# The Effectiveness of Rotating Tank Experiments in Teaching Undergraduate Courses in Atmospheres, Oceans, and Climate Sciences

Kathleen J. Mackin,[1] Nancy Cook-Smith,[2] Lodovica Illari,[3] John Marshall,[3,a] and Philip Sadler[2]

## ABSTRACT

While it is commonly recognized that laboratory experiments and demonstrations have made a considerable contribution to our understanding of fluid dynamics, few U.S. universities that offer courses in meteorology and/or oceanography provide opportunities for students to observe fluid experiments in the classroom. This article explores the evaluation results of a three-year, NSF-funded project in partnership with the Massachusetts Institute of Technology (MIT) and five universities nationally, to provide laboratory demonstrations, equipment, and curriculum materials for use in the teaching of atmospheres, oceans, and climate. The aim of the project was to offer instructors a repertoire of rotating tank experiments and a curriculum in fluid dynamics to better assist students in learning how to move between phenomena in the real world and basic principles of rotating fluid dynamics, which play a central role in determining the climate of the planet. The evaluation highlights the overwhelmingly positive responses from instructors and students who used the experiments, citing that the *Weather in a Tank* curriculum offered a less passive and more engaged and interactive teaching and learning environment. Results of three years of pre- and posttesting on measures of content related to atmospheres, oceans, and climate sciences with over 900 students in treatment and comparison conditions, revealed that the treatment groups consistently made greater gains at the posttest than the comparison groups, especially those students in introductory level courses and lab courses. *© 2012 National Association of Geoscience Teachers*. [DOI: 10.5408/10-194.1]

*Key words*: education testing and evaluation, education science experiments, education undergraduate, atmospheric and oceanic sciences, teaching and curriculum, climate, fluids, rotating fluids, geophysical fluid dynamics

## INTRODUCTION

The study of atmospheres, oceans, and climate is rooted in a special kind of fluid dynamics known as Geophysical Fluid Dynamics (GFD). GFD combines rotation and stratification, and explores the often counterintuitive properties of rotating, stratified fluids and how those properties manifest themselves in the circulation of the atmosphere and ocean. With the realization that understanding climate is one of the most pressing challenges facing humankind (Ki-moon, 2007), it is very important that GFD should not be considered a specialist subject, but instead the foundation of a complete education in atmospheres, oceans, and climate sciences for science majors and non-science majors alike (Knox, 2008).

The study of atmosphere–ocean dynamics is often introduced to students as a branch of applied mathematics in which the governing equations are written down in a frame of reference rotating with the earth, simplified by artful scaling assumptions and solved in simplified settings. Sometimes, and one is tempted to say more often than not, the physical connection to real phenomena in the atmosphere and ocean is lost in the process. Like much teaching in the physical sciences, courses are often characterized as passive learning environments dominated by a lecture format instead of more active environments that engage students and promote motivation and learning (Sokoloff and Thornton, 1997; Pandya et al., 2004).

In an attempt to enrich the learning experience of students of atmospheres, oceans, and climate, educators Dr. Lodovica Illari and Dr. John Marshall of the Department of Earth, Atmospheric, and Planetary Sciences at the Massachusetts Institute of Technology (MIT), decided to explore the use of simplified laboratory experiments to help bridge the gap between observations and theory. The intent was to better help students learn how to move between phenomena in the real world, laboratory abstractions, and theory (Turner, 2000). The NSF-funded project, known as *Weather in a Tank* (2006–2009), is described in Illari et al. (2009) and on the project Web site, http://paoc.mit.edu/labguide, together with an associated textbook by Marshall and Plumb (2008), *Atmosphere, Ocean and Climate Dynamics: An Introductory Text*. The purpose of the project was to provide professors at several sites nationally with an educational resource, including laboratory experiments, associated equipment, curricular materials, and real-world examples, to assist them in their teaching, and to evaluate whether teaching with *Weather in a Tank* was effective pedagogy.

In this paper we present the evaluation results of the project. To place our study in context, we begin in "Background and Context" by reviewing previous explorations of the use of laboratory experiments in the teaching of atmosphere/ocean science. In "Teaching with Weather in a Tank" we describe how we teach with *Weather in a Tank* and, in "Implementation at Collaborating Universities," how the project was implemented at collaborating universities. A formative and summative evaluation of the success of the approach is presented in "Results of the Evaluation," Finally, in "Conclusions and Implications for Instruction and

*Received 14 July 2010; revised 1 August 2011; accepted 26 August 2011; published online 16 February 2012.*

[1]Mackin Education Consulting, Stratham, New Hampshire 03885, USA
[2]Science Education Department, Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA
[3]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
[a]Author to whom correspondence should be addressed. Electronic mail: johncmarshall54@gmail.com.

