007 combined (n = 118). Comparisons used Fisher Exact tests.

	for all responses DEC 2000-2005 versus DEC 2006-2007	**2000-2005 n=274**	**2006-2007 n=118**	**Δ**	**P value**
A	**scored 8, 9, or 10**	**0.03**	**0.54**	**0.51**	**P<0.001**
	➢ significantly more students excelled on this question following the course revision				
B	**TREE just grew (ontogeny)**	**0.43**	**0.23**	**-0.20**	**P<0.001**
C	**DINO just grew (ontogeny)**	**0.02**	**0.01**	**-0.01**	**N.S.**
	➢ significantly fewer students in 2006-2007 maintained post course that the trees and Dino's merely grew; thus, more students articulated a mechanism for change other than ontogeny following course revision				
D	**smaller TREEs eaten more**	**0.41**	**0.74**	**0.33**	**P<0.001**
E	**shorter DINOs starved**	**0.88**	**0.96**	**0.08**	**P<0.01**
	➢ significantly more students post revision recognized the key point that tree height and Dino neck length were under selection				
F	**TREE pheno's vary in time**	**0.89**	**0.91**	**0.02**	**N.S.**
G	**DINO pheno's vary in time**	**0.91**	**0.92**	**0.01**	**N.S.**
H	**coexisting TREEs vary**	**0.18**	**0.68**	**0.50**	**P<0.001**
I	**coexisting DINOs vary**	**0.32**	**0.86**	**0.54**	**P<0.001**
	➢ in 2006-2007 there were huge and significant increases in students' recognition of the key concept that coexisting trees and Dino's must exhibit variation in order for natural selection to distinguish among them, which shows huge gains under the revised course design				
J	**TREE NEEDs caused the Δ**	**0.26**	**0.16**	**-0.10**	**P<0.02**
K	**DINO NEEDs caused the Δ**	**0.62**	**0.24**	**-0.38**	**P<0.001**
	➢ significantly fewer students maintained that teleologically based needs caused the evolution of taller trees and longer Dino necks following course revision				
L	**progeny of tall TREEs are tall**	**0.03**	**0.40**	**0.37**	**P<0.001**
M	**progeny of long neck DINOs are long**	**0.12**	**0.56**	**0.44**	**P<0.001**
	➢ significantly more students noted that offspring characteristics are related to their parents'				

N	**TREE w/ mutation or any genetic change**	**0.07**	**0.73**	**0.66**	**P<0.001**
O	**DINO w/ mutation or any genetic change**	**0.17**	**0.84**	**0.67**	**P<0.001**
	➢ there were huge and significant increases in the numbers of students who implicated any role of mutation or genetic change in their narrative post revision				
P	**TREE w/ directed mutation or genetic Δ**	**0.02**	**0.06**	**0.04**	**P<0.05**
Q	**DINO w/ directed mutation or genetic Δ**	**0.09**	**0.10**	**0.01**	**N.S.**
	➢ course redesign did not affect that about 10% of the students still maintain that directed evolution affects variation in Dino neck length, and there is even a slight increase in this thinking among some to explain variation in tree height				
R	**TREE w/ random mutation or genetic Δ**	**0.00**	**0.45**	**0.45**	**P<0.001**
S	**DINO w/ random mutation or genetic Δ**	**0.02**	**0.47**	**0.46**	**P<0.001**
T	**TREE w/ mutation followed by selection**	**0.02**	**0.56**	**0.54**	**P<0.001**
U	**DINO w/ mutation followed by selection**	**0.04**	**0.63**	**0.59**	**P<0.001**
	➢ following course redesign, there were significant increases in the numbers of students who implicated the expert understanding that random mutation was the initial source of variation in tree height and neck length, and also who correctly sequenced events of selection as following and acting upon these mutant varieties				
V	**explanation way off**	**0.13**	**0.07**	**-0.06**	**P<0.05**
W	**poor writing skills**	**0.09**	**0.07**	**-0.03**	**N.S.**
X	**less than 5 line answer**	**0.32**	**0.03**	**-0.28**	**P<0.001**
	➢ significantly fewer students were classified as "way off" and significantly more students offered at least 5 lines for their answer post course redesign				

