

# Understanding Atmospheric Carbon Budgets: Teaching Students Conservation of Mass

Collin Reichert,<sup>1</sup> Cinzia Cervato,<sup>2,a</sup> Dale Niederhauser,<sup>3</sup> and Michael D. Larsen<sup>4</sup>

## ABSTRACT

In this paper we describe student use of a series of connected online problem-solving activities to remediate atmospheric carbon budget misconceptions held by undergraduate university students. In particular, activities were designed to address a common misconception about conservation of mass when students assume a simplistic, direct relationship between atmospheric CO<sub>2</sub> concentrations and carbon emissions. This particular misconception was challenged through an instructional intervention applying constructivist learning theory principles in an effort to prompt cognitive dissonance and induce conceptual change. This study is based on 1 y of data collected from a survey completed by introductory physical geology students ( $n = 176$ ), divided into a control group ( $n = 127$ ) and an experimental group ( $n = 49$ ). The students in the experimental group worked on an instructional intervention targeting identified misconceptions during a laboratory session. Both the control group and the experimental group were presented information targeting the same misconception through a traditional lecture. Students completing the instructional intervention demonstrated significant increases in learning and reductions of misconceptions relative to students in the control group. However, some aspects of the misconceptions seemed to persist. © 2015 National Association of Geoscience Teachers. [DOI: 10.5408/14-055.1]

**Key words:** Introductory physical geology, budgets, atmospheric carbon dioxide

## INTRODUCTION

Recent research on undergraduate and graduate student ideas regarding budgets (or stock-flow systems) have demonstrated a generally poor understanding of budgets in general and atmospheric carbon budgets in particular (Cronin and Gonzales, 2007, 2009; Sterman and Sweeney, 2007; Sweeney and Sterman, 2007; Sterman, 2008). These misunderstandings lead many students to think that stock levels are controlled solely by the inflow to a system, especially when the inflow–outflow rates are presented to students graphically (this phenomenon is referred to as a “pattern matching” misconception). Thus, many undergraduate students wrongly believe that simply stabilizing carbon dioxide (CO<sub>2</sub>) emissions at their current levels would stop the increase of atmospheric CO<sub>2</sub>. These misunderstandings of budgets persist even among highly educated graduate students at prestigious universities like the Massachusetts Institute of Technology (e.g., Sterman and Sweeney, 2007).

Understanding atmospheric carbon budgets is important for the public if people are to understand climate change and make informed decisions about supporting or rejecting national policies that address these issues. To that end, researchers and educators have sought effective ways to teach scientific climate change principles to postsecondary

students (cf. September 2014 issue of the *Journal of Geoscience Education*). Many researchers have advocated the implementation of constructivist learning principles when teaching climate change (Meadows and Wiesenmayer, 1999; Huntoon and Ridky, 2002; Rebich and Gautier, 2005; Bardsley and Bardsley, 2007; Harrington, 2008; McCaffrey and Buhr, 2008; Moxnes and Sarsel, 2009; DeWaters et al., 2014), as well as the incorporation of andragogy (adult education) principles (Arndt and Laude, 2008; Schuster et al., 2008). In designing the treatment for the present research, we included these principles, as well as drawing on principles associated with cognitive flexibility theory and conceptual change theory.

## PEDAGOGICAL FOUNDATION OF THIS STUDY

Constructivist learning theory has been greatly influenced by the work of Jean Piaget. Piaget held that knowledge is generated by the creation of mental representations of the world, or schemas, that change over time based on an individual’s experiences (Piaget, 1963; Driver et al., 1994; Woolfolk, 2007). According to Piaget, human development is a meaning-making process that involves continuous attempts at equilibration, or testing the adequacy of existing schemas, in integrating new information that the individual experiences. New experiences can be integrated into existing schemas, a process called assimilation. Alternatively, if the experience cannot be adequately explained using existing schemas (and the experience is not dismissed for some reason), the person must go through a process of accommodation, in which new schemas are created or existing schemas are adapted or replaced to incorporate the new information (Piaget, 1963; Woolfolk, 2007).

Received 25 September 2014; revised 11 March 2015; accepted 16 June 2015; published online 14 September 2015.

<sup>1</sup>Ames Middle School, 3915 Mortensen Rd., Ames, Iowa 50011, USA

<sup>2</sup>253 Science I, Department of Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa 50011, USA

<sup>3</sup>Department of Curriculum & Instruction/Literacy Studies, West Virginia University, 355 Oakland Dr., PO Box 6122, Morgantown, Virginia 26506, USA

<sup>4</sup>Department of Statistics and Biostatistics Center, George Washington University, 6110 Executive Boulevard, Suite 750, Rockville, Maryland 20852, USA

<sup>a</sup>Author to whom correspondence should be addressed. Electronic mail: cinzia@iastate.edu. Tel.: 515-294-7583. Fax: 515-294-6049

Conceptual change research suggests two conditions that promote conceptual change learning. Students must be provided with an appropriate conceptual pathway, which provides a logical chain of reasoning that takes them from their current erroneous or naive understandings to the desired conceptual understanding and that builds on accurate prior knowledge possessed by the learners; in addition, inaccurate conceptions should be directly confronted to prompt the cognitive effort necessary for accommodation (Posner et al., 1982; Scott et al., 1991). In our study, students had inaccurate schemas associating carbon emissions with atmospheric carbon concentrations (Reichert et al., 2014). These schemas were perturbed through a series of activities designed to create disequilibrium. Our prediction was that when students' inadequate schemas were challenged, the students would accommodate new information in ways that would promote the development of more scientifically accurate schemas. Thus, in this conceptual change model, providing a scientifically accurate conceptualization of the natural phenomenon is of vital importance. It allows learners to abandon inaccurate schemas while providing students with a better explanation to account for the phenomenon under examination.

Most educational research has focused on the way children learn, but some work has also described differences between effective instructional practices for children (pedagogy) and effective instructional practices for adults (andragogy). The following principles are associated with andragogy (Knowles et al., 2005):

- (1) Learners draw more heavily on prior experience.
- (2) Concepts being taught must have some relation to real life.
- (3) Learners prefer to learn through problem solving.

We incorporated these learning principles in several ways, including explicitly using scenarios and examples familiar to the students (e.g., filling a bathtub with water or examining cash flow through a bank account) to draw on prior experience, embedding activities in a scenario that involved running a small business and having students go to meteorological Web sites to find data to enhance relevance and provide a real-world context, and requiring students to work through a series of challenges to engage them in complex problem solving.