Learning" we present our conclusions and the wider implications of our study for the teaching of weather and climate in our schools and universities.

## BACKGROUND AND CONTEXT

### Students' Understanding of Scientific Phenomena

To teach successful college courses in atmospheres, oceans, and climate, it is important to be aware that students come to college with their own unique conceptual framework that often includes misconceptions about weather and climate phenomena. Student conceptions about weather and the physical properties of the atmosphere and oceans, prior to instruction, have been reported extensively in the research literature, but primarily at the precollege level (Henriques, 2002; Libarkin and Anderson, 2005).

While almost all students hold misconceptions about scientific phenomena, younger students' views are qualitatively different from college-age students and adults. Young students often attribute natural phenomena to supernatural actions or conceptualize phenomena in a way that violates physical laws (Piaget et al., 2007). The reasoning of older students often employs "phenomenological primitives" (p-prim) knowledge elements that abstract a particular phenomenon, but are often misapplied (Hammer, 1996). Several authors have provided common examples of these misapplications among high school and college-age students (Rollins et al., 1983; Aron et al., 1994; Stepans, 1994; Harrington, 2008). For instance, Aron et al. (1994) report on students' tendency to conflate density of air with their sensory experience of it, thinking that because humid air feels "oppressive," it is actually denser than dry air. Many college-age students correctly assume that the sun's radiation plays a role in weather, but often mistakenly believe that the seasons result from the earth's elliptical orbit about the sun (Rollins et al., 1983). The seasons, instead, are a consequence of the tilt of the earth's spin axis from the vertical.

### Students' Conceptions of Rotating Systems

Less is known about student misconceptions concerning rotating systems, especially as applied to the oceans and atmosphere. The Coriolis force is subtle (see, for example, Persson, 1998). Much as the ancients believed that circular motions were "natural," students think that particles following a curved track will, when free to move in any direction, continue to curve rather than move in a straight line (Gunstone, 1971; Hestenes et al., 1992). From common experiences students often conclude that Coriolis forces impact the direction of drainage from bathtubs differently for northern and southern hemispheres, even though the effect is negligible at this scale (Nelson et al., 1992). Barowy and Lochhead (1980) state that proper conceptualization of rotating systems generally requires that students ignore superficial perceptual features and instead concentrate on constructing physically relevant representations. For such systems, compelling physical phenomena sometimes offer few clues to forces acting on component particles.

### Limitations of Conventional Approaches to Teaching the Geosciences

Conventional approaches to teaching that include lecture and simulations in the geosciences often do not result in changes in student misconceptions (Gyllenhaal and Perry, 1998; Hay et al., 2000). Classroom demonstrations are used with some frequency by science teachers, but often do not have a major impact on student understanding of science concepts (Crouch et al., 2004). This is most likely due to the fact that they are often poorly integrated in to the content of the course or taught in large enrollment classes that allow students little hands-on experience (Kahl, 2008). Students' lack of mastery of relevant theoretical or representational frameworks can also limit the impact of demonstrations. Moreover, there is often a lack of opportunity for students to interact with the demonstrations or propose modifications to test their ideas (Roth et al., 1997). Strategies that appear effective in teaching in the geosciences include the utility of students' drawings as a preassessment tool (Gobert, 2005). Also, in laboratory settings, students are typically freer to repeat or modify elements of an experiment or engage in more open-ended investigations that encourage deeper understanding (Hofstein and Lunetta, 2003).

### Use of Laboratory Experiments in Teaching

Many educators agree that the potential for learning is greatly enhanced through exposure to laboratory experiments (Thornton, 1996; Turner, 2000; Roebber, 2005). For example the Technology Enabled Active Learning (TEAL) project at MIT has shown that students tend to learn better in a more active environment (Dori and Belcher, 2005). The MIT experience in teaching undergraduate physics in small groups with advanced technology indicates the benefit of interactivity, visualization, and hands-on experiments. Providing opportunities for students to visualize phenomena and engage in inquiry-based science instruction maximizes the learning experience and encourages more abstract and theoretical thinking (National Research Council, 1996, 2000; Hofstein and Lunetta, 2004).

