JOURNAL OF RESEARCH IN SCIENCE TEACHING VOL. 46, NO. 3, PP. 333–356 (2009)

Written Extended-Response Questions as Classroom Assessment Tools for Meaningful Understanding of Evolutionary Theory

Martina Nieswandt,[1] Katherine Bellomo[2]

[1]Illinois Institute of Technology, Mathematics & Science Education Department, 3424 S. State Street, Room 4007 South Bldg., Chicago, Illinois 60616
[2]Ontario Institute for Studies in Education of the University of Toronto, Department of Curriculum, Teaching and Learning, Toronto, ON, Canada

Received 17 September 2005; Accepted 9 May 2008

Abstract: This qualitative study analyzed grade 12 biology students' answers to written extended-response questions that describe hypothetical scenarios of animals' evolution. We investigated whether these type of questions are suitable for students ($n = 24$) to express a meaningful understanding of evolutionary theory. Meaningful understanding is comprised of factual, procedural (rules, algorithms), schematic (''knowing why''), and strategic knowledge (when, where and how to apply knowledge). Evolutionary theory as a multi-level concept includes concepts on three different levels (descriptive, hypothetical, and theoretical). Students' answers are examined as to whether they reflect the meaningful linking of all concepts through appropriate use of scientific language. Results showed that students (a) predominantly linked descriptive concepts and, although expected, (b) demonstrated only some cross-concept-level links (theoretical–descriptive), (c) exhibited even fewer multi-concept-level links (theoretical–descriptive–hypothetical), and (d) avoided the linking of hypothetical concepts with theoretical ones. All these results showed the lack of explanations and reasoning (absence of schematic and strategic knowledge) and knowledge of how to link concepts about evolutionary theory meaningfully. The results indicate further that written extended-response questions are only partially suitable for demonstrating meaningful understanding. Implications for teaching of evolutionary theory are discussed. © 2008 Wiley Periodicals, Inc. J Res Sci Teach 46: 333–356, 2009
Keywords: biology; evolution; performance assessment

The aim of science education is to achieve a high degree of scientific literacy (Council of Ministers of Education, 1997; National Research Council, 1996). Vital to this goal is the development of meaningful understanding of scientific concepts. Such an understanding goes beyond rote memorization toward the ability to construct connections among various pieces of information and to explain everyday phenomena with current scientific knowledge. Research demonstrates that developing such an understanding is difficult and often not successful (see for example studies on conceptual change in Pfundt & Duit, 2004). Assessing a meaningful understanding is also a difficult endeavor as teachers often lack the knowledge, skills, and time to develop assessment strategies that allow students to demonstrate their meaningful understanding rather than their memorization skills. Science education researchers have identified a variety of assessment strategies as having potential to assess meaningful understanding, such as for example, concept maps, which are visual, structured heuristics (Novak, 1990, 1998) that represent the interrelationship of concepts and have the potential to elicit students' meaningful understanding (Novak, Mintzes, & Wandersee, 2000). Other examples are structured interviews (Southerland, Smith, & Cummins, 2000) in which students are asked to construct detailed personal explanations of concepts, or tape-recorded verbal interactions (Hogan & Fisherkeller, 2000), which like the structured interviews assess students' reasoning skills. However, both researcher groups suggest that these assessment strategies are impractical for classroom use, are time consuming, require technical support and depend largely on students' oral skills. In contrast, commonly used

Correspondence to: M. Nieswandt; E-mail: mnieswan@iit.edu
DOI 10.1002/tea.20271
Published online 24 November 2008 in Wiley InterScience (www.interscience.wiley.com).

in classrooms are multiple-choice, short-answer and essay-answer questions. In this study we will focus on a special type of essay question that describes a hypothetical scenario and asks students, as they articulate an answer, to re-organize and synthesize various newly learned scientific concepts ‘‘into an integrated web of meaning’’ (Rivard & Straw, 2000, p. 568). We label them as *written extended-response questions* in order to distinguish them from other types of essay questions that for example, refer to simple tasks such as defining, listing or describing and which require the learner to focus on one concept in isolation. While teachers frequently use examples of written extended-response questions as summative assessment of a unit, science education researchers use them in studies focusing on identifying misconceptions (Bishop & Anderson, 1990) and in particular, in studies stressing students' conceptual change using a pre-post-test design (Jensen & Finley, 1996; Jiménez-Aleixandre, 1992; Settlage, 1994). In contrast to these studies, the purpose of our study is to determine whether *this type of assessment strategy* is suitable for grade 12 biology students to demonstrate their understanding of evolutionary theory by examining their answers to extended-response questions at the end of a unit on evolution.

Evolution a Unifying Theme and a Controversial School Topic

The topic of evolution was chosen because it is a central idea and unifying theme in biology, which is reflected in the prominent role that this topic is given in various national and international science standards and benchmarks (American Association for the Advancement of Science, 1989; 1993; Council of Ministers of Education, 1997; National Research Council, 1996), international curricula such as the Spanish and British national curriculum (CEC, 1994; QCA, 2007) and professional teacher associations (National Association of Biology Teachers, 1995; National Science Teacher Association, 1997). The theory of evolution is important for students studying biology at the high school level as it provides a foundation upon which other concepts can be built and a framework into which complex biological systems fit. Evolution is rich in opportunities to teach students aspects of the nature of science, of the role of evidence to support claims and of the process of the development of theory in the discipline of science/biology. As Dobzhansky (1973) wrote: ‘‘Nothing in biology makes sense except in light of evolution.’’ Thus, for students to become biologically literate and to pursue further studies in biology at the post-secondary level, it is essential that they develop a meaningful and scientifically accepted understanding of evolutionary theory. Nevertheless, regular polls (Goodstein, 2005) as well as educational research show that knowledge about evolutionary theory is limited and controversial among high school science students (Jensen & Finley, 1996) as well as across adult North Americans. A large number of people reject evolutionary theory (Smith, Siegel, & McInerney, 1995) and various religious and community groups oppose its teaching in high school classes (Goldsmith, 2000; Moore, 2000). Similar discussions are being reported from other countries such as the United Kingdom (Allgaier & Holliman, 2006), however the magnitude of this discussion in the U.S.A. is unique. Various studies pointed out that U.S. biology teachers spend only a few instructional hours, if any at all, on the topic as a result of community pressure or their own beliefs that evolutionary theory should only be taught in parallel to creationism or intelligent design. For example, 55% of teachers in Texas and 60% of teachers in Louisiana spend less than 5 instructional days on teaching evolution (Aguillard, 1999; Shankar & Skoog, 1993). Donnelly and Boone (2007) investigated 229 Indiana teachers' attitudes towards the state and the evolution standards and the emphasis given to teaching evolution. They found a strong correlation between teachers' attitudes toward the Indiana and evolution standards and teachers' use of these standards. For example, teachers' attitudes toward the evolution standards were a strong predictor of their evolution practice; they spend an average of 14.3 days teaching evolution and also regarded the evolution standards as helpful in justifying teaching evolution to parents and administration.

While earlier science education research focused on students' cognitive understanding of evolutionary theory (Anderson, Fisher, & Norman, 2002; Bishop & Anderson, 1990; Demastes, Settlage, & Good, 1995), the heated discussion of evolution versus creationism resulted in research examining ‘‘paracognitive factors’’ (Dagher & BouJaoude, 2005, p. 379) such as personal or religious beliefs, which influence student engagement with the content (Brem, Ranney, & Schindel, 2003; Fysh & Lucas, 1998; McKeachie, Lin, & Straver, 2002; Roth & Lucas, 1997; Sinatra, Southerland, McConaughy, & Demastes, 2003). Hokayem and BouJaoude (2008) explored 11 college students' beliefs about evolutionary theory after taking an undergraduate course on this topic. Students' positions about evolutionary theory ranged from complete acceptance to complete rejection despite all students' acceptance of the validity of scientific explanations in life situations and in nature. In addition, students' views differed from the course instructor's strictly scientific notions and explanations. Based on these results the authors stress the importance of identifying and understanding students' personal beliefs when teaching evolutionary theory. These results are in contrast to Ingram and Nelson's (2006) study. These authors investigated university students' acceptance or rejection of creation and evolution in upper-level evolution courses in a pre-post-course test design. They found that students had positive attitudes towards evolution which became more positive following the course and in particular, for students who were initially undecided about their attitudes toward evolution. Furthermore, independent of whether students accepted evolution or rejected it, their attitudes had little influence on achievement. Ingram and Nelson concluded that understanding evolution is more important than accepting it. Anderson (2007) raised the question of whether it is of concern what students come to *learn* about evolutionary theory or what they come to *believe* about evolutionary theory. He answered his question by making reference to the audience and the context of the discussion. In instructional contexts what matters is what students know about evolutionary theory, which can be tested with various assessment strategies. What students believe about evolutionary theory, and whether or not they believe it is true, cannot be tested and as Anderson stresses ''would be difficult to do with certainty even if the instructor wished to do so'' (p. 674). In a New York Times article Kenneth Miller was quoted saying to a student, ''Belief is never an issue . . . I just want you to know what it is and how it works'' (Dean, 2005 quoted in Anderson, 2007: 674). Our study focuses on students' content knowledge about evolutionary theory and more specifically, whether particular assessment tools (written extended-response questions) are suitable for students to express their meaningful understanding of evolutionary theory.