However, developing learning activities that draw on principles associated with conceptual change and andragogy is likely necessary but not sufficient for addressing atmospheric carbon budget concepts, which constitute a complex, and often confusing, instructional challenge. Students exposed to an introductory-level understanding of a particular content area often demonstrate an inability to transfer learned knowledge to new or more complex scenarios (Spiro et al., 1988, 1992). Due to the difficulty of generalization, cognitive flexibility theory advocates a case-based learning environment that provides students with novel and ill-structured problem-solving tasks. By "crisscrossing the conceptual landscape" (Spiro and Jehng, 1990), students can develop skills associated with advanced knowledge acquisition, namely, understanding the complexity of content and its applicability in other domains. Cognitive flexibility principles that were incorporated into

the instructional intervention include the following (Spiro et al., 1988):

- (1) Avoiding oversimplification of content by creating complex, ill-structured learning domains
- (2) Using multiple representations of content to encourage different applications of the concept
- (3) Using cases to teach the concept (avoiding abstract representations)
- (4) Making multiple interconnections between cases that exemplify the content
- (5) Providing opportunities for the learner to construct knowledge rather than relying on transmission of knowledge by the instructor

In this paper, we build on the body of data that documented the extent of students' budget misconceptions described in Reichert et al. (2014). Here, we describe our instructional intervention and our efforts to help students overcome their misconceptions regarding stock-flow systems. We also document the persistence of students' misconceptions as they learn this difficult concept in an introductory geoscience setting.

## METHOD

In this study, we compared changes in student understanding of atmospheric carbon budgets when students were taught through lecture alone with those of students who were taught through lecture with an accompanying 2-h instructional laboratory experience. Students in both groups answered budget questions on an initial pretest, participated in instructional activities, and responded to the questions again on the course's final exam. The study was reviewed by the institutional review board and deemed exempt.

## PARTICIPANTS

Study participants included 176 students enrolled in an introductory physical geology course at a large midwestern university in fall 2009. All students involved in the study were enrolled in the liberal arts and sciences (LAS) college. Table I compares the experimental and control groups. Restriction to the LAS students in both experimental and control groups largely balances other demographic variables. The experimental group includes slightly more juniors and fewer freshmen than the control group. The difference, however, is not statistically significant (chi-square test,  $p = 0.32$ ).

Forty-nine students participated in the instructional intervention, and the remaining 127 students served as a control group. Students in the treatment group were enrolled in the introductory geology lab, as well as in the lecture course. The lab is a separate course chosen by students who need a science lab and is required for geology majors. Students enrolled in the lab but not in the lecture course during the same semester were not included in the treatment group. In essence, by enrolling in the lab students self-selected to be part of the treatment group and were allowed to opt out if they did not want their data included in analyses for this research. The results of the study of budget understanding for all students are described in Reichert et al. (2014).

**TABLE I: Demographic characteristics of control and experimental groups.**

|            | Control ( <i>n</i> = 127) | Experiment ( <i>n</i> = 49) |
|------------|---------------------------|-----------------------------|
| Gender     |                           |                             |
| Female     | 53%                       | 53%                         |
| Male       | 47%                       | 47%                         |
| Age        |                           |                             |
| Under 19   | 24%                       | 16%                         |
| 19–21      | 69%                       | 71%                         |
| 22–24      | 6%                        | 6%                          |
| Over 24    | 2%                        | 6%                          |
| Year       |                           |                             |
| Freshmen   | 25%                       | 18%                         |
| Sophomores | 48%                       | 47%                         |
| Juniors    | 16%                       | 27%                         |
| Seniors    | 12%                       | 8%                          |

## INSTRUMENTS

### Demographic Questionnaire

Students provided demographic information on gender, age, major, college, year in school, interest in science, concern for the environment, and any actions they had taken to protect the environment through a questionnaire.

### Pretest

Five questions (see Appendix A, available online at <http://dx.doi.org/10.5408/14-055s1>) were used to assess student knowledge of atmospheric carbon budgets. The five questions required students to (1) recognize emissions would have to drop below removal rates to decrease atmospheric carbon levels; (2) and (3) examine a text-based scenario of emissions and carbon removal rates to determine when decreasing, stable, and maximum atmospheric carbon levels would occur; and (4) and (5) examine a graph of emissions and carbon removal, to determine points at which maximum or minimum carbon levels would occur. Pretest questions were presented with the demographic questionnaire and scored manually.

### Posttest

Posttest questions included the five questions from the pretest. Posttest items were integrated into the final examination for the course. Final exam data were collected using bubble sheet response forms and scored using a Scantron reader.

Atmospheric carbon budget knowledge items that were used as pretest and posttest measures were examined for content and construct validity. Six items were initially developed for the pretest and posttest; however, one item was removed when we discovered that the question prompted the correct response. This item was not included on the posttest or used in any analysis.

Validity review team members included a professor in atmospheric sciences, an expert psychometrician who specializes in survey construction, and one geoscience graduate student who was not involved in the study. These subject matter experts evaluated content validity by exam-

ining how fully pretest and posttest items assessed atmospheric carbon budget knowledge and concluded that the range of items provided in our measure addressed all key atmospheric carbon budget concepts. To address construct validity, the same team reviewed pretest and posttest questions relative to current theories of budget-driven models that explain changes in atmospheric carbon levels. They concluded that correct responses were consistent with current theoretical budget-driven explanations of increases in atmospheric carbon, while distractor responses were not.

Another concern centers on whether knowledge or skills that are not directly related to the content under study influence whether participants can answer questions correctly. For example, Questions 4 and 5 might be measures of students' ability to read graphs rather than their understanding of atmospheric carbon budgets. While a basic understanding of how to read a graph is necessary to answer the question, participants are university students who have had experiences with reading graphs in this course. Furthermore, the graph in the present study includes only two lines, and understanding the relationship between the two lines is necessary if one is to arrive at the correct answer.

## INSTRUCTIONAL MATERIALS

The treatment was grounded in an instructional case that addressed key misunderstandings identified through the surveys. Thinkspace, an online problem-solving e-learning system, was used as the environment to develop and present the instructional intervention. A new version of the e-learning platform is being developed and tested in summer 2015. Potential users should contact the corresponding author for further details.

The intervention included a scenario in which students took on the role of a local snow-cone business owner in an effort to situate instructional activities in a real-life situation. Students completed four tasks designed to improve their understanding of budget problems in different contexts. A screencast of the intervention is available (<http://screencast.com/t/MTI4MDQzM>).

Part of the rationale for developing the instructional case through Thinkspace was to test the efficacy of a stand-alone remediation implemented in a large lecture course where significant instructor–student interaction is limited. In implementing the remediation, instructor interaction was limited to simply assisting with technical issues and encouraging students. Thus, any feedback provided to students was intended to come from the Thinkspace program. Four tasks were completed by students, each requiring an application of budget concepts to new scenarios. Feedback was provided in two settings: (1) through a simulation in Task 1, where students would see in real time the effect of changing inflows on overall stock levels, and (2) in Task 3, where students' prior pattern-matching misconception was specifically targeted. In this second source of feedback, the group responded to the feedback through a short-answer question requiring the explanation for the disagreement between what is expected when the misconception is applied and what is observed.