Comparing students' responses at the end of the course on the "Dino Neck" and "definition of evolution" questions reveals an important disconnect in learning that was significantly mitigated by course redesign. Figure 3.10 shows the counts of students by score for the "definition of evolution" question (columns) vs. counts by score for the "Dino Neck" question (rows). In 2000-2005 (left), note that a large number of students who excelled at defining evolution (scoring 8, 9, or 10) showed little or no ability to apply this definition to the conceptual application in the "Dino Neck" question (about 50% lie in the "disconnect" circle in Figure 3.10, left). In fact, students' pairwise responses were uncorrelated ($R^2 = 0.012$, $n = 247$, $\underline{P} > 0.05$). This suggests that another important layer of latent misconceptions is held by my students about how to apply knowledge across cognitive domains, e. g. the application of factual knowledge to challenges of knowledge synthesis.

The same data collected from students after the course revision in 2006-2007 (Figure 3.10, right) reveals a significant reduction in this disconnect ("in" defined as $8 \leq x \leq 10$ and $y \leq 7$, 2*2 Fisher's Exact test, $\underline{P}<0.001$). Indeed, most of the students performed well on both questions (scores of 8, 9, or 10) in contrast to the dispersion of scores in the 2000-2005 era. Thus, the combination of course revisions also affected an important outcome in student learning across cognitive domains in 2006 – 2007, i.e. students were better able to apply factual knowledge about evolution to the challenges of knowledge synthesis in a narrative explanation of the process of evolution by natural selection. Devising teaching methods that enable students to make this type of connection across cognitive domains (as in Bloom's Taxonomy {link to *TIEE Glossary*}) pose another layer of important research questions for this project.

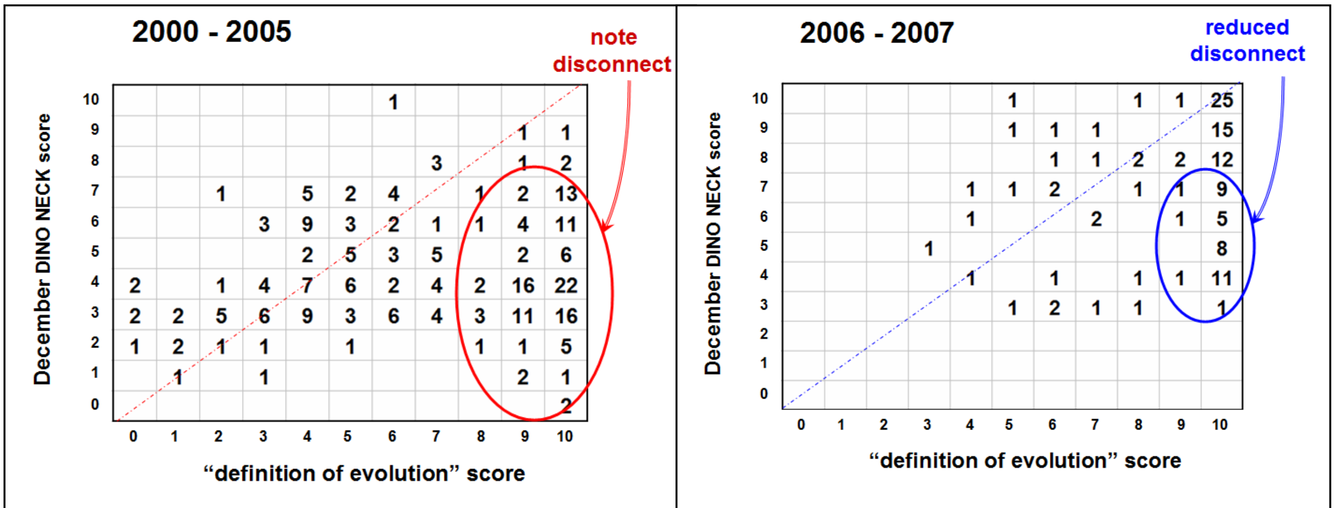


Figure 3.10. Comparison of final exam "Dino Neck" vs. "definition of Evolution" scores for all students in 2000-2005 (left, n = 247) and 2006-2007 (right, n = 118).

(4) My "Personal Theory of Practice" to Account for My Students' Conceptual Change in Bio161.