Theoretical Perspective

Meaningful Understanding

Meaningful understanding is a complex phenomenon. It is comprised of different types of knowledge and the ability to link these knowledge types. Meaningful understanding of biology in general and evolutionary theory specifically incorporates an understanding of single concepts such as population or of more complex concepts such as ecosystems (*declarative or factual knowledge*), which following particular rules and models, combines multiple individual concepts (e.g., producer, consumers, decomposers, food chain, energy flow) resulting in a new concept. The latter describes *procedural knowledge* (concepts, rules, algorithms), which together with declarative knowledge comprises parts of meaningful understanding. Furthermore, meaningful understanding includes *schematic knowledge* (guiding principles, schemes, mental models), ''knowing why,'' such as why ecosystems are an important factor for evolutionary change. Finally, meaningful understanding implies *strategic* or *conditional knowledge*, the understanding of when, where, and how to employ and apply knowledge and why it is important to do so (Paris, Cross, & Lipson, 1984; Shavelson, Ruiz-Primo, & Wiley, 2005). For example, physicists know when to apply Newton's first law given a problem dealing with force and motion whereas novices are attracted to the surface features of the problem (Chi, Feltovich, & Glaser, 1981). Physics students or as in our study, grade 12 biology students are not seen as experts but neither are they novices. Thus, they are likely to develop strategic knowledge but it is questionable to what extent (quantity and quality) and how well this knowledge is structured or organized. In the context of Ausubel's (1977) view of meaningful understanding we define high school students' strategic knowledge as the ability to apply the newly learned concepts to new situations or scenarios during class time, and to draw on the biological concepts more consistently and appropriately when confronted with scientific phenomena in the classroom and in everyday life. This includes for example, the ability to recognize new information as something different from one's current understanding, to identify inconsistencies and to construct explanations to reconcile knowledge conflicts or to seek connections among diverse pieces of information (Chan, Burtis, & Bereiter, 1997).

Levels of Scientific Concepts

Science concepts are often complex and abstract in nature. This can be a barrier for students to develop a meaningful understanding of such concepts. Lawson, Alkhoury, Benford, Clark, and Falconcer (2000) developed a classification system for scientific concepts, which we used for this study. They classified scientific concepts as descriptive, theoretical or hypothetical; what follows is a brief summary of this work. (See Lawson et al., 2000, for a fuller explanation and for the theoretical underpinnings that support these terms.) *Descriptive concepts* allow us to order and describe experience. They are expected to be the easiest to learn since meaning comes from experience and they should be learned first when new knowledge is tackled. Examples of descriptive concepts are ‘‘a chair’’ or ‘‘walking’’ or the concept that an animal is ‘‘nocturnal’’ as these can be observed and described. The second type of concepts according to Lawson and colleagues are *theoretical* ones. These concepts cannot be directly observed but are imagined and accepted based on an understanding of other information and theories. For example, ‘‘an atom’’ or ‘‘air pressure’’ cannot be observed no matter how long an observer waits and their meaning develops from indirect evidence and theories. The third type of concept is intermediate to descriptive and theoretical concepts and is called *hypothetical*. It cannot be observed but it is proposed that if one could wait for an extended period of time then it would be observed and described. Lawson and colleagues give examples of hypothetical concepts such as natural selection or species and we would add here the concept of ‘‘ecological succession.’’

Evolutionary Theory—A Multi-Level Concept

Using Lawson et al. (2000) three levels of concepts, evolutionary theory is comprised of various concepts on multiple levels. Concepts such as phenotype (fur color, lung capacity, and speed), food chains (predator–prey relationships) and environmental factors would be classified as descriptive concepts (D). Gene, genotype and mutation would be theoretical concepts (T) and as such the most difficult for students to comprehend. Concepts that are intermediate in terms of ease of comprehension as they are in part theoretical but could explain observations if it were possible to observe for an extended period of time are classified as hypothetical concepts (H). Examples are species, evolution, natural selection, and fossil.

A meaningful understanding of evolutionary theory pre-supposes that students can combine the discreet concepts given as examples above in a coherent explanation of a situation that is observed in nature. Specific to the complexity of evolutionary theory, many more concepts, as listed in Table 1, have to be combined to reflect a meaningful understanding of the current modern synthesis of Darwinian theory and Mendelian genetics. This synthesis is often labeled neo-Darwinian understanding. A neo-Darwinian understanding demands the combining of all three levels of concepts. The neo-Darwinian perspective includes mostly theoretical concepts (e.g., genes, mutations, gene pool), while the Darwinian theory has mostly descriptive concepts (e.g., variation, number of offspring, changes in the environment) and both include hypothetical concepts (e.g., species, natural selection, genetic drift).

Assessment of Meaningful Understanding

The focus on written tests in educational settings assumes that such tests reveal what students know or understand and therefore, that the test items are reliable and valid. Furthermore, such tests imply that the student refers to the same context and scientific discourse as the teacher (or researcher) who developed the

Table 1
Basic tenets comprising the multi-level concept “evolutionary theory”

Populations (D) have a great deal of variation (D) due to random genetic mutations (T)
Populations produce more offspring than the environment can support (D)
Within an ecosystem (D) there is competition for limited resources (D) and a differential survival among members of the population (D)
Changes in the environment result in more acute competition for limited resources (D)
Individuals who by chance have some form of variation (D if phenotype; T if genotype), which makes them more suited to the new conditions have a better survival rate and will leave more offspring (D)
Over a very long period of time the population will change (H) with respect to its appearance (phenotype; D) and its gene pool (genotype; T)
Over a very long period of time a new species may immerge, if the population is isolated from other populations who have not undergone this change (H)

questions. For example, the question ‘‘What is water?’’ can be answered from the perspective of a chemist who looks at the molecular structure of water, or the intermolecular forces between water molecules. A biologist might look at water in the context of the water cycle in nature, the physicist might approach water as an energy source, and the sociologist focuses on the meaning of water in different cultures. Although we can assume that students in general know in which context they are supposed to answer the test questions (biology, chemistry, physics, or social sciences), it is debatable whether we can assume that they are able to ‘‘talk science’’ (Lemke, 1990, 1995), or whether they are able to use, in oral or written language, the particular scientific discourse that is described in the test item. Research conducted by Johnstone and Cassels (1978) and Cassels and Johnstone (1979) looked at students’ language understanding of multiple-choice questions. This study demonstrated that even first language speakers had difficulties with a large body of non-scientific and scientific vocabulary. They either did not know the term or thought that it meant something different than the appropriate meaning. Another study (Clerk & Rutherford, 2000) examined a 20-item written multiple-choice test on Newtonian mechanics. A group of nine high school science students were interviewed after they had taken the multiple-choice test and asked to explain why they chose a particular answer. The results showed that students who had chosen a false answer representing a misconception often had language difficulties (e.g., misinterpreted the question text), instead of a lack of scientific understanding. They recommended giving language in science education more importance by looking at the extent of language interference in the diagnosis of misconceptions. Although our study is not looking at the presence of misconceptions, the findings of these studies highlight how important it is to phrase the test items in an unambiguous language and to use scientific terms that are known to the students.

Multiple-choice and short-answer questions would be appropriate to assess students’ declarative knowledge including the memorization of scientific terms, but these assessment strategies fall short when procedural, schematic and strategic knowledge or the ability to talk science is being assessed. While we can conceptually distinguish between the different types of knowledge, assessment strategies are not straightforward. Conducting a laboratory activity that compares hominid skulls to find out how facial features changed as hominids evolved, requires knowing, for example, what ‘‘hominid’’ means. An assessment strategy for such an activity, which focuses on students’ procedural and schematic knowledge, supposes that students know the term hominid. Thus, assessment strategies often do not align with discreet knowledge types and instead imply combining different types of knowledge. The hominid skull laboratory activity also implies students’ understanding of different types of concepts. While ‘‘hominid’’ falls into the category of descriptive concepts, evolutionary theory is a multi-level concept including descriptive, theoretical, and hypothetical concepts assuming students knowledge of all the single concepts that comprise the theory. As with knowledge types the different scientific concepts are often difficult to categorize and assess in distinct levels.

The test items upon which we will concentrate in this study are *written extended-response questions*. These questions ask students to consider a situation or scenario that is rich in detail, contextualized, and hypothetical in contrast to other forms of assessment such as multiple-choice, matching or true–false items. Students have to consider several concepts on different levels (descriptive, hypothetical, and theoretical) at once, to apply the concepts in a new situation and to synthesize them into a comprehensive answer that includes all types of knowledge. When formulating an answer to such a question, students not only must have developed an understanding of scientific concepts, they must also be able to select specific concepts that apply to the particular test item and combine them ‘‘into an integrated web of meaning’’ (Rivard & Straw, 2000, p. 568) through appropriate use of scientific language. Thus, the answers to written extended-response questions have the potential to and should illustrate students’ meaningful understanding and this, in the context of evolutionary theory as a multi-level concept, is the focus of our study.

Anderson and colleagues’ study (Anderson et al., 2002) assessing college students’ understanding of evolution suggests a multiple-choice test based on questions within realistic contexts. However the specific scenarios that were part of their questions provided copious and very detailed information that we feared would be distracting and overwhelming for high school students. For the participants in this study (high school students with first formal exposure to the topic of evolution) we are suggesting that it is more appropriate to use written answers to scenarios involving changes in characteristics of animals.