In the field of meteorology and oceanography, however, it appears that laboratory experiments currently play only a minor role in the education of our students (Gynnild et al., 2007). This is probably due to the misconception that relevant laboratory experiments are too complicated to be carried out in classes and require the backup of a sophisticated fluid laboratory. Although the internet has made meteorological data widely available, to the great benefit of all, it is unlikely that many students are being exposed to real fluids in their undergraduate education.[1] They see data manipulated over the Web, movie loops from numerical models, but rarely get their hands wet with a real fluid. One significant indication of its rarity is that, prior to Marshall and Plumb (2008), there was no textbook that

[1]In 2006, Marshall and Sadler sent out a questionnaire to approximately 1,000 U.S. Professors teaching Atmospheric Science to gauge their interest in the use of laboratory experiments in teaching. Of the 5% who replied, over 75% were very positive. A more detailed analysis showed that 91% are in favor of using simple laboratory experiments in class demonstrations, 55% are in favor of setting up a laboratory course, and 88% are interested in trying out some of the experiments. These findings were very encouraging, especially if interpreted in the context of the type and size of the classes taught by our targeted professors: 13% were teaching large classes (>50), 24% small classes (<20) and 63% medium classes (20–50). The survey suggests that in classes that are sufficiently large, experiments are best carried out in a demonstration mode. However, many also expressed interest in setting up hands-on laboratory courses. These broad conclusions are borne out by the *Weather in a Tank* project results presented here.

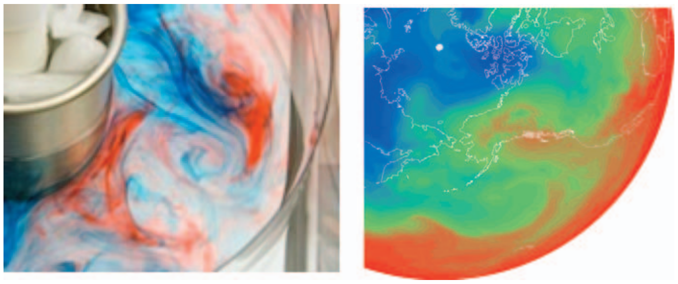

**FIGURE 1: (Left) A laboratory analogue of weather systems in which an ice can placed in the middle of a rotating tank of water induces a radial temperature gradient. The presence of eddies in the tank, analogous to the atmospheric weather systems shown on the right, can be seen through the swirling dye patterns. (Right) A view of temperature variations at a height of 2 km showing swirling regions of warm (red) and cold (blue) air associated with synoptic-scale weather systems. The North Pole is indicated by the white dot at the upper left of the figure. See http://paoc.mit.edu/labguide. See online article to view color version of figure.**

makes comprehensive use of laboratory experiments or draws them into its content. Thus, teachers lack access to resources or equipment to support the use of experiments in the classroom. Moreover, when laboratory experiments are used there is little empirical evidence demonstrating the effect of experiments on student learning in these fields or whether experiments have any impact on student engagement in the scientific process (Fox and Hackerman, 2003; Singer, et. al., 2005; Nelson et. al, 2010).

It was in the context of the above research that we began the *Weather in a Tank* project. We developed simple, transparent experiments, focused on fundamentals that had an immediate connection to the real world (e.g., the "rotating annulus" experiment shown in Fig. 1), together with atmospheric data illustrating weather systems. Our hope was that these would benefit the learning of all students, irrespective of their background or sophistication with mathematics or physics. Because Earth's rotation has such a profound influence on the circulation of the atmosphere and ocean, and yet is a difficult concept to teach effectively, we also focused attention on rotating laboratory experiments.

The *Weather in a Tank* project, then, (1) created a resource to enable professors to use demonstrations in their teaching and (2) engaged an external evaluator, Dr. Kathleen Mackin, to evaluate the three-year implementation of the project to determine whether there was an impact on classroom teaching and student learning. Additional statistical analyses of student outcome data were conducted in collaboration with Dr. Nancy Cook-Smith and Dr. Philip Sadler of the Science Education Department (SED) of the Harvard-Smithsonian Center for Astrophysics. Here we report on our three-year examination of undergraduate student learning after the introduction of a series of structured rotating tank experiments across a spectrum of large- and small-enrollment atmospheric and oceanic science courses at six universities. Our investigation includes extensive formative and summative evaluations that examine:

1. The extent to which instructors valued and used the *Weather in a Tank* experiments and curriculum in the college classroom;
2. The effect of the *Weather in a Tank* curriculum on teaching in undergraduate atmospheric and oceanic science courses;
3. The learning impact among students introduced to the experiments across college courses and universities involved in the project; and
4. The reactions of students to the *Weather in a Tank* experiments and curriculum used in their courses.