### Task 1: Water Tank Problem

In the first task, students were told that they needed to meet a health code requirement that utensils used to make



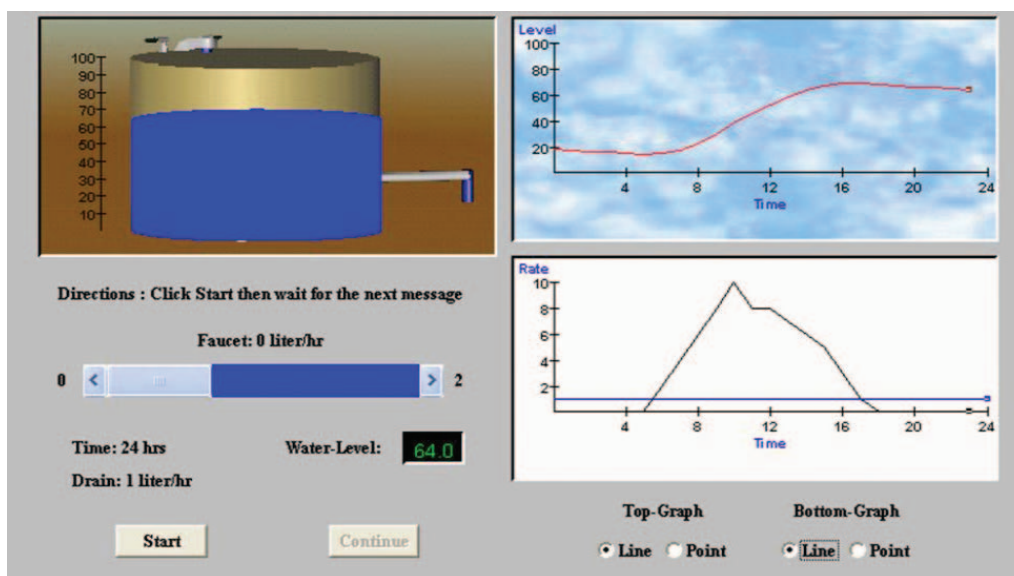


FIGURE 1: Simulation used by students in the instructional intervention. Students adjusted the inflow rate during business hours, and graphs displayed the data for a 24-h period. Students answered questions requiring interpretation of the data recorded on the graphs. From Thinkspace.

and serve the snow cones be stored in a container with continuously flowing potable water. This activity was designed to help students connect to their prior knowledge on using a faucet to control water levels in sinks and bathtubs. Students were informed that the business was equipped with a sink that had variable input rates on the faucet (0.0–10.0 L/h), and a constant output in the drain (1.0 L/h). Students were asked to maintain inflow rate such that water in the sink would reach a maximum level of 70 L. To help in accomplishing this task, students were provided with access to a simulation (Fig. 1) that allowed them to explore the effects of changing inflow rates on the water level in the tank. Students were allowed to control inflow rate during business hours (6 a.m. to 6 p.m.) by adjusting the inflow  $\pm 2.0$  L/h at the beginning of each hour. Graphs were used to provide a visual representation of inflow and outflow rates, and overall water levels, for a 24-h period. Students had had an opportunity to explore the simulation and were asked to interpret simulation data by answering a series of multiple-choice questions about the effects of changing inflow rates (e.g., At what point was the water level at its lowest? Under what conditions did the water level increase?).

### Task 2: Radiation Problem

For the second task, students were asked to anticipate inventory needs by determining yearly temperature variation and assuming that their snow-cone sales would be correlated with it (i.e. higher temperatures result in higher sales). Students read a brief explanation of the connection between temperature and radiation budgets and were asked to determine yearly radiation and temperature variation for their location. Students were provided with access to a series of radiation and temperature graphs, climate data from the National Climate Data Center, and an animation depicting the axis orientation of Earth as it orbits the sun. Again, when students were done exploring the materials, they were asked to answer a series of multiple-choice questions based on the relationship between radiation and temperature. These

questions were essentially identical to those asked in the first task except for the change in budget topic.

### Task 3: Bank Account Problem

The third task required students to consider cash flow through their snow-cone business with respect to the timing of planned renovations. Bank account records were presented in tables and graphs for the past year's income and expenses (Fig. 2). Graphs presented dollar amounts for deposits, withdrawals, and a running balance over the course of the year. Students were asked to consider why the point at which the account had maximum balance did not align with the point at which they were making maximum deposits. In this scenario students were required to apply what they had learned about budgets in a novel situation (as advocated by cognitive flexibility theory). Furthermore, the task promoted cognitive dissonance (as advocated by constructivist and conceptual change theory) by explicitly confronting student pattern-matching misconceptions that were identified in the pretest by having them account for a contradictory example.

### Task 4: Atmospheric Carbon Problem

In the fourth task, students projected future snow-cone sales for the 21st century, assuming correlation between snow-cone sales and projected temperature increases due to increases in atmospheric carbon, and whether this would justify expanding the business. This task included many resources and was the most intentionally ill structured of all tasks. The majority of resources were based on the Intergovernmental Panel on Climate Change's (IPCC's) 2007 report and projections. Students were asked to consider the effect of a 20%–40% reduction in emissions on atmospheric carbon levels and to construct a response explaining why, during a period of stable emissions in the 1990s, greenhouse gas concentrations continued to rise. Students also considered emission scenarios generated by the IPCC (2007) and implications of those emission