Results indicate that the suite of course revisions I instituted in fall 2006 significantly decreased students' misconceptions and improved their learning about evolution and the process of evolution by natural selection. The revisions included course content modification and reduction and radical redesign of the presentation modality of this course content. In particular, I used student pre-test responses to estimate class-wide frequencies of misconceptions about evolution and the process of evolution by natural selection. I then presented frequency histograms of their misconceptions directly to them in class, engaged them in reflective discourse to address these misconceptions, and then I rearranged and presented course content to systematically confront, address, and displace these misconceptions as "expert" knowledge and ways of knowing were transmitted. Thus, the students' demonstrated misconceptions drove the course outline and methods of content delivery.

Results indicate that major types of misconceptions decreased in students' post-test responses following instruction in the revised modality. For example, more students had recognized the importance and nature of phenotypic variation as occurring among co-existing individuals, more students understood the role of random mutation as the origin of these within population variants, more students appreciated the action of natural selection as selecting among these coexisting variants, and more students had jettisoned teleological/ Lamarckian explanations for evolutionary change based on need. Finally, more students demonstrated improved cognitive capacity at applying factual knowledge across cognitive domains to synthesize information in the context of an evolutionary narrative. In sum, more of them got it and I claim that they did so in large part because they were actively engaged in the research process into their own learning.

As Howard Gardner (2004) so eloquently relayed, “It is my belief that, until recently, those of us involved in education have not appreciated the strength of the initial conceptions, stereotypes, and ‘scripts’ that students bring to their school learning nor the difficulty of refashioning or eradicating them.” (p. 5) The challenge of identifying students’ misconceptions and designing pedagogies to dislodge them is an important design principle of constructivist classrooms (Fox 2001). A very large literature base exists applying these principles to address students’ misconceptions about ecology and evolution (Munson 1994, Dagher and BouJaoude 1997, Alters and Nelson 2002, Anderson *et al*. 2002, Bowman 2008, Hokayem and Boujaoude 2008, and references therein). Upon examination, all of the misconceptions categories I have detected as prevalent in my class have been previously identified in the literature as being deeply resistant to instruction. However, I know of no other studies that have undertaken as intensive a program of systematic long-term practitioner research on evolution as I report here, nor one that uses misconceptions inventories to drive the curriculum as I describe here.

This project underscores the value of long-term practitioner research in two major ways. Firstly, data from pre-post tests in 2000-2005 showed that important concepts in evolutionary biology were not being learned, and in spite of repeated instruction, I was not displacing their easily identifiable and deeply entrenched misconceptions about evolution. These data both mandated and molded the course revision. Secondly, the practitioner research methodology provides guidance away from a positivist research paradigm and more towards a program of reflection and engagement (Stenhouse 1975, Jarvis 1999, Allwright 2005). Whereas positivism had very successfully led to identifying exactly why my curricula and instructional methods were less effective than I desired, practitioner research unveiled a pedagogy for my course towards a modality of misconceptions confrontation and reflective learning by my students that was truly transformative not only to them but also to me (Cochran-Smith and Lytle 1999). In sum, I brought their data directly to them in an “action research” model to enlist and engage them in reflective problem solving in their own learning (Wandersee *et al*. 1994, Hollingsworth 1997, Allwright 2005).

I offer that the approach of confronting student misconceptions was effective for several reasons. Firstly, I believe that my students perceived at least a partial transfer of ownership of the traditional lecture material from my domain, where I determined what was presented how and when, to a modality in which their attributes (prior knowledge, misconceptions, AND struggles to learn new content and concepts) were elevated in prominence. The message was clear from me that what they knew and didn’t know mattered to me and greatly affected what we “covered,” when we covered it, and how we covered it in class.

Perhaps it is because this is a very narcissistic generation that lights up when they hear talk about themselves. But, I enjoin more likely it is because most of my students may have yet to really learn how to be active listeners and participants in a lecture-based classroom, and thus many feel lost as soon as they walk in the room. But when I projected a histogram of misconceptions (Figure 3.5) and I talked about them <u>to them</u> in class, they would all be looking at me, listening intently, focused upon my observations about THEM, and with their reflections on their pre-tests in hand, they were engaged in the challenge of figuring out what needed to happen in and out of class for each person to improve in the course. Importantly, I believe that I cared more about them because I started seeing that what I was doing was having more of an effect. Positive reinforcement matters to me, too. Thus, I offer in closing that a major reason why student performance increased was because the course revisions made the learning process more personal, relevant, and accessible to them – their struggle with learning this material was really about their struggle with learning about themselves — and, the same applies to me, which I will elaborate upon next.