## TEACHING WITH WEATHER IN A TANK

The *Weather in a Tank* project has developed an extensible list of teaching modules (currently there are a dozen or more) focusing on carefully chosen phenomena that are central to our understanding of atmospheres, oceans, and climate.[2] The modules provide a unique, integrated program linking hands-on experience in the laboratory with real-world examples and relevant theory, enabling students to gain a deeper understanding of the physics of these systems. They were designed to be used in classroom and laboratory settings to create an active learning environment and engage students in all aspects of the learning process from activation of prior knowledge to engagement in further inquiry. Many experiments are also useful in outreach activities to non-specialists and the public.

The modules integrate and describe

1. detailed information on how one sets up and runs the laboratory experiment;
2. demonstration material for use by teachers in the classroom;
3. write-ups on the relevant theory; and
4. descriptions and data from real-world situations that are dominated by (or have a significant contribution from) the phenomenon being studied.

A comprehensive discussion of the use of the curriculum materials can be found in Illari et al. (2009). Information about how to obtain low-cost equipment required to carry out the associated experiments can be found there and at the project Web site.

We now describe one particular module in more detail, the "balanced vortex" experiment, to give a better feel for what is involved and the approach we are advocating.

### Description of the Balanced Vortex (Radial Inflow) Project

The balanced vortex project was designed to illustrate the dynamics of vortices, from familiar examples such as the swirl associated with the flow of water down a drain hole to huge, destructive vortices such as hurricanes in the atmosphere. The curriculum materials developed to study vortices are (1) rooted in basic physical principals such as conservation of angular momentum, (2) illustrated by a rotating fluid experiment, and (3) make use of real-world

[2]See Web site http://paoc.mit.edu/labguide/projects.html for list and descriptions.