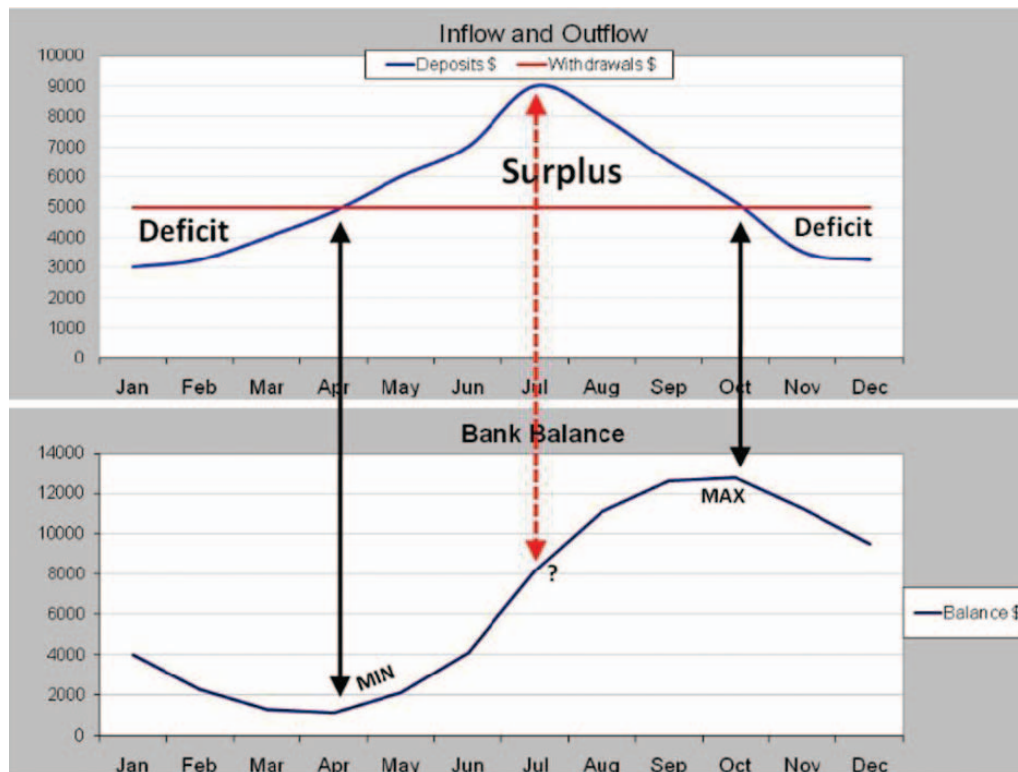


FIGURE 2: Bank account records presented to students depicting dollar amount on the vertical axis and month on the horizontal axis. Deposits are represented by the blue line in the top graph, and withdrawals are in red; overall balance is depicted in the bottom graph. Students were asked to account for why the maximum balance occurred in October but the maximum deposits were in July. From Thinkspace.

scenarios for temperature increases throughout the 21st century—requiring students to evaluate multiple emission scenarios in which carbon removal was projected to continue at its current value (see Fig. 3 for an example of resources students could access to help them solve this problem).

## PROCEDURE

Participants completed the demographic questionnaire and pretest within the first 2 weeks of the semester. Instruction was provided as part of two sections of a semester-long, three-credit introductory physical geology course taught by the same instructor. The course met three times per week for 50 min and covered a range of topics, including plate tectonics, geologic time, natural hazards, energy and mineral resources, and climate change. Students were encouraged to work in small groups and answer simple questions or solve problems. When the lecture class (which included students from both control and experimental groups) covered climate change, information and the graph used on the survey were explained to students during one of the lecture sessions. This lecture took place the week before Thanksgiving break, when students were asked to monitor and collect data on their carbon footprint during the 9 d of the break and report it for a homework assignment. This homework was followed by two more weeks of instruction and the final exam.

The lab consisted of three sections taught by graduate teaching assistants using a traditional lab manual. The lab manual did not include any assignment or instruction on

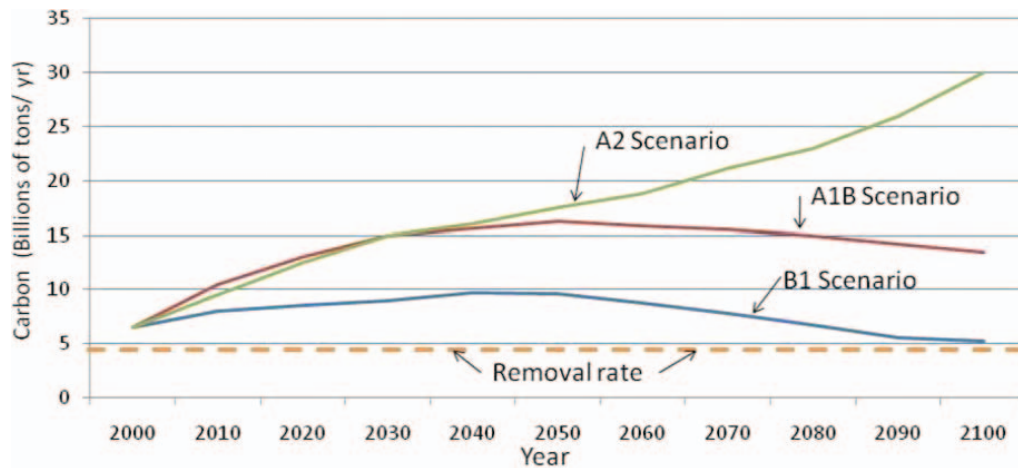
climate change, and all lab sections participated in the Thinkspace intervention. The intervention was completed by students during one lab period in a computer lab in groups of two to three students. The first author was present during all treatment lab sessions to provide students with technical assistance and to give occasional encouragement. At the end of the semester, and within two weeks of the experimental group's completion of the instructional intervention, students answered the posttest questions on the final exam. All participants answered the five viable budget questions from the pretest on their final exam (see Appendix A).

## RESULTS

Three phases of data analysis were conducted, and the results of each phase are reported in separate sections here. The first phase of analysis compared the experimental and control group equality of means on the pretest score and differences in demographic data. Next, growth in budget understanding from pretest to posttest was analyzed and compared across the control and experimental groups. Then, the experimental and control groups' performance on individual budget questions was analyzed and compared to identify specific learning gain differences.

### Group Comparisons

An independent sample *t*-test (two-tailed,  $p < 0.05$  criterion) was conducted using group assignment (experimental versus control) as the independent variable and pretest score mean (range = 0–5) as the dependent variable



This graph shows three of the scenarios developed by the IPCC for future greenhouse gas emissions in the 21<sup>st</sup> century. The removal of carbon from the atmosphere by oceans, plant-life and other factors currently absorbs about 50% (4 billion tons/ yr) of the current emissions (8 billion tons/ yr). Scientist think that the ability of natural sources to absorb carbon from the atmosphere will decrease over time, but here it is represented as constant.

FIGURE 3: Example of a resource that students could access when examining atmospheric carbon budgets. From IPCC (2007).

to determine whether there was a systematic difference in background knowledge between the two groups. The independent *t*-test was repeated using gender (male versus female) as the independent variable and pretest score mean as the dependent variable to determine whether there was a systematic difference in background knowledge between males and females in the sample. A series of *t*-tests were then conducted comparing experimental and control groups' interests and beliefs about the topic addressed in the study. Group assignment (experimental versus control) was used as the independent variable, and average interest in science (0–4 scale), environmental concern (0–4 scale), and actions taken to protect the environment (0–8 scale) served as dependent variables.

There was a significant difference in performance on pretest budget questions between groups ( $t(174) = 2.23, p < 0.05$ ). Students in the experimental group (mean = 1.55, SD = 1.08) scored higher than the control group (mean = 1.17, SD = 0.98) on the pretest measure. Pretest scores were also significantly different based on gender ( $t(174) = 2.91, p < 0.01$ ), with males (mean = 1.51, SD = 1.17) scoring higher than females (mean = 1.07, SD = 0.82). The distributions of males and females in the control and experimental groups were identical (47% male and 53% female in both groups).