Among the most important findings I report to date is a very personal one – I believe that over the past eight years while researching my teaching and student learning in this course, I have learned several key analytical approaches that have equipped me with important and immediate positive responses to when my students do not perform up to my hopes and expectations. Instead of resorting to deficit explanations (Valencia 1997, Elmersky and Tobin 2005) when students fail, blaming either them, me, others around me, or all of the above, I now have a clear set of evidence-rich alternative directions in which to focus my efforts and design the next round of course revisions. Grading quizzes and exams for example has become much more of an intriguing data collection activity and much less of a mind-numbing and depressing drudgery. I have regained a lot of the thrill of discovery learning that got me into this profession to begin with. For me, I think this is the most important outcome of "practitioner research."

Many however would counter that it's all well and good for me to feel better about my teaching, but consequences matter the most – what is the evidence that my students are learning evolutionary ecology better?

The evidence I presented to address this question comes from my efforts to teach the meaning of the term "evolution" and the process of "evolution by natural selection." Results indicate that when I added specific types of classroom instruction which I call "reflective constructivism" in fall 2006 as part of the TIEE "practitioner research" program, there were reduced frequencies of major misconceptions and significant improvements in students' learning of the content and process of evolution by natural selection. However, I am still at best only seeing 60-80% of my students demonstrate understanding of these fundamental issues in evolutionary biology. Clearly, I have a long way to go, and I am not alone. This need for teaching teachers how to conduct practitioner research was articulated in the USA more than a decade ago in the National Science Education Standards in which a key facet of teacher professional development was to change teachers from "consumers of knowledge about teaching" to "producers of knowledge about teaching" (NRC 1996).

Future Directions:

(1) My principal next steps involve a substantial redesign of the reflective learning components of Biology 161. I seek to improve and give structure to my students' metacognitive processing abilities and cause them to reflect upon their own learning to displace inaccurate misconceptions they harbor not only about the course content but more importantly I need them to face their own misconceptions about how they learn and enhance their self efficacy (Baldwin *et al.* 1999) and self actualization as learners. What in-class activities, readings, study habits, peer-networks, and assignments will

make them more self aware of their own learning and the effects of their prior knowledge and ways of knowing on their accomplishing the tasks I set before them?

(2) I need to make greater use of published misconceptions inventories (e.g. Anderson *et al*. 2002) and analysis methods (e. g. ethnographic analysis, Tobin *et al*. 2001) to improve my identification and classification of the misconceptions my students harbor as well as enable my students to self diagnose through online assessments and course discussion forums.

(3) I need to make greater use of multivariate statistics to cluster “errors” and identify underlying “latent misconceptions” not only to reduce the pure number of issues to address, but also to adopt more of a systems approach to designing misconceptions interventions.

(4) I need to explore experimentally, by making use of the different sections of the Bio161 course, what methods or combination of methods best displaces students’ diverse ecology misconceptions beyond just those related to evolution.

(5) The pattern of temporal variation in individual students’ scores on the “definition of Evolution” question in Figure 2.1 above and in Supplemental Figure S2 is both astonishing and alarming. Why do students exhibit such dynamic temporal variation in performance on the same question over such a short time? How can I improve my use of class time and design better assignments to decrease this variation and improve all students’ scores?

(6) How do I use misconceptions pre-tests to identify students at risk (or who are advanced) and design interventions to improve their success and retention?

(7) How do I pursue the next logical step in “practitioner research” to engage my students in genuine “action research” into their own learning (Hollingsworth 1997, Elmesky and Tobin 2005)? I would model such a program on published advances in freshman learning communities and theories of social constructivism, and teach my students to research their own learning styles, processes, and outcomes in parallel to my standard measures of course performance. My bet is that the social dimension will accelerate metacognitive development and thereby affect increases in academic performance, agency, and self-efficacy.

Collaborators on any of the above topics would be more than welcome! Please email me at bwgrant@widener.edu to get involved.