Methodology

Participants

The study was conducted at grade 12 level Biology in two high schools[1] in southern Ontario (Canada); one, Smith Academy, an urban all-boys' independent school,[2] the other, St. Claudia, a suburban Catholic school.[3] Both schools have heterogeneous populations; our participants consisted of students with European, Asian, African, and South American ancestry. Grade 12 biology classes at both schools were involved in a research project conducted by the first author, which explored how students' motivation and interest in science influences the understanding of science concepts. For the present study, data were used from students' answers to written extended-response questions at the end of the evolution unit in grade 12. Because not all students were present on the day of the data collection or had not given their consent to use their school work for the research project, data from a total of 24 students (7 male and 17 female students) were collected for this study. The majority of students were from St. Claudia (total of 21) and three students were from Smith Academy.

Design

All grade 12 biology students were given two written extended-response questions (see Table 2) at the end of a 25-hour unit on evolution. These questions were modified from studies concerning students' conceptions about evolution (Bishop & Anderson, 1990; Settlage, 1994) and from similar types of questions regularly used in assessing students' understanding of evolutionary theory in the classroom (Blake et al., 2002; DiGiuseppe et al., 2003; Johnson & Raven, 2001). In a previously conducted pilot study we used questions that were taken ''as is'' from studies concerning students' conceptions (Bishop & Anderson, 1990; Settlage, 1994). The analysis of students' answers showed rudimentary and incomplete answers, which often did not address major components described in the question. Furthermore, these questions lead students to answer solely on the descriptive level using concepts such as variation and number of offspring (Darwinian understanding). Based on the results from the pilot study we revised the questions used in the current study (a) to make them more comprehensible by addressing students directly instead of asking them to image an anonymous biologist working on an issue as the original question did, and (b) to provide students with obvious ''hints'' by the inclusion of key words that would lead them to reflect the multi-level dimension of evolution (neo-Darwinian understanding) in their answers. For example, in the ''Seal-question'' students were asked to incorporate additionally the concepts of mutation and genetic drift, both theoretical concepts and in the ''Polar Bear-question'' to include the theory of natural selection, a hypothetical concept.

Students answered the written extended-response questions during regular class time at the end of the unit on evolutionary theory and without the teacher being present but under the first author's supervision. No time limit was imposed and students completed the test on average in 25 minutes. Students were asked to read the questions carefully, to ask for clarifications if needed, and then to answer the questions using their knowledge from the unit on evolutionary theory using the space provided (half page) and the back side of the test page, if necessary.

Prior to administering the questions, the teachers confirmed that based on the content covered in class, students would be able to answer them, However, just because the teachers told us that they covered the content (implemented curriculum), we do not know whether students actually learned and what they learned.

Table 2
Written extended-response questions

Seal scenario
Seals remain underwater without breathing for nearly 45 minutes as they hunt for fish but their ancestors could stay underwater only a couple of minutes. Use the concepts of mutations and genetic drift to explain how the ability to not breath for long periods of time has evolved
Polar Bear scenario
White hair in animals (such as in polar bears) is due to the absence of a chemical pigment. Biologists think that ancestors of polar bears had dark fur. Use the theory of natural selection (e.g., directional selection) to explain how polar bears evolved to have white fur

Only appropriate assessment tools would provide information about whether students actually learned the content (achieved curriculum; Tarr, Chavez, Reys, & Reys, 2006). Although it was not the main focus of this study, the step-by-step analysis of our written extended-response questions should give us information about whether the implemented curriculum was also achieved.

In order to use test questions that would not change students' behavior during the test situation and thus not affect their performance but rather measure what we want to measure (Webb, Campbell, Schwartz, & Sechrest, 1966), our participants were made familiar with written extended-response questions similar to the ones that we used in this study, throughout the unit on evolution. Examples of such in-class questions are the following two scenarios: ''In zoos, lions and tigers can mate and produce fertile offspring. Evidence shows that lions and tigers lived in the same areas 1000 of years ago, and yet there is no evidence that they interbred. How can you explain this phenomenon?'' and ''Consider a cluster of three isolated islands. On the first island, most of the ground-dwelling animals are gray in color; the second island contains similar ground-dwelling species but they are mostly brown in color; and on the third island, the same types of animals appear but they are reddish in color. Suggest an explanation for these observations.'' Teachers either used these questions during the class (e.g., in whole class discussions or as group tasks), gave them as a homework assignment, or they were part of quizzes and the unit test. Based on random checks of the type of homework assignments and quizzes that teachers used throughout the unit, we presumed that students had sufficient interaction with written extended-response questions throughout the unit. Thus we assumed that the teacher created a classroom assessment environment (Stiggins, 1999) with which students were familiar. Although the teachers' assessment of students' answers were different from our multi-step analysis of the written extended-response questions, based on their own grading system and not on our deep analyses, teachers' oral and written feedback had the potential to make students aware of what they had already learned, what they were able to apply and of areas in which they still lacked the knowledge needed to explain the scenarios fully and appropriately.

Either directly after they had answered the question or a day later, students were shown their answers to the written extended-response items in individual interview sessions. They were asked whether they were satisfied with their answers or whether they wanted to add more information or otherwise change their answers. Students' answers were audio-taped and transcribed verbatim. Although students talked about their answers, the analysis of the transcripts showed that they neither changed nor added more information to their written answers. We therefore, discarded the interview data for this study.

In addition to students' answers to the written extended-response questions, we collected samples of students' work throughout the unit in order to assess what they have learned. Although teachers had agreed on a few common assignments and students had agreed to allow us to have a copy of their graded assignments, by the end of the unit we did not have a critical mass of the same assignments to be able to link individual students' answers to the written extended-response questions. Therefore, we decided not to use students' sample of work as additional measures of students' understanding of evolutionary theory.

Classroom Context

The unit preceding the data collection focused on evolutionary theory as described in the official curriculum in Ontario for grade 12 biology (Ministry of Education, 2000). At the start of the research project none of the participating science teachers had previously taught this particular unit. This was not surprising. Ontario's curriculum had undergone a revision based on the reduction of 5 years of high school to 4 years, which resulted in rearrangement of topics and the integration of new units into the grade 12 biology course. Prior to starting the research project, the participating teachers familiarized themselves with the new curriculum on evolutionary theory and then developed, in collaboration with the first author, an approach and a particular order of topics for the unit (see Table 3). This approach was adapted from a U.S. National Academy of Science resource on evolution and the nature of science (National Academy of Science, 1998). The unit was preceded by a unit on genetics and began with a historical overview followed by examination of various forms of evidence for evolution, exploration of the nature of science in order to provide a context for mechanisms of evolution, and finally, discussion of some contemporary examples such as population growth and the construction of phylogenetic trees using DNA evidence.

Table 3
Overview of topics and description of objectives of unit on evolutionary theory

Topic	Description/Objective
Theory of biological evolution including historical perspectives	Develop an understanding of the history and nature of science, e.g., show students that scientific explanations must be open to new ideas and criticisms Introduce students to the individuals who attempted to explain biological evolution, e.g., Jean Lamarck, Charles Darwin, Alfred Wallace
Evidence for evolution	Observe, describe, and evaluate features showing common ancestry from, e.g., fossils, anatomy, embryology, and biochemistry
Nature of science as it relates to evolution	Read, critically analyze and précis articles in order to: Correctly communicate scientific methodology and interpretations Think critically and logically when comparing evidence presented and explanations given Develop a questioning mind when reading journals, magazines, newspaper articles, etc. Compare and contrast opposing views Recognize and examine alternative explanations
Mechanisms of evolution	Explore possible mechanisms for evolution such as genetic variation and natural selection Explore the consequences of interactions within ecosystems (e.g., differential survival and reproduction) and that population dynamics are constrained by finite resources through predator–prey simulation
Contemporary examples:	Enhance students' mathematical skills as they investigate population growth of organisms
Population growth	Discuss overpopulation and the implications of current human population growth on natural ecosystems
Common descent	Investigate the evolutionary relationship between apes and humans through DNA sequences

As the teachers implemented the unit they used the following agreed-upon teaching strategies, which emphasized guided inquiry-based, student-centered activities:

- *Small group discussions on a variety of topics*: For example, students analyzed and critiqued the theoretical basis for theory building in evolutionary biology; they proposed lineages, and then given data from DNA sequences they created phylogenetic trees, which they compared to their proposals.
- *Small group activities to develop pictorial representations of concepts*: For example, students represented concepts such as genetic drift and the founder effect using diagrams of imaginary organism populations.
- *Individual written assignments*: For example, students wrote short papers on the evolution of diseases and on readings by Stephen J. Gould.
- *Individual textbook-based tasks*: For example, students made concept maps to describe their understanding of the mechanisms and sources of evidence for evolution.

Data Analysis

Drawing on the second author's expertise in biology and evolutionary biology in particular, we developed exemplar answers to both questions (see Table 4). We integrated all the expected concepts (e.g., mutation, genetic drift or directional selection) in combination with concepts necessary to reflect a meaningful understanding of the specific evolution of seals and polar bears. The exemplar answers are written in a way that we would expect student to be able to answer after 2 months of instruction on the topic of evolution.