Results of interest and belief differences between the experimental and the control groups are summarized in Table II. Not surprisingly, students who self-selected into the experimental conditions were more interested in science (mean = 2.86, SD = 1.22) than were students in the control group (mean = 2.20, SD = 1.04). The experimental group students also expressed greater concern for the environment (mean = 3.18, SD = 0.70) than did students in the control group (mean = 2.72, SD = 1.01). Likewise, experimental group students took more actions to protect the environ-

ment (mean = 6.43, SD = 1.10) than did students in the control group (mean = 5.60, SD = 1.45).

These results indicate clear systematic differences between the experimental and the control groups. Students assigned to the experimental group (enrolled in both the lecture and the laboratory geology course) possessed greater knowledge of budgets on the pretest than did students in the control group (enrolled only in the lecture portion of the geology course). Experimental group students also identified themselves as being more interested in science, more concerned about the environment, and more active in protecting the environment than did students in the control group. Though males tended to exhibit greater budget knowledge than did females on the pretest, the distribution of gender was identical for both groups. Thus, one can assume gender did not play a role in the differing budget knowledge between the experimental and the control groups.

### Growth in Budget Knowledge

The difference between posttest and pretest scores was computed for each student to assess their learning of the budget concepts. Overall, students performed better (mean gain = 0.64, SD gain = 1.13,  $t(175) = 7.52, p < 0.001$ ) on the posttest (mean = 1.92, SD = 1.10) than they did on the pretest (mean = 1.28, SD = 1.02), with students in both groups showing statistically significant improvement from pretest to posttest. In the control group, the average gain was 0.53 points (SD 1.11,  $t(126) = 5.35, p < 0.001$ ). In the experimental group, the average gain was 0.94 points (SD 1.15,  $t(48) = 5.74, p < 0.001$ ). The larger gain in scores for the experimental group relative to the control group was also statistically significant (difference in average gain = 0.41,  $t(174) = 2.18, p = 0.03$ ). This means students exposed to the



TABLE II: Demographic differences of control ( $n = 127$ ) and experimental ( $n = 49$ ) student groups.  $t$  values compared mean values between groups.

| Science Interest                               | Very Interested (4) | Somewhat Interested (3) | Neutral (2) | Not Very Interested (1) | Not at All Interested (0) |            | $t(174) = 3.59, p < 0.001$ |
|--|---------------------|-------------------------|-------------|-------------------------|---------------------------|------------|----------------------------|
| Control  | 10%                 | 31%                     | 31%         | 24%                     | 4%                        |            | Mean = 2.20                |
| Experimental                                   | 37%                 | 35%                     | 14%         | 6%                      | 8%                        |            | Mean = 2.86                |
| Concern for the Environment                    | Very Concerned (4)  | Somewhat Concerned (3)  | Neutral (2) | Not Very Concerned (1)  | Not at All Concerned (0)  |            | $t(174) = 2.92, p < 0.01$  |
| Control  | 21%                 | 47%                     | 17%         | 13%                     | 2%                        |            | Mean = 2.72                |
| Experimental                                   | 33%                 | 55%                     | 10%         | 2%                      | 0%                        |            | Mean = 3.18                |
| Total out of Eight Environmental Actions Taken | 8                   | 7                       | 6           | 5                       | 4                         | 3 or fewer | $t(174) = 3.63, p < 0.001$ |
| Control  | 1%                  | 31%                     | 27%         | 24%                     | 12%                       | 6%         | Mean = 5.60                |
| Experimental                                   | 6%                  | 59%                     | 16%         | 10%                     | 6%                        | 2%         | Mean = 6.43                |

instructional intervention in the experimental group learned more about budgets than did students in the control group. Thus, the instructional intervention appears to have been effective in helping students learn about budgets.

Linear models were used to adjust for the impact of covariate factors when estimating the impact of the experimental group on the gain in scores on the final exam (posttest) over the pretest. Table III presents results of fitting six models. The first model uses no covariate in addition to the experimental group. The second through fifth models use one covariate each in addition to the experimental group. The sixth model uses only the covariate “interest in science.” The table reports degrees of freedom and, in the last column, an Akaike’s information criterion (AIC) value for each model (Sakamoto *et al.*, 1986). AIC equals  $-2$  times the log likelihood for the model plus twice the number of parameters in the model. Small values are preferred because they occur when the model fits the data better with fewer parameters.

The estimate for the impact of the experimental group is consistent across the first four models: 0.41–0.43 and statistically significant. The covariates in Models 2–4 do not improve the model and are not statistically significant. In the fifth model, the impact of the interest-in-science covariate on the coefficient for experimental group is to decrease the coefficient so that it is not statistically significant. For further comparison, the sixth model reports the fit with interest in science as the only predictor. It seems that part of the impact of the experimental group is related to students with more interest in science choosing the lab. Models with the experimental group only, with interest in science only, and with both of these variables have the best AIC values, indicating that these models are preferable to models with the other covariates.

The Sobel test is a formal test of mediation (Sobel, 1982, 1986; Baron and Kenny, 1986) that was used to assess whether interest in science significantly mediated the impact of experimental group on the gain in test scores. In this application, although the inclusion of interest in science in the model reduced the estimated coefficient for the impact of the experiment, interest in science did not qualify as a mediating variable, because it was not a statistically significant predictor in Model 6. Furthermore, the Sobel test produced a nonsignificant test statistic of 1.54 ( $p = 0.12$ ). Overall, these results suggest that the instructional intervention appears to have been effective in helping students learn about budgets despite some apparent lessening of the effect due to a tendency for students interested in science to enroll in the experimental group.

### Question Type and Growth in Budget Knowledge

Student pretest and final examination posttest performances on budget items are presented in Fig. 4. All questions showed an increase in scores from the pretest to the final. On average, and out of a total score of 5, students in the control group scored 1.17 on the pretest and 1.70 on the final. In the experimental group, the averages were 1.55 on the pretest and 2.49 on the final. The statistical significance of the increase can be assessed with McNemar’s test for correlated proportions (Lachin, 2010). Question 1 (pretest 48%, final 86%,  $p < 0.001$ ) and Question 4 (pretest 6%, final 22%,  $p < 0.001$ ) showed statistically significant increases in the percentage of correct answers. Question 2

**TABLE III: Impact of covariates on the estimate of impact of the experiment on gain in the total score on five questions from pretest to final.**

| Model | df  | Covariate             | Experiment Impact |      |          |          | Covariate Impact |      |          |          | Model AIC |
|-------|-----|-----------------------|-------------------|------|----------|----------|------------------|------|----------|----------|-----------|
|       |     |                       | Est.              | SE   | <i>t</i> | <i>p</i> | Est.             | SE   | <i>t</i> | <i>p</i> |           |
| 1     | 174 | None                  | 0.41              | 0.19 | 2.18     | 0.03     | —                |      |          |          | 543.51    |
| 2     | 173 | Gender                | 0.41              | 0.19 | 2.18     | 0.03     | 0.03             | 0.17 | 0.17     | 0.86     | 545.48    |
| 3     | 173 | Environmental concern | 0.43              | 0.19 | 2.22     | 0.03     | −0.04            | 0.09 | −0.43    | 0.67     | 545.32    |
| 4     | 173 | Actions taken         | 0.41              | 0.20 | 2.08     | 0.04     | 0.00             | 0.06 | 0.06     | 0.95     | 545.51    |
| 5     | 173 | Science interest      | 0.32              | 0.19 | 1.67     | 0.10     | 0.13             | 0.08 | 1.70     | 0.09     | 542.58    |
| 6     | 174 | Only science interest | —                 |      |          |          | 0.17             | 0.07 | 2.21     | 0.03     | 543.40    |

<sup>1</sup>df = degrees of freedom; Est. = estimate of impact; SE = standard error.