Acknowledgements:

This paper is based on (1) the presentation “Practitioner Research as a ‘Way of Knowing’ to Improve Student Reflective Learning in Ecological Education. Invited oral presentation at the special session *“Blue Sky Thinking in Ecological Education: Starting from Scratch”* organized by David Slingsby for the annual meeting of the British Ecological Society, Imperial College, London, UK, 3 – 5 September 2008, and (2) the presentation “Practitioner Research Improved My Students’ Understanding of Evolution by Natural Selection in an Introductory Biology Course at the National Research Council Board On Science Education workshop “Linking Evidence and Promising Practices in STEM Undergraduate Education,” 30 June 2008, at the National Academy of Sciences, Washington, DC. The sections on my case study on Bio161 were improved by comments by Chris Beck.

I thank members of the National Research Council, Center for Education, Board on Science Education for the invitation to the 30 June 2008 NAS workshop and for the opportunity to submit this paper.

This research was supported in part from grants from the National Science Foundation NSF DUE-CCLI-EMD 04-43714, 01-27388, and 99-52347, upon all three of which I am co-PI with Charlene D’Avanzo (Hampshire College). In addition, support came from the Mathematics Science Partnership of Greater Philadelphia (MSP-GP) NSF DUE- MSP Targeted Awards 03-14806, on which Widener University is a collaborative institution of higher education, as well as MSP Start Partnership, NSF DUE MSP-START 08-32049 upon which I am co-PI with Steve Madigosky (Widener University), Victor Donnay (Bryn Mawr College), Thomas Schrand (Philadelphia University), and William Keilbaugh (Haverford SD).

Literature Cited:

Alder, J. 2008. Surprise. *Educational Theory* 58(2): 149-173.

Alters, B. J., and C. E. Nelson. 2002. Perspective: Teaching evolution in higher education. *Evolution* 56: 1891–1901.

Allwright, D. 1984. Why don’t learners learn what teachers teach?—The interaction hypothesis. In D. M. Singleton & D. G. Little (eds.), *Language Teaching in Formal and Informal Contexts* (pp. 3–18). Dublin, Ireland: BAAL/IRAAL.

Allwright, D. 2005. Developing principles for practitioner research: The case of exploratory practice. *The Modern Language Journal* 89: 353–366.

Anderson, D. L., K. M. Fisher, and G. J. Norman. 2002. Development and evaluation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching* 39: 952–978.

Ayala, F. 2008. Coda: Science as a way of knowing. Counterbalance http://www.counterbalance.net/evolution/know-body.html {accessed 10 July 2008}

Baldwin, J. A., D. Ebert-May, and D. J. Burns. 1999. The development of a college biology self-efficacy instrument for non-majors. *Journal of Research in Science Teaching* 36: 397-408.

Barton, A. C. 2001. Science education in urban settings: Seeking new ways of praxis through critical ethnography. *Journal of Research in Science Teaching* 38(8): 899-917.

Bassey, M. 1992. Creating education through research. *British Educational Research Journal* 18(1): 3-16.

Bishop, B., and C. Anderson. 1990. Students' conceptions of natural selection and its role in evolution. *Journal of Research in Science Teaching* 27: 415–427.

Bowman, K. L. 2008. The evolution battles in high-school science classes: who is teaching what? *Frontiers in Ecology and the Environment* 6: 69–74.

Bransford, J. D., A. L. Brown, & R. R. Cocking. 2000. *How People Learn: Brain, Mind, Experience, and School*. Washington, DC: National Academies Press.

Brem, S. K., M. Ranney, and J. Schindel. 2003 . Perceived consequences of evolution: College students perceive negative personal and social impact in evolutionary theory. *Science Education* 87: 181–206.

Carr, W. 2006. Philosophy, methodology and action research. *Journal of the Philosophy of Education* 40 (4): 421-435.

Chandler, D., and B. Torbert. 2003. Transforming inquiry and action: Interweaving 27 flavors of action research. *Action Research* 1: 133-152.

Claxton, G. 1997. *Hare Brain Tortoise Mind: Why Intelligence Increases When You Think Less*. London, Fourth Estate.

Cochran-Smith, M., and S. L. Lytle. 1999. The teacher research movement: A decade later. *Educational Researcher* 28: 15-25.

Dagher, Z. R., and S. BouJaoude. 1997. Scientific views and religious beliefs of college students: The case of biological evolution. *Journal of Research in Science Teaching* 34: 429–445.