We both independently analyzed each student's response qualitatively using a six-step procedure that we developed for this study. An example of this analysis is shown in Tables 5 and 6:

Table 4
Exemplary scientific answers to extended-response questions

Seal question
Ancestors of the seals could stay underwater for only a short time, but within this ancestral population there would be some individuals who were able to stay underwater for a much longer period of time. This ability would be due to a chance mutation resulting in a ‘‘stay underwater longer’’ gene. This mutation/change in DNA would go unnoticed unless it became an advantage in a changed environment due to competition for limited resources. For example, if there was a food shortage and the only available food was to be found very deep in the water, then seals that have the ‘‘stay-underwater-longer’’ gene/mutation would be able to exploit this new food source. Over time, it would be only these individuals who would be healthy enough to reproduce and at least some of their offspring would have the ‘‘stay underwater longer’’ gene. If a small population of these seals became isolated, by chance, from the rest of the group then, this example of genetic drift, would result in a population of seals that could stay underwater for long periods of time
Polar bear question
The ancestral bear population would have had some individuals with dark and some with white fur. (The white fur would be due to a mutation in the ancestral population that resulted in no pigment being produced.) These bears with white fur might be rare until there was a change in the environment, such as more snow for long periods of time. In the changed environment, white bears can hide or be camouflaged in the snow and the white fur is a selective advantage perhaps because the bears blend into the landscape and can hide from prey, or perhaps because the white fur is thicker and so the bears are warmer in the cold snow. This is an example of directional selection. Over time other bears would not be able to compete for resources (food, shelter) with the white ones and would die off. The white bears would survive to maturity and produce offspring that have the white fur gene (or trait)

(1) Students’ answers were divided into segments reflecting an individual idea. The segments ranged in length from a phrase to a complete sentence.
(2) In each segment a scientific term was highlighted if stated, or if a term was described the scientific term implied was named.
(3) Each term or description of term was then classified into one of the three levels of concepts (descriptive, hypothetical, and theoretical).
(4) Each segment in which the scientific term or description of term appeared was compared to the exemplary answers (see Table 5) and then classified as either matching part of the answer or not. If there was a partial match, then what was missing in the particular segment was noted.
(5) Students’ answers were judged holistically and classified as to whether the scientific terms or descriptions of terms were effectively linked between or among levels of concepts to reflect a comprehensive understanding. Concepts that were missing were noted. The linking of concepts on the same level (e.g., only descriptive concepts) is labeled as ‘‘one-concept-level links,’’ the linking of concepts on two different levels (e.g., theoretical and hypothetical concepts) as ‘‘cross-concept-level links’’ and the linking of concepts on all three levels (descriptive, hypothetical, and theoretical) as ‘‘multi-concept-level links.’’
(6) Patterns among students’ answers within each question as well as patterns across questions were highlighted.

After the individual analysis we compared our coding and discussed and resolved discrepancies. This discussion/resolution procedure was repeated until the inter-rater agreement reached 95%.

Results

To address our research question of whether written extended-response questions are an appropriate assessment strategy that allow students to demonstrate their meaningful understanding of evolutionary theory as a multi-level concept, we used the multi-step process to analyze students’ answers. We examined whether the answers reflect students’ ability to combine and link all necessary concepts through appropriate use of scientific language and regardless of the three conceptual levels [descriptive (D), hypothetical (H), theoretical (T)]. First, we describe the results of students’ answers to each written extended-response question by looking at each analysis’ steps and second, highlight patterns among students’ answers to each question and across both questions. Examples of students’ responses will be used to illustrate these results.

Table 5
Coding of exemplary scientific answer of Seal scenario

Answer	Scientific Terms/ Description of Terms	Concept Classification: Theoretical (T), Hypothetical (H), Descriptive (D)	Comparison to Sample Answer: Scientific or Non-Scientific	Holistic judgment: Ability to Combine Concepts Between/Among Concept Levels
"Ancestors of the seals could stay underwater for only a short time"	Ancestral phenotype	D	Scientific (S)	
"[B]ut within this ancestral population there would be some individuals who were able to stay underwater for a much longer period of time"	Variation	D	S	
"This ability would be due to a chance mutation"	Mutation (random)	T	S	
"[R]esulting in a "stay underwater longer" gene"	Gene	T	S	
"This mutation/change in DNA would go unnoticed unless it became an advantage in a changed environment"	Selective advantage	H	S	All concepts are combined among different concept levels
"[D]ue to competition for limited resources"	Competition (for limited resources)	H	S	
"For example, if there was a food shortage and the only available food was to be found very deep in the water"	Limited resource	D	S	
"[T]hen seals that have the "stay-underwater-longer" gene/ mutation would be able to exploit this new food source"	Selective advantage	H	S	
"Over time it would be only these individuals who would be healthy enough to reproduce"	Differential survival	H	S	
"[A]nd at least some of their offspring would have the "stay underwater longer" gene"	Passing on gene to next generation	D	S	
"If a small population of these seals became isolated, by chance, from the rest of the group,"	Geographic isolation	D	S	
"[T]hen this example of genetic drift"	Genetic drift	H	S	
"[W]ould result in a population of seals that could stay underwater for long periods of time"	New phenotype	D	S	

Table 6
Coding of exemplary scientific answer of Polar Bear scenario

Answer	Scientific Terms/ Description of Terms	Concept Classification: Theoretical (T), Hypothetical (H), Descriptive (D)	Comparison to Sample Answer: Scientific or Non-Scientific	Holistic Judgment: Ability to Combine Concepts Between/Among Concept Levels
"The ancestral bear population"	Ancestors	D	Scientific (S)	
"[W]ould have had some individuals with dark and some with white fur"	Existing variation	D	S	
"(The white fur would be due to a mutation in the ancestral population that resulted in no pigment being produced)"	Mutation	T	S	
"These bears with white fur might be rare until there was a change in the environment, such as more snow for long periods of time"	Change in the environment	D	S	All concepts are combined among different concept levels
"In the changed environment, white bears can hide or be camouflaged in the snow"	Competition advantage	D	S	
[A]nd the white fur is a selective advantage perhaps because the bears blend into the landscape and can hide from prey,"	Selective advantage	H	S	
"[O]r perhaps because the white fur is thicker and so the bears are warmer in the cold snow"	Trait	D	S	
"This is an example of directional selection"	Directional selection	H	S	
"Over time other bears would not be able to compete for resources (food, shelter) with the white ones and would die off"	Competition for limited resources	H	S	
"The white bears would survive to maturity and produce offspring"	Reproduction	D	S	
"[T]hat have the white fur gene (or trait)"	Trait or gene	D or T	S	

Table 7
Scientific terms or description of scientific terms as they appeared in students' answers (n = 24)

Terms	Mutation	Genetic Drift	Natural Selection	Survival	Selective Advantage	Gene/ Trait	Phenotype	Variation in Population	Reproduction/ Offspring	Environmental Change/ Limited Resources
Question										
Seal Scenario	15	6	—	13	0	9	0	14	9	3
Polar Bear Scenario	2	—	10	13	8	3/4	11	8	9	2

Students' answers may include one or more of the listed terms or description of terms.

Seal-Scenario

To express a meaningful understanding of the ''Seal-scenario'' we expected students to use many of the following scientific terms or a description of these terms reflecting the different conceptual levels: mutation (T), gene (T), environmental change (D), competition for limited resources (D), change over generations (H), reproduction or offspring (D), survival (D), genetic drift (H), or variation in the phenotype (D) (see Table 7 for a summary of terms in students' answers).

Out of 23 answers (one student out of a total of 24 did not answer this question) 15 included the word mutation, of which four also included the term genetic drift. Eight answers did not contain mutation but two of the eight included the term genetic drift. Only 3 students' answers mentioned either environmental change or limited resources, 13 used the term survival, 9 the description of the terms reproduction or offspring, 9 the phrase ''passing on genes,'' and 14 included the term variation. Based on these results we would expect students' answers to reflect a meaningful understanding of the evolution of seals. However, the fourth level of our analysis showed that although the terms were included, when compared to the sample answer they were incorporated meaninglessly, most often not even reflecting a factual knowledge based on memorization. For instance, the terms mutation, genetic drift or ''evolved'' were scattered into segments of the answer but mostly not defined, explained, or elaborated upon as seen in the following sample segments of student's responses:

> The genetic drift of breathing times accommodates the mutation of longer breathing times (male, #8).
> The seals' ancestors may not be able (sic) but the seals may have mutated over time to have that trait (female, #6).
> Well, I believe that seals had to have evolved from some other species of fish (female, #22).
> The ability to not breath for a long time has evolved (male, #11).

Even when two or three segments were read together and compared to the sample answer most answers lacked specific details with respect to the Seal scenario or lacked an explanation of the scientific terms as seen in the following examples:

> The ability to not breath for a long time has evolved because mutations allow populations to survive in an environment, if there is a change (male, #11).
> The seals probably accumulated many beneficial mutations, which made them gradually develop their breathing habits (female, #21).
> Survival of the fittest allows specific mutations to carry through and become common in the population (male, #10).

These answers reflect an absence of schematic and strategic knowledge, for example, knowing why mutations lead to variation and increased survival rate and how this knowledge is important for the evolution of the seals, or how genetic drift changes the gene frequency of a population.

When the answer is read holistically (fifth analysis level) we examined if students' answers reflected an ability to combine the different conceptual levels. As our sample answer reflects, many ''cross-concept-level'' links are necessary for a complete answer that addresses both the scenario in the question and also reflects a meaningful understanding of evolution. We specifically asked in the question that students use the concepts of mutation (T) and genetic drift (H) to explain seal evolution. In order to be able to express an understanding of mutation, students would need to link in a meaningful way the theoretical concept gene (or genotype) with the descriptive concept variation (or phenotype). An understanding of genetic drift would require linking of the theoretical concept of gene or allele frequency with the descriptive concept of population. Out of the 15 (of 23 total) answers in which the word mutation appeared, 10 also included the word genotype or phenotype. But only one answer out of these 10 reflected this kind of cross-concept-level link of the theoretical concept ‘‘gene’’ with variation (or phenotype), a descriptive concept:

> Mutations are also random changes that occur in the DNA of a population. If a beneficial mutation occurred that allowed the seal to remain under water longer, it would be more successful in hunting (female, #23).