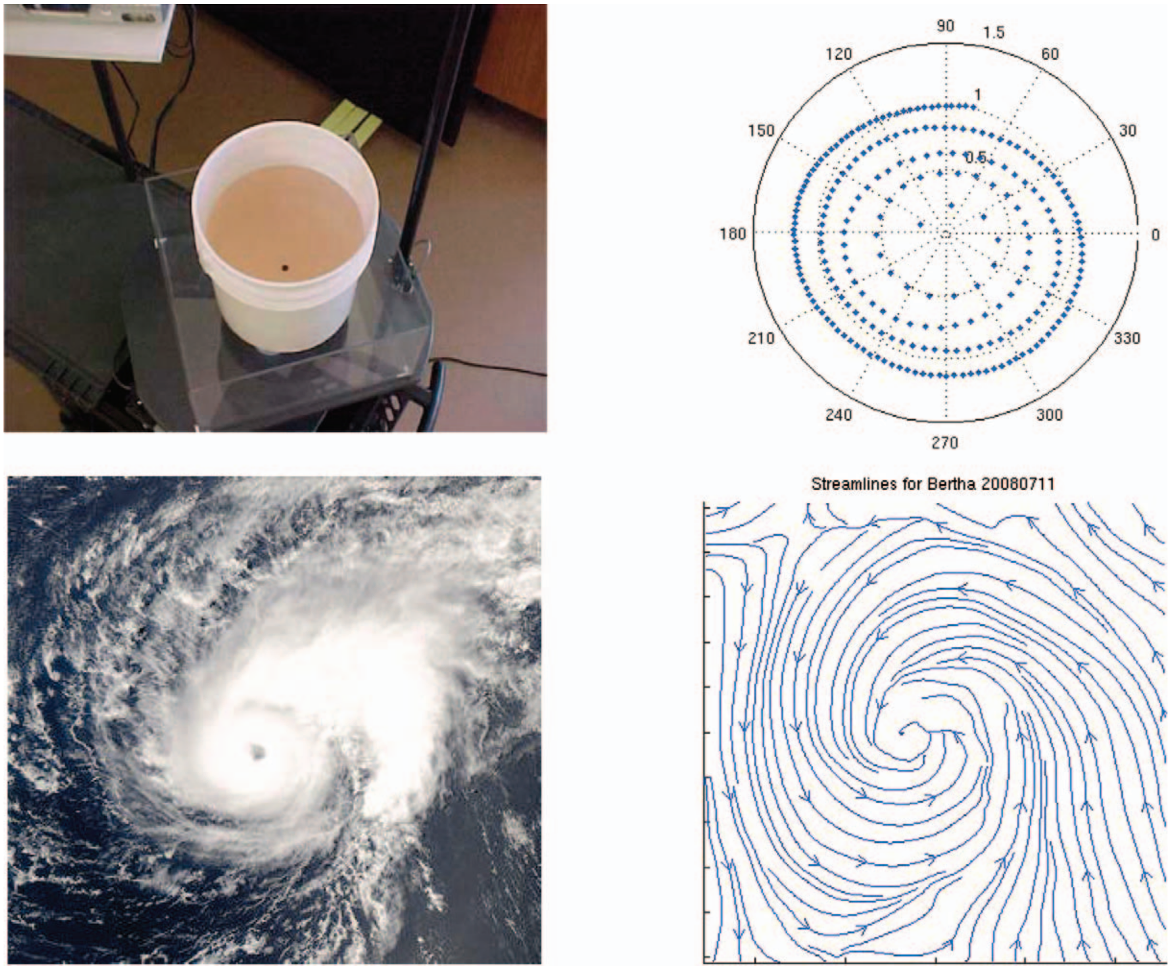


**FIGURE 2: (Top left) The plastic bucket on the dial of a rotating turntable in which a laboratory vortex is formed. (Top right) The trajectory of a paper dot floating on the free surface of the vortex formed from water spiraling inward toward the drain hole. (Bottom left) Spiral cloud formations associated with Hurricane Bertha on September 7, 2008. (Bottom right) Streamlines of surface flow around the eye of Hurricane Bertha on September 7, as revealed by QuikScat wind data.**

example of vortices (e.g., hurricanes) using datasets available over the Web.

The rotating tank laboratory experiment centers on a cylindrical tank filled with water rotating about its vertical axis: the cylinder has a circular drain hole in the center of its bottom blocked by a stopper. After solid body rotation is achieved the stopper is released and the water moves inward, conserving angular momentum and, in so doing, acquires a swirling motion. The swirling motion can become very vigorous even if the cylinder is rotated slowly (a few rpm), because the angular momentum of the cylinder is "concentrated" by inward flowing rings of fluid. At modest rotation rates the effect of rotation is pronounced and parcels complete many circuits before finally exiting through the drain hole (Fig. 2 [top right]). At high rotation the free surface becomes markedly curved, high at the periphery and plunging downwards toward the hole in the center, providing a strong analogy to the eye of a hurricane.

Using particle tracker software, students observe and track the trajectory of particles floating on the surface of the vortex as they spin into the drain hole, compute azimuthal and radial current speeds in inertial and rotating frames, and interpret in terms of theory. The behavior of the laboratory vortex is compared to meteorological phenomena, such as observations of winds associated with Hurricane Bertha, (Fig. 2 [bottom]). Satellite winds from the QuikSCAT scatterometer clearly show vigorous, low level swirling motion around the low-pressure region at the center of the hurricane (Fig. 2 [bottom right]).

This balanced vortex module not only addresses a number of important principles of rotating fluid dynamics (e.g., frames of reference, conservation of angular momentum) but also provides a route to more advanced ideas, such as the balance of forces associated with geostrophic, gradient wind, and cyclostrophic balance. Moreover, it gives a context in which to introduce and discuss the Rossby number, the key nondimensional number governing rotating flows.

The Rossby number is a dimensionless measure of the ratio of inertial forces to Coriolis forces in a rotating fluid (see Fig. 3 and discussion in Marshall and Plumb, 2008, Ch. 6). In the balanced vortex module, the Rossby number is also