D'Avanzo, C. 2003. Research on learning: Potential for improving college ecology teaching. *Frontiers in Ecology and the Environment* 1: 533–40.

D'Avanzo, C, B. W. Grant, D. Morris, S. Musante, J. Taylor, J. Riney, and D. Udovic. 2006. Design, evaluation, and research for TIEE, a peer-reviewed electronic teaching resource. *Frontiers in Ecology and the Environment* 4:189-195.

Dunning, D., K. Johnson, J. Ehrlinger, and J. Kruger. 2003. Why people fail to recognize their own incompetence. *Current Directions In Psychological Science* 12(3): 83-87.

Ebert-May, D. 2000. Workshop #2: Assessment of Student Learning: Implementing the Scholarship of Teaching and Research in Ecology (Biology) Education. Offered 4-5 August 2000, Annual Meeting of the Ecological Society of America, Snowbird, Utah.

Elliott, J. 1987. Educational theory, practical philosophy and action research. *British Journal of Educational Studies* 35(2): 149–169.

Elmesky, R., and K. Tobin. 2005. Expanding our understandings of urban science education by expanding the roles of students as researchers. *Journal of Research in Science Teaching* 42: 807-828.

Feynman, R. P. 1999. *The Pleasure of Finding Things Out: The Best Short Works of Richard P. Feynman*. Perseus Publishing. NY, NY. 270 pages.

Flavell, J.H. 1976. Metacognitive aspects of problem solving. In L. Resnick, ed., *The Nature of Intelligence*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Fox, R. 2001. Constructivism examined. *Oxford Review of Education* 27: 23-35.

Frankfurt, H. G. 2005. *On Bullshit.* Princeton, NJ, Princeton University Press. 80 p.

Freire, P. 1998. *Pedagogy of Freedom: Ethics, Democracy, and Civic Courage*. Rowman & Littlefield, Lanham, MD, 176 p.

Gardner, H. 2004. *The Unschooled Mind: How Children Think and How Schools Should Teach* (Tenth Anniversary Edition). Basic Books. 303 pages.

Grady, M. P. 1998. *Qualitative and Action Research: A Practitioner Handbook.* Phi Delta Kappa Educational Foundation. 55 p.

Griffith, J. A, and S. K. Brem. 2004. Teaching evolutionary biology: Pressures, stress, and coping. *Journal of Research in Science Teaching* 41:791–809.

Handelsman, J., D. Ebert-May, R. Beichner, P. Bruns, A. Chang, R. DeHaan, J. Gentile, S. Lauffer, J. Stewart, S. M. Tilghman, and W. B. Wood. 2004. Scientific teaching. *Science* 304: 521–522.

Hein, T. 1999. Using writing to confront student misconceptions in physics. *European Journal of Physics* 20: 137–141

Hokayem, H., and S. Boujaoude. 2008. College students' perceptions of the theory of evolution. *Journal of Research in Science Teaching* 45: 395–419.

Hollingsworth, S. 1997. *International Action Research: A Casebook for Educational Reform*. Washington, DC: Falmer Press.

Isaacson, R. M., and F. Fujita. 2006. Metacognitive knowledge monitoring and self-regulated learning: Academic success and reflections on learning. *Journal of the Scholarship of Teaching and Learning* 6(1): 39 - 55.

Jarvis, P. 1999. *The Practitioner-Researcher: Developing Theory from Practice.* San Francisco: Jossey-Bass.

Kemmis, S. 1988. Action research. In: J. P. Keeves (ed.) *Educational Research, Methodology and Measurement: An International Handbook*. Oxford, Pergamon Press.

King, P., and K. S. Kitchener. 1994. *Developing Reflective Judgment: Understanding and Promoting Intellectual Growth and Critical Thinking in Adolescents and Adults*. San Francisco: Jossey-Bass.

King, P. M., and K. S. Kitchener. 2004. Reflective judgment: Theory and research on the development of epistemic assumptions through adulthood. *Educational Psychologist* 39: 5-18.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self assessments. *Journal of Personality and Social Psychology 77:* 1121–1134.

Lederman, N. G., F. Abd-El-Khalick, R. L. Bell, and R. S. Schwartz. 2002. Views of Nature of Science Questionnaire: Toward valid and meaningful assessment of learner's conceptions of nature of science. *Journal of Research in Science Teaching* 39: 497–521.