It would also be necessary for meaningful answers to reflect the cross-concept-level link between genetic drift (H) and mutation (T). While four of the 10 answers contained both terms, genetic drift and mutation were linked in a meaningful way in only one student’s answer as shown:

> A chance mutation could have occurred in the ancestral species, which somehow affected the amount of oxygen it could retain (e.g., solubility in blood.)... As a result, these individuals contributed proportionally more offspring to the population and the trait become more common. If the population was small genetic drift could occur and change allele frequencies (female, #19).

Although this student made the link between genetic drift and mutation, the reason given for genetic drift is incomplete since a small population, although a condition for, is not the reason that genetic drift occurs. Despite the lack of a fully complete understanding, this answer is the only one that attempts a multi-concept-level link of population (D), mutation (T), and genetic drift (H).

Out of the 8 (of 23 total) answers in which the word mutation did not appear two incorporate the term genetic drift but in a manner that is meaningless and out of context as seen in this example:

> Eventually seals who can hold their breath become the majority population. Genetic drift favors this type of seal (male, #12).

To explain the seal scenario meaningfully some basic tenets of evolutionary theory are expected to be integrated into the answer. After a 2-month instructional unit on evolution, we would expect students to be able to include that populations have a great deal of variation (D) due to random genetic mutations (T) in their answers. From our analysis of students’ answers we found that only seven were able to link both concepts, however, mutation is used descriptively in that there is no mention of the genetic component which makes it a theoretical concept. The following answers are typical examples of this incomplete cross-concept-level link:

> For long swimming seals, the first mutation could have been something that resulted in a larger lung span or some other effect that would increase the seal swimming times (female, #14).
>
> A mutation may have occurred that caused an increase in lung capacity for the seals (female, #9).

Two answers stated both terms (variation in populations and mutations) but did not link them and 10 stated either one or the other but not both. In four answers the terms did not appear at all.

That populations change in terms of genotype and phenotype over time, is another basic tenet of evolutionary theory. None of the 23 answers included this tenet in a complete way, although one answer did combine populations changing with respect to phenotype.

Three students’ answers included another tenet, that a change in the environment (D) and competition for limited resources (D) are necessary for evolution, but only one of the three answers linked them in a one-concept-level link:

> The ancestors of the present seals may have had an abundance of fish, thus they only needed a few minutes of breathing under water. Due to migration of fish or fish loss selective pressures on present day seals may include to breathe longer under water (female, #13).

This answer demonstrates that the student is aware of the significance of limited resources as a precondition for evolution but gives no mechanism for change and rather implies that the change will occur when it is needed for survival.

Another tenet that is important for an understanding of evolutionary theory states that individuals who by chance have some form of variation (D on the phenotype level and T on the genotype level), which makes them more suited to new conditions (D) have a better survival rate (D) and will leave more offspring (D) and pass on their genes (T). None of the students' answers reflected this cross-concept-level link in its entirety. Yet out of the 23 answers, 10 linked effectively two of the following concepts: survival (D)—offspring (D) (four answers in a one-concept-level link), and survival (D)—passing on genes (T) (six answers in a cross-concept-level link). An additional three answers linked all three concepts, however each answer lacked either explanations of the concepts or the explanation was non-scientific:

> The ability to not breath for a long period of time may have happen (sic) over a long time. The seals' ancestors may not be able but the seals may have mutated over time to have that trait. There could have been ancestors, which could stay under (water) longer then the rest and they were the ones that were stronger and passed they genes on to offspring (female, #6).
>
> Seals that could stay under water for longer periods of time got the meals more often. Thus they were able to survive and breed more often than normal seals. This caused the allele frequency for that mutation to increase leading to a dominance in longer breathing time alleles (male, #8).
>
> A chance mutation could've occurred in the ancestral species, which somehow affected the amount of oxygen (e.g., solubility in blood). Individuals with this variation would be selected for because this undoubtedly increased hunting success and thus survival. As a result, these individuals contributed proportionally more offspring to the population and the trait became more common (female, #19).

From the remaining 10 (out of 23) answers, six included survival on the level of individual seals and not as a selective advantage for future generations, and 4 answers made no mention of any of the terms survival, offspring or passing on genes.

In summary, the complex six-step analysis of the Seal-question demonstrated that students' answers state key terms necessary to explain the seal evolution but these terms are often not explained, elaborated upon or put into the context of the particular scenario of the question. While this suggests a lack of meaningful understanding, the results of the holistic text analysis with respect to combining and linking concepts of the same and of different levels reveal that in the vast majority of answers one-concept-level linking, cross-concept-level linking as well as multi-concept-level linking are absent. However, the wording of our extended-response questions demanded students to combine and link concepts. Before we discuss why students seemed to lack this ability, we will give an overview of the results of the analysis of the second written extended-response question, the Polar Bear scenario.

Polar Bear Scenario

As in the seal scenario, we expected that to show a meaningful understanding of the "Polar Bear" scenario, students would use many of the following scientific terms or a description of these terms, reflecting the different conceptual levels: mutation (T), gene (T), environmental change (D), change over generations (H), selective advantage (H), survival (D), reproduction or offspring (D), competition for limited resources (D), or variation in the phenotype (D) (see Table 7 for a summary of terms in students' answers). Only 2 out of 24 answers showed the term mutation or a description of it; the term or description of survival is seen in 13 out of 24 answers, selective advantage appeared in 8, the term gene is seen in 3 and the term trait in 5 out of 24 answers. Nine out of 24 answers either included the terms or the description of the terms reproduction and offspring, but only 2 out of the 24 answers showed the term environmental change. The question also asked students to explain the evolution of white polar bears from brown ancestors using the theory of natural selection (H), thus we expected students to either include the term or a description of this term. Out of

24 answers 10 showed natural selection; 8 out of the 24 included variation in population, a notion that is part of the theory of natural selection; 11 described a phenotype; and 5 answers had nothing related to either. Based on the frequency of the appearance of terms in the answers, we would assume some meaningful understanding of the polar bear evolution. Yet, analyses of the segments in which these terms appear (fourth level of our analysis) revealed that most terms are not explained or elaborated upon and often are included in the text meaninglessly. One example is natural selection. Only 3 out of the 10 answers in which the term appeared provided an explanation as the following example demonstrates:

> Therefore, natural selection tells us that the population slowly eliminated the darker haired ancestors and the ones with light hair or white hair survived to reproduce and pass on this trait to their offspring (female, #23).

In students' answers that showed the term, it was dropped into the text without further explanation as seen in the following examples:

> The dominant colour became white because of natural selection (female, #4).
> Natural selection for white fur may have begun when there were dark furred animals, in the artic for example (female, #14).
> Theory of selection and directional selection may have caused the absence of pigmentation in present day polar bears (female, #13).

Another example is variation in a population. Out of the six answers in which this description appeared, five of them gave no reasons for the variation (mutation), either in the segment in which the description appeared, or in the combination of two or three segments as the first two examples show, while the third example is the only answer in which mutation is mentioned as the source of variation:

> Although the ancestors are believed to have dark fur, the fur probably varied in darkness (female, #15).
> The ancestors were thought to be dark furred, and over the years polar bears have lost their pigment (female, #1).
> This would have occurred because a mutation might have occurred, which caused the light color of the fur (female, #16).

Out of the 13 answers that included or described the term survival, four gave no reasons for survival, for example:

> As more and more white-furred polar bears are born, these are the ones that survived (female, #20).
> The white hair is useful, so that the selection will have more white hair (male, #2).
> This situation also involves survival of the fittest (male, #10).

When survival is explained then the reasons given referred to hunting, hiding from predators (although this concept of prey–predator is reversed), or being camouflaged:

> These polar bears wouldn't be able to catch food or hide from hunters, so they would die off. The polar bears with white fur would survive (female, #5).
> The animals with dark fur would be at greater risk with predators, causing white fur to prevail (female, #9).
> If a polar bear had dark fur in a snow covered environment they could not be able to camouflage themselves against human predators or the prey they need to survive (female, #23).

As in the seal question these answers reflected an absence of schematic knowledge, which indicates "knowing why." For example, the answers did not reflect the knowledge that in order for natural selection to occur there must be a pre-existing variation in a population and that in such a population a white haired bear might have an advantage over a brown one.

The examination of the answer holistically focused on the ability to link concepts on different levels. As seen in our sample answer (see Table 4), we expected students to cross-link concepts on different levels. We particularly asked students to use the concept of natural selection (H), which could result in a Darwinian perspective as a minimum answer or potentially in a neo-Darwinian perspective from some students who included genetic factors in their answer. Although each perspective was taught, we would anticipate students' answers to reflect different cross-concept-level links, depending on which perspective they were using. The neo-Darwinian answers would include genetic factors [mutations (T), or changes in gene pools (T)], while the Darwinian answers would include inheritance of traits (D). More specifically, cross-concept-level links such as mutation (T) with variation (D) might appear in a neo-Darwinian answer and selective advantage (H) with change in the environment (D) might appear in a Darwinian answer. Additionally, the answers are expected to reflect a multi-concept-level link such as linking directional selection (H) with competition for limited resources (D) and offspring with the favorable gene (T) in a neo-Darwinian answer but the concept of favorable gene might be substituted with favorable trait (D) in a Darwinian answer.