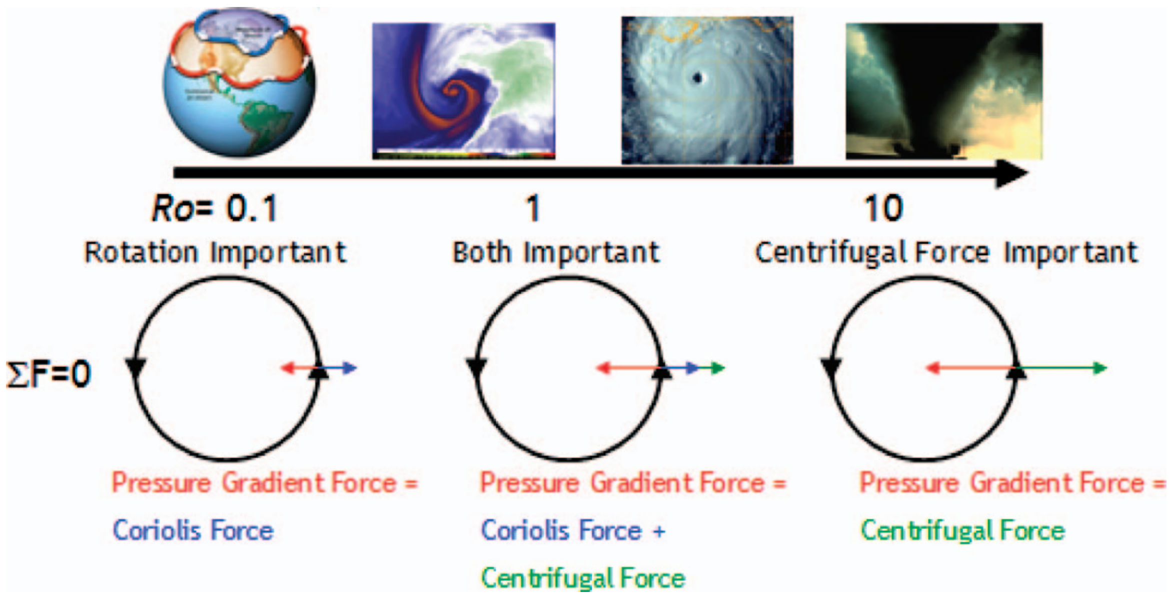


**FIGURE 3: Significance of the Rossby number as it relates to balance of forces, F. Planetary-scale flows on the earth have small Rossby numbers, indicating a balance between pressure gradient and Coriolis forces. Smaller-scale weather systems and fronts have Rossby numbers of order unity, indicating that centrifugal forces also play a role. In hurricanes and tornadoes indicated on the right, Rossby numbers become very large (exceeding 10) and Coriolis plays a negligible role in the force balance. The balanced vortex laboratory experiment illustrates the balance of forces across this whole range of Rossby number.**

presented as a quantity that measures the ratio of two timescales: the period of the rotating system relative to, in this context, the time it takes for a parcel of fluid to circle the tank. If the parcel takes many rotation periods to circle the tank, then the Rossby number (Ro) is small and rotation has a profound influence on the motion and vice versa. At different radii from the center of the bucket, our laboratory vortex displays vastly different Rossby numbers, small on the outside and large toward the center. By measuring the speed of the particles as a function of radius[3] students experimentally determine the Rossby number and compare it against theoretical predictions based on angular momentum conservation. Importantly, they also compare measurements and predictions, appropriately scaled, to observations of balanced vortices in the atmosphere, such as hurricanes (see Figs. 2 and 4). Indeed, an insightful way to interpret the data in terms of theory is to plot the measured Rossby number as a function of radius, both for the laboratory vortex and the hurricane. Figure 4 (left) shows a graph of experimental laboratory data. Using observed data from hurricane Bertha, the Rossby number as a function of radius measured from the eye of the hurricane to the periphery of the system is shown in Figure 4 (right). The similarity between the two graphs is striking and intriguing to students. They go on to quantify and interpret in terms of simple theory (see Marshall and Plumb, 2008, Sections 6.6.1 and 7.1.3 for examples).

The module briefly described above, as in all of the activities developed in *Weather in a Tank*, can be readily customized by teachers to best target the skills and interests of their students, whether they are beginners or more advanced.

[3]Students record video footage of the path of a paper dot swirling in the vortex using an overhead camera corotating with the turntable. They then track the vortex using a particle tracker that can be downloaded from the Web at http://ravela.net/particletracker.html.

## Classroom Implementation

The *Weather in a Tank* project provides a resource and is not meant to be prescriptive. Teachers can select experiments and material from the modules that they feel are most appropriate to fit their particular educational setting and students. However, to set a context in which to evaluate the effectiveness of the teaching strategy being advocated, instructors participating in the project were encouraged to use at least four modules throughout the semester to familiarize students with the experiential approach and to introduce them to the scientific process. It was also important that the professor teaching the class actually be involved in conducting the experiment, in so doing emphasizing the centrality of the experiment to the content of the course.

The following elements were encouraged in using the *Weather in a Tank* modules, constituting a comprehensive implementation of the project experiments and curriculum materials. Many science instructors and researchers encourage these same methods to engage students intellectually, promote more analogical reasoning and conceptual change, and establish a more inquiry-based science approach in the classroom (Sokoloff and Thornton, 1997; Gentner, et al., 2003).

1. **Student Predictions.** Encourage individual students to make predictions and record their conjectures. To avoid any hesitation on the part of the students, assure