Liem, T. 1981. *Invitations to Science Inquiry*. Lexington, MA: Ginn Press.

Mason, M. 2005. On bullshit. *Journal of Philosophy of Education* 39(4): 701-705.

McNiff, J. 2002. Action research for professional development. Accessed online 24 Nov 2007 at http://www.jeanmcniff.com/booklet1.html

McTaggart, Robin. 1997. *Participatory Action Research: International Contexts and Consequences*. State University of New York Press. 283 pages.

Merriam, S. B. 1998. *Qualitative Research and Case Study Applications in Education: Revised and Expanded From Case Study Research in Education*. Jossey-Bass Publishers, San Francisco, CA.

Miller, J. D., E. C. Scott, and S. Okamoto. 2006. Science Communication: Public acceptance of evolution. *Science* 313: 765–766.

Moore, John A., and Others. 1984. Science as a Way of Knowing I — Evolutionary Biology. Proceedings of the 1st Annual Session of Science as a Way of Knowing (Philadelphia, PA, 27-30 December 1983). *American Zoologist* 24(2): 1-115.

Morris, D. A., C. D'Avanzo, and B. W. Grant. 2007. The TIEE Research Practitioners Project: Faculty Investigating Active Teaching and Student Learning. *Teaching Issues and Experiments in Ecology*, Vol. 5: Research #7 [online]. http://tiee.ecoed.net/vol/v5/research/morris/abstract.html

Munson, B. H. 1994. Ecological misconceptions. *Journal of Environmental Education* 25: 30-34.

National Academy of Sciences and Institute of Medicine. 2008. *Science, Evolution, and Creationism*. Washington, DC: National Academies Press.

National Research Council. 1996. *National Science Education Standards*. Washington, DC: National Academies Press.

Passmore, C., and J. Stewart. 2002. A modeling approach to teaching evolutionary biology. *Journal of Research in Science Teaching* 39: 185–204.

Polanyi, M. 1958. *Personal Knowledge: Towards a Postcritical Philosophy*. London, Routledge & Kegan Paul.

Postman, N. 1969. Bullshit and the art of crap-detection. Paper presented at the National Council of Teachers of English Convention (Washington, DC).

Postman, Neil. 1979. Propaganda. *Et cetera* 36(2): 128-133.

Reason, P, and H. Bradbury. 2001. Introduction: Inquiry & participation in search of a world worthy of human aspiration. In Peter Reason and Hilary Bradbury (eds.) *Handbook of Action Research Participative Inquiry and Practice*. Sage Publications.

Reason P., and H. Bradbury. 2006. *Handbook of Action Research*. Sage Publications, London.

Rutledge, M., and M. Warden. 1999. The development and validation of the measure of acceptance of the theory of evolution instrument. *School Science and Mathematics* 99: 13–18.

Ryle, G. 1949. *The Concept of Mind*. London, Hutchinson.

Schmitt, F. F., and R. Lahroodi. 2008. The epistemic value of curiosity. *Educational Theory* 58(2): 125-148.

Schön, D. 1983. *The Reflective Practitioner*. New York: Basic Books.

Schraw, G. 1998. Promoting general metacognitive awareness. *Journal of Instructional Science* 26(1-2): 113-125.

Sinatra, G. M., S. A. Southerland, F. McConaughy, and J. W. Demastes. 2003. Intentions and beliefs in students' understanding and acceptance of biological evolution. *Journal of Research in Science Teaching* 40: 510–528.

Somekh, B. 2006. *Action Research: A Methodology for Change and Development Maidenhead*. Open University Press.

Smith, C. L, and L. Wenk. 2006. Relations among three aspects of first-year college students' epistemologies of science. *Journal of Research in Science Teaching* 43: 747–785.

Stenhouse, L. 1975. *An Introduction to Curriculum Research and Development*. London: Heinemann.

Stenhouse, L. 1985. A note on case study and educational practice. Pages 263-271 *in* R. Burgess (Ed) *Field Methods in the Study of Education*. London: Falmer.

Sternberg, R. J. 1998. Metacognition, abilities, and developing expertise: What makes an expert student? *Journal Instructional Science* 26(1-2): 127-140.

Stringer, E. 2008. *Action Research in Education*, (2nd edition). Pearson Prentice Hall, NY, NY. 242 pp.