The neo-Darwinian cross-concept-level link ''mutation-variation'' did not appear in any of the 24 answers. Instead a variety of answers demonstrated a descriptive one-concept-level link such as ''variation-survival'' (three answers), or ''variation-offspring'' (two answers), or ''variation-survival-passing-on-trait'' (nine answers). The answers below show this one-concept-level link:

> Polar bears could have evolved to have white fur if one of its ancestors had white fur. That bear with the white fur could have passed that trait down to other bears and so on and so forth until there were a whole new species of white bears (male, #24).
>
> Early polar bears would stand out and be more susceptible to predators. White fur became more efficient as it allowed them to blend with the snow, thus causing the trait to filter through the population (male, #10).

These answers also illustrate, as we found in most students' answers, that a variation in the population of bears is implied but not explicitly expressed.

Four out of the 24 answers reflected a descriptive-theoretical cross-concept-link such as ''variation-survival-passing-on-genes/alleles,'' which is seen in the following examples:

> The ancestors to the Polar bear with lighter fur were less visible in the snow than those with dark fur. The dark ones were hunted and killed, not passing on their genes. The dominant colour became white because of natural selection (female, #4).
>
> Most likely, when the polar bears migrated to Artic lands, they needed more camouflage to survive predators and to hunt. Those polar bears, which were dark, didn't fit into the environment and thus, were easy prey/hungry. This led to a higher survival rate for white polar bears and more common appearance of white fur. White hair alleles were passed on more than dark hair, thus nature ''selected'' the white polar bears to become common (male, #8).

The last answer is also an example of a multi-concept-level link because it links the concept of selective advantage, which is hypothetical, with the theoretical concept of alleles and the descriptive concepts variation and survival. Two other students' answers showed this link.

Finally, we saw three answers with the multi-concept-level link mutation (T), variation (D), and selective advantage (H), of which two are shown below:

> This would have occurred because a mutation might have occurred, which caused the light color of the fur. The environment would have favoured and the mutation might have been beneficial. So, the dark-fur polar bears went extinct and the light-furred polar bears reproduced successfully (female, #16).
>
> The polar bear first of all usually lives in places where most of the year is covered in snow. Therefore, if a mutation occurred that a polar bear was white-furred, then it would be able to camouflage better from mankind, who was not always there. As more and more white-furred polar bears are born, these are the ones that survived and natural selection put pressure on the dark-furred polar bears, until these were extinct (female, #20).

The last answer states a snowy environment as the habitat of the polar bears' ancestors, which is then linked to a selective advantage. We found this descriptive-hypothetical concept link in altogether 9 out of 24 answers. However, stating the environment is not sufficient for an evolutionary understanding, we expected the cross-concept-level linking of *change of environment* (D) with selective advantage (H). We found only four answers, which linked these two concepts as shown in the example below:

> The ancestors of polar bears probably had dark fur because they didn't live in the same environment. But a change in their habitat or migration to a colder climate would cause the population to experience the pressures of directional selection where one extreme variation is favoured over the current population average. Thus, the trait for white fur was selected for and became more numerous as the bears were more successful (i.e., hunting success was the result of being able to camouflage in the presence of prey.) (female, #19).

The other 11 answers either make no link at all (9 answers), or link selective advantage with an unspecified environment that causes the change (2 answers).

In summary, the analysis of the polar bear scenario demonstrated that students integrated key terms in their answers but these terms were often not explained, elaborated upon or put into the context of the scenario. The results of the holistic analysis of the text confirmed this result. While a meaningful understanding would combine various types of cross-concept-level and types of multi-concept-level links (see sample answer in Table 4), none of the students' answers demonstrated this. Instead a majority of answers showed one descriptive one-concept-level link and a few students' answers had either one type of cross-concept-level link or one type of multi-concept-level link but never multiple types of links in combination.

Commonalities and Differences

The comparison of students' answers across both scenarios showed some patterns common to both answers. First, a majority of answers included the following descriptive concepts: individuals vary from one another; a trait or variation may result in survival; and producing offspring, mating or reproduction are important. However, most answers lacked reasons, which would demonstrate meaningful understanding of the seal or polar bear evolution. In the first instance, variation (D) should be attributed to random mutations (T) and also described as existing within a population. When describing a trait (D) as resulting in survival (D) then the conclusion that a favorable trait results in a selective advantage (H), which then leads to survival, needs to be included. The notion of reproduction, mating or producing offspring (D) is only complete when the answer includes that the favorable trait (D) or gene (T) is passed on to the next generation (H). This absence of reasons and the persistence of descriptive-level-concepts in isolation indicate that students lacked schematic knowledge. They cannot explain "why," which only would be given through linking concepts that are descriptive with concepts which are theoretical or hypothetical. This pattern also demonstrates the lack of linking a theoretical concept with a hypothetical concept.

A second common pattern found in students' answers to both scenarios is the absence of time as a factor in evolution. This factor could be expressed in a phrase such as "over generations," which would be a hypothetical concept, or would be implied in concepts such as genetic drift (H), yet it is seldom mentioned. Another pattern showed that almost none of the answers included the basic tenet that evolution happens at the level of the population's gene pool (T) but rather focused on "individuals changing." Both patterns underline that almost all students omit hypothetical and theoretical concepts that are key to a meaningful explanation of evolutionary theory.

Another commonality in most answers to both scenarios is the absence of the basic tenet that changes in the environment or environmental pressures (D) result in competition for limited resources (D). Many of the polar bear answers included a description of the environment or the habitat as a sort of back-drop within which evolution happens but most omitted the influence of a change in the environment. In contrast, only a few of the seal answers included any mention of the environment but several did explain the competition for limited resources (usually expressed as limited food or limited fish) as a factor or driving force in seal evolution but none of the answers explained this in the necessary context. Again, the "why" is missing in both sets of answers.

In order to show a meaningful understanding of evolution of seals or polar bears, multi-concept-level links (H-D-T) are necessary. However, we found only a few students' answers reflecting such links; the majority of answers lacked them. Instead we found that some answers link either theoretical with descriptive or hypothetical with descriptive concepts.

In summary, our results demonstrate the dominance of linking descriptive concepts, some cross-concept-level links such as T-D and H-T, few multi-concept-level links (T-D-H), and the avoidance of linking hypothetical concepts with theoretical ones. Furthermore, scientific terms that were often combined did not represent meaningful linking because reasons, explanations or the ''why'' needed to show schematic knowledge were missing.

Discussion

The written extended-response questions required students to link various theoretical concepts such as mutation or genotype with the hypothetical concepts genetic drift or natural selection or with descriptive concepts such as variation or change in the environment. For example, in the seal scenario students were expected to describe the process of seal evolution linking the concepts of genes (T) and mutations (T), population (D), survival (D) and passing on genes (T) or the concepts of gene (T), population (D), genetic drift (H), survival (D), offspring (D). Similar links were expected in answering the question of the evolution of the polar bears. The multi-step analysis demonstrated that many of the students' answers were adequate when linking or explaining descriptive concepts (D) such as the notion of survival and offspring. This result is in agreement with Lawson and colleagues' findings (2000) that students had the least problem with descriptive concepts. But in contrast to Lawson and colleagues' results (2000), our students had more difficulties in comprehending hypothetical than theoretical concepts. While Lawson and colleagues assessed students' demonstration of factual knowledge, our study went beyond this and assessed for students' demonstration of meaningful understanding, which includes all knowledge types and which is expressed through linking concepts on different levels. Our results showed that students not only had difficulties explaining theoretical and hypothetical concepts thus, demonstrating limited factual knowledge, but also for the majority, the linking of concepts on different levels was most challenging, which indicates a lack of procedural and schematic knowledge.

The integration of concepts such as mutation and genetic drift in the seal scenario and natural selection and directional selection in the polar bear scenario demanded specific cross-concept-level and multi-concept-level links instead of explanations of discreet concepts. While students were not able to fulfill these specific expectations, their answers did contain links, although most are scientifically meaningless or are solely on the descriptive level, which means that they only stated information that can be deducted from the question. Thus, students knew that they were supposed to link concepts in their answers.