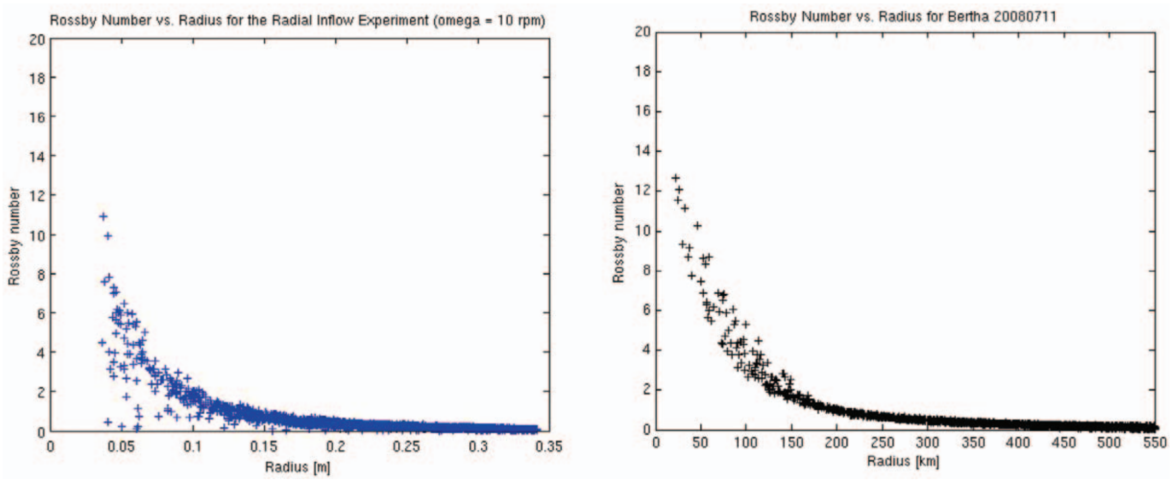


**FIGURE 4: (Left) Rossby number (Ro) plotted as a function of the particle's radius (in cm). (Right) The Rossby number plotted as a function of radius (km) for Hurricane Bertha, using near-surface winds from QuikScat data.**

them that their predictions will not be graded. Making predictions will activate their prior knowledge of the subject, engage them in the phenomena being demonstrated, and ensure that they have a stake in the outcome of the demonstration. The instructor can ask students to volunteer their predictions and draw comparisons across student responses.[4]

2. **Conduct the Experiment.** Conduct the experiment allowing individual students or small groups of students to assist where appropriate. For instance, students can administer the dye or float the paper dots. Encourage students to gather around the tank for better viewing or display the experiment in progress on a large-screen projection monitor.
3. **Observe the Experiment.** Give the students ample time to absorb the effects of the experiment. Let the students simply observe and resist the attempt to explain everything that is happening. This will strengthen their powers of observation and promote their natural interest and curiosity. Simply observing what is happening in the experiment will challenge their assumptions and predictions, and test their conclusions.
4. **Encourage Questioning and Discussion**. Encourage students to talk about the phenomena being observed, either in small groups or as a whole class. Assist students in formulating initial conclusions about their observations and deal with any misconceptions and incorrect assumptions.
5. **Make Connections with Real-World Phenomena.** Assist students in deepening their understanding of phenomena by expanding on the concept that is the focus of the experiment and making connections to the real world using theory, as appropriate. Engage students in a discussion of similar phenomena based on the same concepts. This will allow students to bridge the gap between concrete phenomena and the abstract.
6. **Provide Opportunities for Further Inquiry**. Experiments naturally foster deeper questioning about a phenomenon or concept. Provide opportunities for students to follow up on their questions and test their ideas more thoroughly by repeating experiments or in the design and execution of new experiments. Create opportunities for students to conduct further research to explore a concept or a phenomenon.[5]
7. **Bridge to Mathematical Models.** Allow students to explore mathematical models. For example the use of the Rossby number in the balanced vortex module (Figs. 3 and 4) acts as a unifying concept bridging between the laboratory experiments and real world phenomena, and provides a clear example of how a fundamental theoretical concept can be introduced with the aid of laboratory experiments.

## IMPLEMENTATION AT COLLABORATING UNIVERSITIES

The MIT project staff provided the *Weather in a Tank* apparatus and equipment to collaborating professors in the five universities located in the Northeast and Midwest states, listed in Figure 5, and coordinated efforts to implement the experiments in their courses. Over the 3 y of the project, the *Weather in a Tank* experiments were used instructionally with over 700 students in 26 undergraduate atmospheric

[4]In some projects we have found it useful to make use of, and encourage students to discuss, their results in the context of a 2 × 2 matrix of experiments, as discussed on the project Web site. This allows a sequence of experiments to be discussed together, in which two parameters take on two values (e.g., high and low rotation, high and low temperature gradient), yielding four experiments in all.

[5]See, for example, the MIT Synoptic Laboratory, http://paoc.mit.edu/synoptic/ (Illari, 2001) for real time data; the NASA Earth Observatory, http://earthobservatory.nasa.gov/NaturalHazards/ for beautiful satellite images; and the COMET (UCAR), http://www.comet.ucar.edu for interesting case studies.