(pretest 45%, final 52%,  $p = 0.18$ ), Question 3 (pretest 27%, final 30%,  $p = 0.58$ ), and Question 5 (pretest 2%, final 3%,  $p = 0.68$ ) did not show a statistically significant increase.

Graphs comparing performance on pretest and final budget questions by the experimental group are shown in Fig. 5. The experimental group is the solid line, and the control is the dotted line. The percentage correct increases for each question in the experimental group, but in the control group scores on Questions 3 and 5 are lower on the final than on the pretest. For Questions 1 and 2, the gain from pretest to final is about the same in the two groups. For Questions 3 and 4, the gain is larger in the experimental group than in the control group. The lower level of graphs presents the percent normalized gain for the two groups. Experimental is in blue, and control is in red. The normalized gain score is 100 times the gain in percentage correct divided by one minus the percentage correct on the pretest. Except for Question 2, the normalized gain score is larger for the experimental group (blue) than for the control group.

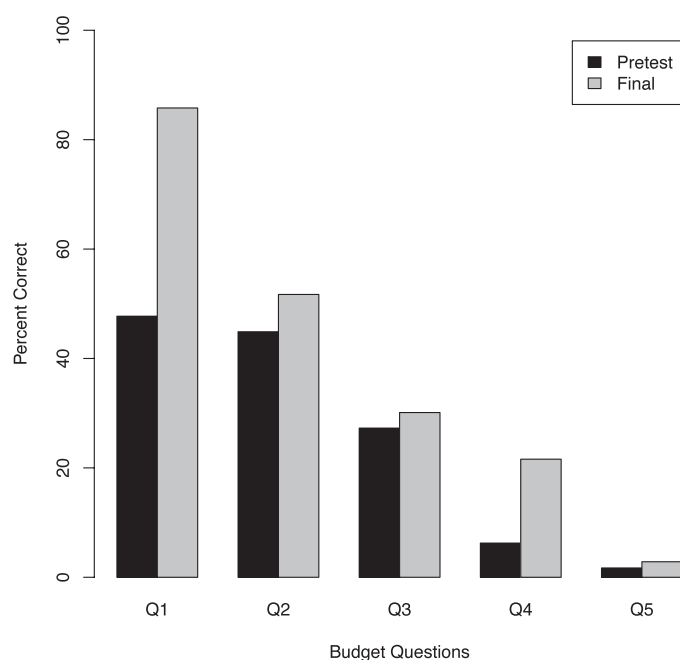
Proportional odds models were fit to assess whether differences are statistically significant. The outcome for each

student on each question is correct (1) or incorrect (0). The difference between a final score and a pretest score is 1 (improvement), 0 (no change), or −1 (worse; correct on pretest, incorrect on final). Table IV contains estimates of parameters for the first four questions. Question 5 had few changes in outcome between pretest and final. As a result, it is not worth making an inferential statement about the coefficient for experimental group status or covariates.

Two other methods were also examined for assessing whether the differences are statistically significant. The differences can be fit by a linear model with a group as predictors, but the residuals will not have a normal distribution. The other method used a logistic regression to predict improvement versus no improvement, where the negative category (scored as zero) includes no change and a worse result. Results were consistent with the conclusions of Table III: only Question 4 had a statistically significant gain related to the experimental treatment.

## DISCUSSION

The results of this study suggest that lecture instruction alone and additional time on a task with specifically designed instruction results in a better understanding of budget concepts, because both groups demonstrated significant growth from pretest to posttest. Additional time on a task with instructional intervention leads to significantly higher learning gains overall compared to lecture alone. When analyzed by individual question, the experimental group performed significantly better than the control group on just one of the five budget questions used to measure students' understanding. This question related to correctly identifying maximum stock levels based on the interpretation of an inflow–outflow graph. This was the concept most specifically addressed in the intervention, so one would expect that students having gone through the intervention would perform better on this task. However, the goal of the intervention was to have students generalize budget concepts to deeply understand them and gain the ability to retrieve this concept and apply it in appropriate situations. The fact that experimental group students did not apply their understanding significantly more than the control group on any of the other questions, particularly Question 5, is a concern, and it suggests that complete generalization of budget concepts did not occur. This means that students in both groups still significantly struggle in accurately understanding the graphical budget questions. Closer examination