Most students earned either an A or a B for this unit on evolutionary theory. The grade was based on various assessment strategies such as tests, quizzes, and written assignments. The tests and quizzes were a mix of multiple-choice questions, short answer questions, essay questions that included written extended-response questions, and problem-solving. The written assignments ranged from laboratory activity reports to position papers on issue-based topics (e.g., Tuberculosis: An evolving disease) to answers to probing questions related to the history of evolutionary theory (e.g., the controversy of Wallace, Darwin and Lamarck's positions). Overall, assessment strategies addressed all four knowledge levels: factual, procedural, schematic, and strategic knowledge, although not necessarily in one single assignment. Students' final grade in the course was consistent with their grade for this particular unit. According to the grades given by their classroom teacher, these are competent students who for the most part have done well in both the unit on evolutionary theory and in the grade 12 biology course. There seems to be a discrepancy between students' grades, which we assume reflect the existence of meaningful understanding, and students' answers to our questions. Our test items were in coherence with the general classroom assessment environment (Stiggins, 1999); students had the opportunity to practice similar written extended-response questions throughout the unit on evolutionary theory and thus, were familiar with this type of assessment tool. However, they seemed less able to fully and meaningfully transfer the knowledge they had gained from previous practice when answering the questions in our study. Research on transfer is in disagreement as to whether and when transfer is possible (for an overview see Barnett & Ceci, 2002). Based on Barnett's and Ceci's (2002) taxonomy for far and near transfer nine key dimensions (learned skill, performance change, memory demands, knowledge domain, physical context, temporal context, functional context, social context, and modality) should be considered in order to assess the likelihood for transfer. Although analyzing our study in the context of this taxonomy goes beyond its scope, it seems that the majority of dimensions did not change with the exception of the functional context. While students were given written extended-response questions throughout the unit on evolutionary theory either as homework or in tests that the teachers used as part of their unit assessment, for our study students answered the items voluntarily and without the pressure of being evaluated. Barnett and Ceci (2002) suggest that differences in functional contexts may "be accompanied by a difference in motivation, which could explain some of the performance difference found" (p. 623).

It seems too simple to conclude that our group of students did not have any understanding of evolutionary theory. The detailed multi-step analyses of their answers showed that they had some kind of understanding but the demonstrated understanding seemed to lack the strategic knowledge that would allow students to employ and apply their knowledge of evolutionary theory in the context of our questions. Furthermore, the revision of the questions did not have the expected effect. The integration of hints such as genetic drift, mutation or natural selection did not help them to use their knowledge of these concepts in explaining the seal or polar bear evolution. Instead, it seemed to cause students to regurgitate factual knowledge based on memorization as illustrated by the inclusion of disconnected scientific terms in their answers. With respect to our research question whether written extended-response questions are suitable for students to express their meaningful understanding of evolutionary theory, we have to conclude that these particular questions were suitable to express students' factual and procedural knowledge about evolution. But for our participants these types of questions seem to be less suitable to express schematic knowledge (knowing why) and in particular, strategic knowledge (knowing when, where, and how to employ knowledge) and thus, a meaningful understanding of evolutionary theory.

Furthermore, the analysis of students' answers unveiled that students lacked the ability to create a hypothetical history explaining the seal and polar bear evolution. This is particularly evident in the absence of the notion of environmental changes resulting in limited resources and/or selective advantage. Instead, students pre-assumed an environmental change by stating for example a snowy environment or they ignored this notion entirely. The written extended-response questions did not ask explicitly for the creation of such a hypothetical history but this expectation was implied and we assumed that students knew what they had to do. The questions assumed that students had the strategic knowledge of when and how to describe the evolutionary history of the seal and polar bear. It seems that referring in the questions to certain processes such as mutation and to different mechanisms such as genetic drift or directional selection that result in evolution, was not sufficient to guide our group of students towards a scientifically appropriate answer, which would demonstrate the different levels of knowledge that they may possess. Similar to Clerk and Rutherford's study (2000) our students had difficulties identifying the intent of the question that our sample answers reflected. Rather than using the wording as a guide and a reminder of key ideas or concepts that the answer should have, students focused their answer on specific details to the exclusion of the "big-picture" and to the exclusion of connecting specific details necessary to explain a complex theory. For example they gave answers that included the term mutation but provided little more in terms of evolutionary theory. They did not use the concept of mutation as a starting point that would logically take them to describe how mutations lead to variation in a population and how variation confers differential survival to those individuals who have a favorable variation in a particular environment.

Although the two questions (the seal and the polar bear scenarios) focused on concepts (mutation and genetic drift versus natural selection) that students were supposed to use as a springboard to explain the animals' evolution, most students ignored them and instead used other simpler concepts for example that reproduction and survival are important. This is particularly curious in the polar bear scenario in which the mechanism of natural selection is named but ignored even when it could explain how reproduction leads to differential survival resulting in changes in populations over long periods of time. This is another example of students' lack of strategic knowledge; they have omitted "when" to include important concepts and "how" to include them.

Conclusions and Implications

The results of our study demonstrated that answering written extended-response questions on evolutionary theory meaningfully is complicated. Although our participants had practice in answering questions similar to the ones used in our study during the unit on evolution, they did not answer them as expected. Students were able to make a few one-, cross- and multi-concept-level links but not generally at the level of sophistication needed to create answers that mirrored our sample answers. Our study is purely qualitative and therefore, we cannot make inferences regarding a general efficacy of written extended-response questions as assessment tools for meaningful understanding. However, we want to stress that these types of assessment tools are commonly used in science classrooms and thus, a practical application for teachers is to be aware of cognitive demands and the subtext in the questions necessary for answering them. To increase the quality and quantity of links between concepts on different levels that are necessary for demonstrating a meaningful understanding of evolution, students need

(a) to have an understanding of major concepts that comprise evolutionary theory;
(b) to know how to link these concepts meaningfully;

and in order to apply this understanding to the questions, they need;

(c) to possess cognitive strategies that allow them to decode the text for implied hints and directions.

The latter would include knowing *what* to do, *why* to do it, *how* to use given concepts as a springboard for the development of an appropriate answer, and *when* and *how* to employ additional information (e.g., a hypothetical history) that may not be provided in the question but are necessary for a meaningful explanation of evolution. Students need to learn these cognitive strategies in personally meaningful contexts (Mayer, 1998). Rather than learning strategies in isolation or context independent, we propose a step-by-step approach embedded in classroom tasks that interest and motivate students to practice these strategies. Teachers should guide their students through the reading and decoding of questions and also of deciphering each specific cognitive strategy that is being used in the different steps. For example, in both of our questions the word ''ancestors'' implies the importance of time as a factor in evolution, and the phrase ''use the concept/theory'' is not synonymous with ''include the term,'' instead it is a hint that the actual concepts should be integrated in the answer meaningfully. Students of course, need to know what ''meaningfully'' means. Thus, the teacher has to demonstrate the linking of individual concepts on different levels that comprise a more complex concept such as mutation or genetic drift. Doing this will make students aware of specific cognitive strategies and will help them to acquire the tools for linking and transferring knowledge, therefore developing strategic and schematic knowledge (Alexander, 1997; Pintrich, Marx, & Boyle, 1993). Pintrich (2003) suggests that when students use specific cognitive study strategies they are more involved with the content, which leads to more knowledge acquisition, which is an important pre-condition for developing a meaningful understanding of such a multi-level concept as evolution.

The teachers in our study used a curriculum that was mostly based on the materials recommended by the National Academy of Science (1998) and their instruction was based on guided-inquiry principles. This approach stressed major concepts of evolutionary theory; incorporated historically rich materials that have been identified as effective for developing an understanding of Darwinian evolution (Jensen & Finley, 1996); and gave students ample opportunities to instrumentally use the knowledge instead of merely passively perceiving it, for example by working on various contextual tasks such as created phylogenetic trees or discussing the evolution of diseases (Rudolph & Stewart, 1998). Despite its depth, the results of our study indicate that this approach needs improvement in order to fill the gap of students' knowledge in areas such as the ability to create a hypothetical history for animal evolution or to realize that evolution happens at the level of the population's gene pool. Passmore and Stewart (2002) stress the importance of working on data-rich hypothetical and real cases, which provide students with a rich context to discuss concepts such as variation and differential survival and in particular, to use those concepts to explain changes in populations over time. Catley (2006) argues for the integration of more macro-evolutionary topics such as issues of bioethics, human origins or cloning into the teaching of evolution. While such topics might increase students' motivation and thus their engagement in the learning of such concepts, we agree with Catley that further research is necessary on ‘‘students’ understanding of challenging macro-evolutionary concepts including species, deep time, natural hierarchy, character evolution, extinction, and evidence of monophyly (synapomorphy)’’ (p. 781).

Besides modifying the curriculum by stressing deeper understanding of specific concepts and their use in real and hypothetical cases, current research emphasizes considering students’ personal epistemological views about science and evolution (Ingram & Nelson, 2006) as well as their religious beliefs (Anderson, 2007) when teaching about evolution. Deniz, Donnelly, and Yilmaz’s (2008) finding of the importance of multivariate factors in explaining the variance in accepting evolutionary theory among Turkish pre-service teachers demonstrates the complexity of this topic. Evidently, more research is necessary to develop and assess teaching approaches that consider students’ worldviews and factors influencing their views in relation to the development of meaningful understanding of evolutionary theory, and the latter being assessed using written extended-response questions.

Notes

[1]Both schools were given pseudonyms in order to assure confidentiality.

[2]Independent schools in Ontario are governed by the Ministry of Education, and charge tuition fees.

[3]Catholic schools are governed by the Ministry of Education, are publicly funded and no tuition is charged.

References

Aguillard, D. (1999). Evolution education in Louisiana public schools: A decade following Edwards v. Aguillard. American Biology Teacher, 61, 182–188.

Alexander, P.A. (1997). Mapping the multidimensional nature of domain learning: The interplay of cognitive, motivational, and strategic forces. In: M.L. Maehr & P.R. Pintrich (Eds.), Advances in motivation and achievement, Vol. 10 (pp. 213–250). Greenwich, CT: JAI Press.

Allgaier, J., & Holliman, R. (2006). The emergence of the controversy around the theory of evolution and creationism in the UK newspaper reports. Curriculum Journal, 17(3), 263–279.

American Association for the Advancement of Science (AAAS). (1989). Science for all Americans. New York: Oxford University Press.