Thomas, G. 1998. The myth of rational research. *British Educational Research Journal* 24(2): 141-161.

Tobin, K., W. M. Roth, and A. Zimmermann. 2001. Learning to teach science in urban schools. *Journal of Research in Science Teaching* 38: 941-964.

Weick, K. E. 1998 . Improvisation as a mindset for organizational analysis. *Organization Science 9*(5): 543-555.

Valencia, R. R. 1997. *The Evolution of Deficit Thinking: Educational Thought and Practice*. Falmer Press, London, 270 pp.

Wandersee, J. H., J. A. Mintzes, and J. D. Novak. 1994. Research on alternative conceptions in science. In: *Handbook of Research on Science Teaching and Learning*, edited by D. L. Gabel. New York: Macmillan, p. 131–210.

Weiner, J. 1994. *The Beak of the Finch: A Story of Evolution in Our Time.* First Vintage Books, New York, NY.

Wills, G. 2004. The Day the Enlightenment Went Out. New York Times. 4 November 2004. http://www.nytimes.com/2004/11/04/opinion/04wills.html?scp=1&sq=enlightenment+went+out&st=nyt {accessed 16 March 2008}

Appendix of Supplementary Figures

Supplemental Figure S1. Bio161 lecture outline for 2000 to 2005 showing the weeks during which evolution and coevolution were prominent concepts.

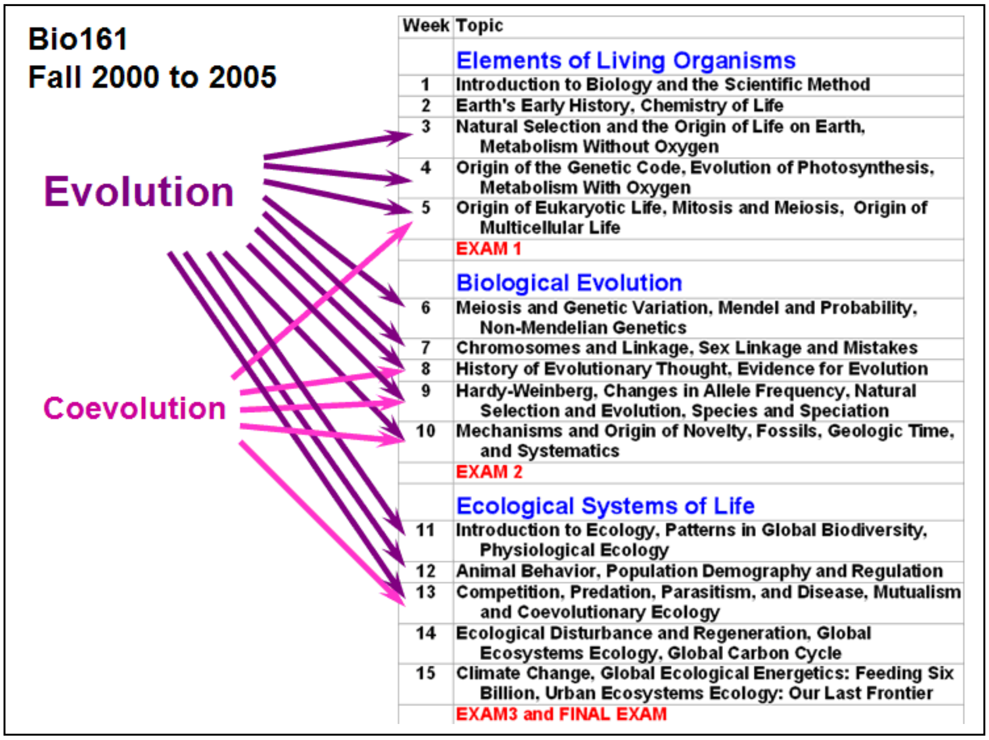


Figure S2. Students' individual scores on the pre-test, quizzes, and exams for (A) fall 2006, and (B) fall 2007 on the "definition of evolution" assessment question. The trajectory of the averages are in green. These data reveal a very important point about student performance on these kinds of assessments — individuals exhibit tremendous variation in performance from one assessment event to the next, and the vast majority of students show paths that bear little resemblance to the class averages. Most people are all over the place, ostensibly winking in and out of understanding. Learning is a highly non-linear process.