**FIGURE 4: Percentage correct on pretest and final for five budget questions.**



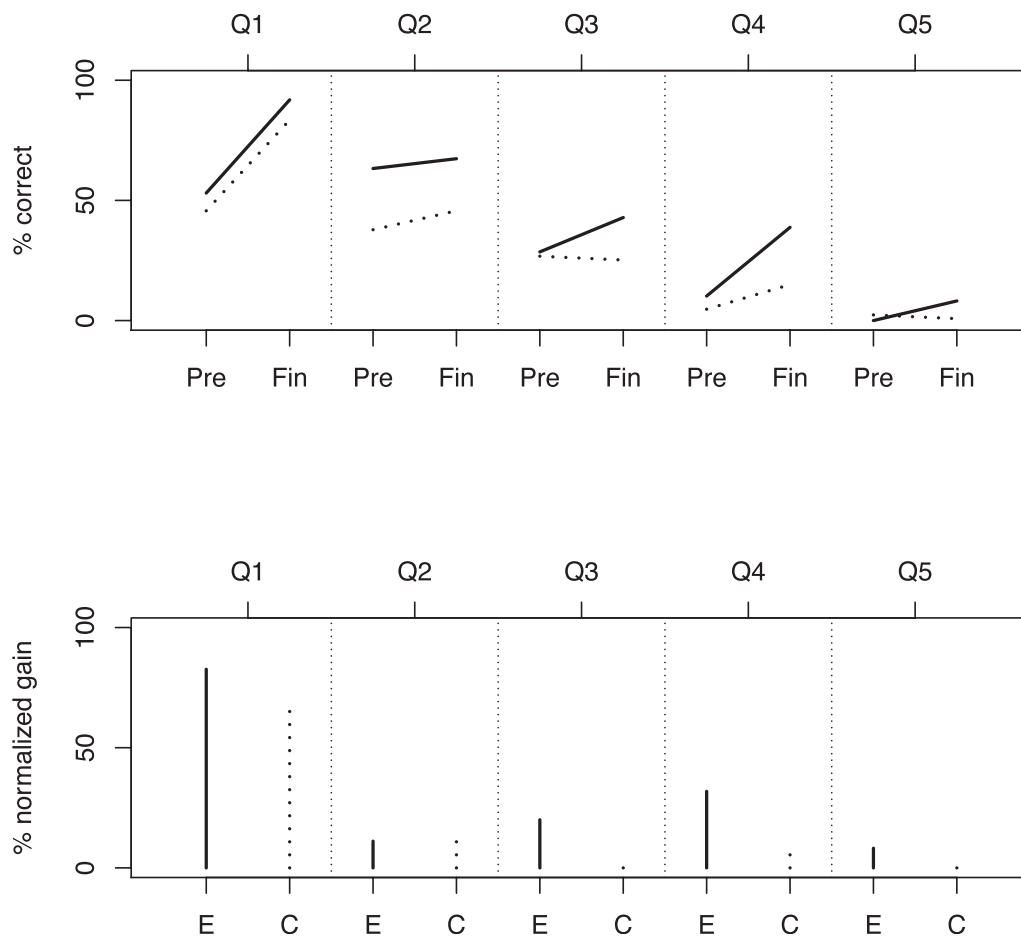


FIGURE 5: Percentage correct for five budget questions for the experimental (E; solid line) and control (C; dotted line) groups on the pretest (Pre) and final (Fin), as well as normalized gain scores.

of the experimental group's performance offers some insight into why this might be the case.

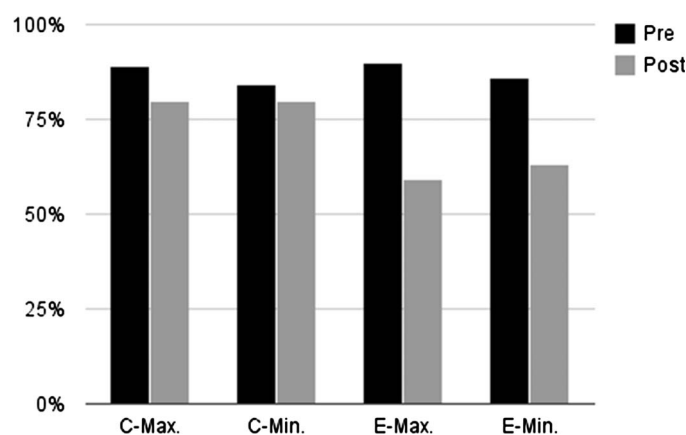
Students in the experimental group performed better on graphical interpretation of maximum stock levels (Question 4) than they did on graphical interpretation of minimum stock levels (Question 5) on the final exam; experimental group performance was quite poor on Question 5. Students in the experimental group were able to more correctly interpret maximum levels from a budget graph but did not fully apply the knowledge when interpreting a minimum value on the same graph. We suggest that this unequal performance may be a result of the instructional intervention requiring students to examine primarily maximum stock levels in the various budget scenarios. Future instruction

could benefit from the use of scenarios in which both maximum and minimum stock levels are examined, as well as providing students with multiple dissonant experiences addressing students' misconceptions in multiple ways (i.e., to address minimum inflow association with minimum stock level, in addition to maximum inflow association with maximum stock level).

Of the students in the experimental group, 90% used the pattern-matching misconception for identifying maximum stock levels on the pretest, while only 59% used it on the posttest, represented by a drop of 31% on the E-Max question (Fig. 6). There was a drop of 23% in the experimental group's use of the misconception for identifying minimum stock levels (E-Min). This compares to

TABLE IV: Results of fitting proportional log odds ratio models for four budget questions. Data are the number of students who had worse, had the same, or got better scores on the final than on the pretest. Questions 1–3 have nonsignificant results. Question 4 has a significant positive impact before and after adjustment for multiple testing. Question 5 did not warrant fitting such a model.

|            | Coefficient of Experimental Treatment | Standard Error | <i>t</i> Value | <i>p</i> Value | Adjusted <i>p</i> Value |
|------------|---------------------------------------|----------------|----------------|----------------|-------------------------|
| Question 1 | −0.06                                 | 0.33           | −0.17          | 0.87           | 1.00                    |
| Question 2 | −0.11                                 | 0.34           | −0.33          | 0.74           | 1.00                    |
| Question 3 | 0.61                                  | 0.36           | 1.69           | 0.09           | 0.45                    |
| Question 4 | 1.13                                  | 0.41           | 2.72           | 0.007          | 0.035                   |



**FIGURE 6:** Percentage of students relying on pattern-matching misconception for the experimental (E) and control (C) groups on the pretest and posttest (final). “Max.” corresponds to students who matched highest inflow on a graph with highest stock levels. “Min.” corresponds to students who matched lowest inflow on a graph with lowest stock levels.

respective drops of 9% and 4% in control group performance on these items. Assuming the reduction in reliance on this misconception is due to the instructional intervention, the data suggest that the experimental group may have experienced greater cognitive dissonance than did the control group, as reflected by the differential decrease in the number of experimental group students who chose the pattern-matching misconception on the posttest after participating in the intervention. We hypothesize that the experimental group students may have partially abandoned reliance on the misconception in interpreting maximum or minimum stock levels from a graph. Future research could address this hypothesis more directly.

Many of these students, however, still have an incomplete accurate conception of stock flow systems to replace the misconception as they fail to answer the graphical interpretation questions correctly. Our instructional intervention, therefore, likely promoted cognitive dissonance for about 30% of the students it targeted but was insufficient for students to deeply understand budget concepts, even for those who did experience cognitive dissonance.

The implication for instruction would be that greater effort at promoting cognitive dissonance should be made in similar instructional interventions, perhaps by requiring students to correctly answer questions and forcing them to recognize the shortcomings of their misconceptions before moving to the next task. In addition, more would need to be done to promote the accurate understanding of budget concepts by perhaps scheduling the intervention immediately before lecture instruction on the concepts. This approach should leave those students who have experienced cognitive dissonance in the intervention more ready to accept the scientific explanation presented during lecture. Furthermore, students should be allowed to ask questions and have their emerging understanding checked as a lecturer provides budget instruction to ensure adequate conceptual development. The use of personal response systems (clickers) in the classroom would assist with this step.