American Association for the Advancement of Science (AAAS). (1993). Benchmarks for science literacy: A Project 2061 report. New York: Oxford University Press.

Anderson, D.L., Fisher, K.M., & Norman, G.J. (2002). Development and evaluation of the conceptual inventory of natural selection. Journal of Research in Science Teaching, 39(10), 952–978.

Anderson, R.D. (2007). Teaching the theory of evolution in social, intellectual, and pedagogical context. Science Education, 91, 664–677.

Ausubel, D.P. (1977). The facilitation of meaningful verbal learning in the classroom. Educational Psychologist, 12(2), 162–178.

Barnett, S.M., & Ceci, S.J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. Psychological Bulletin, 128(4), 612–637.

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

Blake, L., Craven, M., Dobell, D., Flood, N., Jasper, G., Little, C., Mason, A., & Price, G. (2002). Biology 12. Toronto: McGraw-Hill Ryerson.

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

Cassels, J.R.T., & Johnstone, A.H. (1979). Understanding of non-technical words in science. London: The Chemical Society, Education Division.

Catley, K.M. (2006). Darwin’s missing link—A novel paradigm for evolution education. Science Education, 90, 767–783.

CEC (1994) (Aug.). Conselleria d’Educació i Ciéncia de la Generalitat Valenciana. Decret 174/1994 de 19 d’agost pel qual s’estableix el currículum del Batxillerat a la Comunitat Valenciana. Diari Oficial de la Generalitat Valenciana (no. 2356).

Chan, C., Burtis, J., & Bereiter, C. (1997). Knowledge building as a mediator of conflict in conceptual change. Cognition and Instruction, 15, 1–40.

Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorizing and representation of physics problems by experts and novices. Cognitive Science, 5, 121–152.

Clerk, D., & Rutherford, M. (2000). Language as a confounding variable in the diagnosis of misconceptions. International Journal of Science Education, 22(7), 703–717.

Council of Ministers of Education, Canada. (1997). Common framework of science learning outcomes K to 12. Toronto, ON, Canada: Council of Ministers of Education, Canada.

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

Dean, C. (2005, August 23). Scientists speak up on mix of God and science. New York Times, p. A1.

Demastes, S.S., Settlage, J., & Good, R. (1995). Students' conceptions of natural selection and its role in evolution: Cases of replication and comparison. Journal of Research in Science Teaching, 32(5), 535–550.

Deniz, H., Donnelly, L.A., & Yilmaz, I. (2008). Exploring factors related to acceptance of evolutionary theory among Turkish preservice biology teachers: Toward a more informative conceptual ecology for biology evolution. Journal of Research in Science Teaching, 45(4), 420–443.

DiGiuseppe, M., Vavitsa, A., Ritter, B., Fraser, D., Arora, A., & Lisser, B. (2003). Biology 12. Toronto: Nelson.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. The American Biology Teacher, 35, 125–129.

Donnelly, L.A., & Boone, W.J. (2007). Biology teachers' attitudes toward and use of Indiana's evolution standards. Journal of Research in Science Teaching, 44(2), 236–257.

Fysh, R., & Lucas, K. (1998). Religious beliefs in science classrooms. Research in Science Education, 28, 399–427.

Goldsmith, T.E. (2000). The evolution of wars. The Science Teacher, 67, 8.

Goodstein, L. (2005). Teaching of creationism is endorsed in new survey, The New York Times www.nytimes.com/2005/08/31/national/31religion.html (retrieved August 31, 2005).

Hogan, K., & Fisherkeller, J. (2000). Dialogue as data: Assessing students' scientific reasoning with interactive protocols. In: J.J. Mintzes, J.H. Wandersee, & J.D. Novak (Eds.), Assessing Science Understanding. A. Human Constructivist View. (pp. 95–127). San Diego, CA: Academic Press.

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

Ingram, E.L., & Nelson, C.E. (2006). Relationship between achievement and students' acceptance of evolution or creation in an upper-level evolution course. Journal of Research in Science Teaching, 43(1), 7–24.

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

Jiménez-Aleixandre, M.P. (1992). Thinking about theories or thinking with theories?: A classroom study with natural selection. International Journal of Science Education, 14(1), 51–61.

Johnson, G.B., & Raven, P.H. (2001). Biology. Principles & Explorations. Austin: Holt, Rinehart and Winston.

Johnstone, A.H., & Cassels, J.R.T. (1978). What's in a word? New Scientist, 78(103), 432–434.

Lawson, A.E., Alkhoury, S., Benford, R., Clark, B.R., & Falconcer, K.A. (2000). What kinds of scientific concepts exist? Concept construction and intellectual development in college biology. Journal of Research in Science Teaching, 37(9), 996–1018.

Lemke, J.L. (1990). Talking science. Language, learning and values. Norwood, NJ: Ablex.

Lemke, J.L. (1995). Intertextuality and text semantic. In: P.H. Fries & M. Greogory (Eds.), Discourse in society: Systematic functional perspectives. Meaning and choice in language (pp. 85–114). Norwood, NJ: Ablex.

Mayer, R.D. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. Instructional Science, 26, 49–63.

McKeachie, W.J., Lin, Y.-G., & Strayer, J. (2002). Creationist vs. evolutionary beliefs: Effects on learning biology. American Biology Teacher, 64(3), 189–192.

Ministry of Education. (2000). The Ontario Curriculum, Grades 11 and 12: Science. Toronto, ON, Canada: Ministry of Education.

Moore, R. (2000). The revival of creationism in the United States. Journal of Biological Education, 35, 17–21.

National Academy of Science. (1998). Teaching about evolution and the nature of science. Washington D.C.: National Academy Press.

National Association of Biology Teachers (NABT). (1995). NABT Statement on Teaching Evolution. http://www.nabt.org/sites/S1/index.php?p=65.

National Research Council. (1996). National Science Education Standards. Washington DC: National Academy Press.

Novak, J.D. (1990). Concept mapping: A useful tool for science education. Journal of Research in Science Teaching, 27(10), 937–949.

Novak, J.D. (1998). Learning, creating, and using knowledge: Concept maps as facilitating tools in schools and corporations. Mahwah, NJ: Lawrence Erlbaum Associates.

Novak, J.D., Mintzes, J.J., & Wandersee, J.H. (2000). Epilogue: On ways of assessing science understanding. In: J.J. Mintzes, J.H. Wandersee, & J.D. Novak (Eds.), Assessing Science Understanding. A. Human Constructivist View (pp. 355–374). San Diego, CA: Academic Press.

Paris, S.G., Cross, D.R., & Lipson, M.Y. (1984). Informed strategies for learning: A program to improve children's reading awareness and comprehension. Journal of Educational Psychology, 76(6), 1239–1252.

Passmore, C., & Stewart, J. (2002). A modelling approach to teaching evolutionary biology in high schools. Journal of Research in Science Teaching, 39(3), 185–204.

Pintrich, P.R. (2003). A. motivational science perspective on the role of student motivation in learning and teaching contexts. Journal of Educational Psychology, 95, 667–687.

Pintrich, P.R., Marx, R.W., & Boyle, R.A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. Review of Educational Research, 63, 167–199.

Pfundt, H., & Duit, R. (2004). Bibliography. Students' alternative frameworks and science education. Electronic edition. Kiel, Germany: Leibniz Institute for Science Education.

Qualifications and Curriculum Authority (QCA). (2007). Science. Programme of study for key stage 3 and attainment targets. London, UK.

Rivard, L.P., & Straw, S.B. (2000). The effect of talk and writing on learning science: An exploratory study. Science Education, 84, 566–593.

Roth, M.-W., & Lucas, K. (1997). From "truth" to "inverted reality": A discourse analysis of high school physics students' talk about scientific knowledge. Journal of Research in Science Teaching, 34, 145–179.

Rudolph, J.L., & Stewart, J. (1998). Evolution and the nature of science: On the historical discord and its implications for education. Journal of Research in Science Teaching, 35(10), 1069–1089.

Settlage, J. (1994). Conceptions of natural selection: A snapshot of the sense making process. Journal of Research in Science Teaching, 31(5), 449–457.

Shankar, G., & Skoog, G. (1993). Emphasis given evolution and creationism by Texas high school biology teachers. Science Education, 77, 221–233.

Shavelson, R.J., Ruiz-Primo, M.A., & Wiley, E.W. (2005). Windows into the mind. Higher Education, 49, 413–430.

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

Smith, M.U., Siegel, H., & McInerney, J.D. (1995). Foundational issues in evolution education. Science and Education, 4, 23–46.

Stiggins, R.J. (1999). Are you assessment literate? High School Magazine, 6(5), 20–23.

Southerland, S.A., Smith, M.U., & Cummins, C.L. (2000). "What do you mean by that?" Using structured interviews to assess science understanding. In: J.J. Mintzes, J.H. Wandersee, & J.D. Novak (Eds.), Assessing Science Understanding. A. Human Constructivist View (pp. 71–93). San Diego, CA: Academic Press.

Tarr, J.E., Chavez, O., Reys, R.E., & Reys, B.J. (2006). From the written to the enacted curriculum: The intermediary role of middle school mathematics teachers in shaping students' opportunity to learn. School Science and Mathematics, 106(4), 191–201.

Webb, E.J., Campbell, D.T., Schwartz, R.D., & Sechrest, L. (1966). Unobtrusive measures: Nonreactive research in the social sciences. Chicago: Rand-McNally.