## CONCLUSIONS

Budget misconceptions cannot be effectively reduced by relying on lecture presentations alone, as often might be done in introductory science courses (Reichert et al., 2014). When students engage in an ill-structured, real-world case with multiple representations of budget scenarios challenging budget misconceptions, some notable learning gains are made and fewer students rely on misconceptions when answering graphical interpretation questions. However, even after carefully targeting students’ misconceptions, providing feedback in real time, and providing students a chance to modify their thinking in response to feedback, the majority of students still rely on pattern-matching misconceptions when interpreting graphical information. Learning gains and budget misconception reductions require student misunderstandings to be explicitly challenged through experiences providing cognitive dissonance in which students must wrestle with information they can only explain when mass–balance concepts are understood. The results of this study lead us to conclude that some students can learn budget concepts when time on the task is increased and when the students engage in the application of scientific concepts in multiple contexts but that budget misconceptions are also difficult to overcome.

## Acknowledgments

We thank Pete Boysen for his assistance in the creation of the Thinkspace activity and for his technical support. We thank Holly Bender for funding the upgrade of this activity within Thinkspace 2.0. We also thank John Beyers for his helpful insight in the formulation and analysis of the Thinkspace activity and graduate students in the Departments of Geological and Atmospheric Sciences and School of Education at Iowa State University for the evaluation of questions and review of the activity before it was deployed with students. The thorough review and comments by two reviewers and one associate editor, Research Editor Alison Stokes, and Editor-in-Chief Kristen St. John helped us greatly improve this paper.

## REFERENCES

- Arndt, D.S., and Laude, D.S. 2008. Applying concepts of adult education to improve weather and climate literacy. *Physical Geography*, 29(6):487–499.
- Bardsley, D.K., and Bardsley, A.M. 2007. A constructivist approach to climate change teaching and learning. *Geographical Research*, 45(4):329–339.
- Baron, R.M., and Kenny, D.A. 1986. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6):1173–1182, doi:10.1037/0022-3514.51.6.1173. PMID 3806354.
- Cronin, M.A., and Gonzales, C. 2007. Understanding the building blocks of system dynamics. *System Dynamics Review*, 23:1–17.
- Cronin, M.A., and Gonzales, C. 2009. Why don’t well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. *Organizational Behavior and Human Decision Processes*, 108:116–130.
- DeWaters, J.E., Andersen, C., Calderwood, A., and Powers, S.E. 2014. Improving climate literacy with project-based modules rich in educational rigor and relevance. *Journal of Geoscience Education*, 62(3):469–484.
- Driver, R., Asoko, H., Leach, J., Mortimer, E., and Scott, P. 1994.

- Constructing scientific knowledge in the classroom. *Education- al Researcher*, 23(7):5–12.
- Harrington, J., Jr. 2008. Misconceptions: Barriers to improved climate literacy. *Physical Geography*, 29(6):575–584.
- Huntoon, J.E., and Ridky, R.K. 2002. Examining long-term global climate change on the web. *Journal of Geoscience Education*, 50(5):497–514.
- Intergovernmental Panel on Climate Change (IPCC). 2007. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., and Miller, H.L., eds. Cambridge, UK: Cambridge University Press.
- Knowles, M., Holton, E.F., and Swanson, R.A. 2005. The adult learner: The definitive classic in adult education and human resource development, 6th ed. Burlington, MA: Elsevier.
- Lachin, J.M. 2010. Biostatistical methods: The assessment of relative risks, 2nd ed. Hoboken, NJ: John Wiley & Sons, 672p.
- McCaffrey, M.S., and Buhr, S.M. 2008. Clarifying climate confusion: Addressing systematic holes, cognitive gaps, and misconceptions through climate literacy. *Geographical Research*, 29(6):512–528.
- Meadows, G., and Weisenmayer, R. 1999. Identifying and addressing students' alternative conceptions of the causes of global warming: The need for cognitive conflict. *Journal of Science Education and Technology*, 8(3):235–239.
- Moxnes, E., and Saysel, A.K. 2009. Misperceptions of global climate change: Information policies. *Climatic Change*, 93:15–37.
- Piaget, J. 1963. The origins of intelligence in children. New York: Norton.
- Posner, G.J., Strike, K.A., Hewson, P.W., and Gertzog, W.A. 1982. Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2):211–227.
- Rebich, S., and Gautier, C. 2005. Concept mapping to reveal prior knowledge and conceptual change in a mock summit course on global climate change. *Journal of Geoscience Education*, 53(4):355–365.
- Reichert, C., Cervato, C., Larsen, M., and Niederhauser, D. 2014. Students' misconceptions of atmospheric carbon budgets: Undergraduate students' perceptions of mass balance. *Journal of Geoscience Education*, 62:460–468.
- Sakamoto, Y., Ishiguro, M., and Kitagawa, G. 1986. Akaike information criterion statistics. Dordrecht, Germany: D. Reidel.
- Schuster, D.A., Filippelli, G.M., and Thomas, C.W. 2008. Secondary students' subject matter representations of climate change. *Journal of Geoscience Education*, 56(4):307–316.
- Scott, P.H., Asoko, H.M., and Driver, R.H. 1991. Teaching for conceptual change: A review of strategies. In Duit, R., Goldberg, F., and Niederer, H., eds., Research in physics learning: Theoretical issues and empirical studies. Kiel, Germany: Schmidt & Klanning, p. 310–329.
- Sobel, M.E. 1982. Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, 13:290–312. doi:10.2307/270723.
- Sobel, M.E. 1986. Some new results on indirect effects and their standard errors in covariance structure. *Sociological Methodology*, 16:159–186. doi:10.2307/270922.
- Spiro, R.J., Coulson, R.L., Feltovich, P.J., and Anderson, P.J. 1988. Cognitive flexibility theory: Advanced knowledge acquisition in ill-structured domains. In Proceedings of the 10th Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Lawrence Erlbaum Associates, p. 375–383.
- Spiro, R.J., Feltovich, P.J., Jacobsen, M.J., and Coulson, R.L. 1992. Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill structured domains. In Duffy, T.M., and Jonassen, D.H., eds., Constructivism and the technology of instruction: A conversation. Hillsdale, NJ: Lawrence Erlbaum Associates, p. 57–75.
- Spiro, R.J., and Jehng, J.C. 1990. Cognitive flexibility and hypertext: Theory and technology for the nonlinear and multidimensional traversal of complex subject matter. In Nix, D., and Spiro, R., eds., Cognition, education and multimedia: Exploring ideas in high technology. Hillsdale, NJ: Lawrence Erlbaum Associates, p. 163–205.
- Sterman, J.D. 2008. Economics-risk communication on climate: Mental models and mass balance. *Science*, 322:532–533.
- Sterman, J.D., and Sweeney, L. 2007. Understanding public complacency about climate change: Adults' mental models of climate change violate conservation of matter. *Climatic Change*, 80:213–238.
- Sweeney, L., and Sterman, J. 2007. Thinking about systems: Student and teacher conceptions about natural and social systems. *System Dynamics Review*, 23(2/3):285–312.
- Woolfolk, A. 2007. Educational psychology, 10th ed. Boston, MA: Pearson/Allyn and Bacon.