***Weather in a Tank* Collaborators**

**Thomas Haine**, Department of Earth and Planetary Sciences, *Johns Hopkins University*

**Lodovica Illari and John Marshall**, Department of Earth, Atmospheric and Planetary Sciences, *Massachusetts Institute of Technology (MIT)*

**Richard Clark, Ajoy Kumar, and Todd Sikora**, Department of Earth Sciences, *Millersville University*

**Sukyoung Lee, Peter Bannon and Ray Najjar,** Department of Meteorology, *Pennsylvania State University*

**Amit Tandon**, Department of Physics/Department of Estuarine and Ocean Sciences, *University of Massachusetts Dartmouth*

**Galen McKinley and Michael Morgan**, Department of Atmospheric and Oceanic Sciences, *University of Wisconsin, Madison*

FIGURE 5: Collaborating universities and professors.

science and climatology courses at these universities. Courses included introductory offerings, such as Introduction to Atmospheric Science; intermediate and advanced courses targeting majors in the discipline, such as Synoptic Meteorology and Earth and Planetary Fluids; and small lab-based offerings, such as a climate and weather laboratory and a meteorology lab. The students enrolled in these courses were primarily undergraduates majoring in atmospheres, oceans, and climate or related sciences, but also included students majoring in other disciplines, such as psychology, education, and business management.

## Data Collection

### *Instructor Logs*

In order to understand how the instructors used the demonstrations and to what benefit, they were asked to complete a brief weekly log detailing the experiments they used, curriculum materials, student reactions, and any challenges or benefits they experienced in using the apparatus, the experiments, and/or the Web site. This log was submitted electronically to the MIT project staff and to the external evaluator, Dr. Kathleen Mackin, via the project Web site. (A sample of the Instructor Log and other evaluation instruments and protocols can be found in Mackin, 2008).

The weekly log provided critical information to the project staff allowing them to monitor any challenges instructors were having in using the equipment or experiments, make any needed corrections to the Web site, provide additional instructions and guidance, and track the number and type of experiments that were being used in classes each semester. Some of these issues were also shared at the yearly collaborator meetings and during project staff site visits with collaborators. The instructor logs also provided important information about student reactions to the demonstrations from the perspective of instructors (e.g. increased motivation and interest, greater questioning and discussion and/or efforts to conduct further inquiry).

### *Web-based Surveys*

In order to corroborate information gleaned from the instructor logs, students were also asked to respond to an anonymous Web-based survey that requested their opinions of the impact of these laboratory fluid experiments on their learning as well as their perceptions of any improvements in classroom atmosphere and instructional quality as a result of using the *Weather in a Tank* experiments and curriculum.

## Assessment of Student Learning

To understand student gains in content knowledge as a result of instruction using the experiments, we designed and implemented a pre/posttest measure that was given to students at the beginning and end of courses each semester (see Fig. 6).

On the pre- and posttest Scantron form accompanying the test, students were asked to supply additional demographic data (e.g. date of birth, major, SAT/ACT math and verbal scores, college level, and gender). Collection of demographic data allowed the external evaluator to not only match pre- and posttests by student to measure gains, but also analyze outcome data by subgroups of students.

To ensure robustness of the study and to support or negate the assumption that student learning gains could be linked to the *Weather in a Tank* experimental approach, we enlisted additional students enrolled in similar courses at the collaborating colleges to serve as a comparison group, students who did not have access to the experiments. We used a comparison group instead of a control group because control groups in education are fraught with problems that are difficult to overcome. For instance, a true control group requires randomization and matches by types of students on a number of variables, such as gender and level of educational achievement, factors that are impossible to control during the normal course enrollment process. Also, educators are naturally reluctant to exclude one set of students from potentially promising educational interventions while providing them to others.

Table I displays the number of matched samples of student pre- and posttests that were obtained for analysis purposes for the Treatment and Comparison groups. Over the course of the 3 y, 914 matched[6] pre- and posttest scores were collected and analyzed from undergraduate students, 458 and 456, respectively, in the Treatment and Comparison groups. As illustrated in Table I, there were considerably more students enrolled in introductory courses in the Comparison group than in the Treatment group, 345 and 240 respectively. The number of students enrolled in intermediate/advanced courses in the Treatment and Comparison conditions was similar, 174 and 111 respectively; no lab classes were represented in the Comparison group. In selecting a comparison group for this study we did encounter some of the problems associated with nonrandomized student samples. For instance, there are major differences between students in introductory and intermediate/advanced courses, particularly in the gender makeup, fraction of freshmen, and choice of major. While the three intermediate/advanced groups are rather similar in background, the introductory Comparison and Treatment groups differ considerably in gender makeup, SAT/ACT scores, fraction of freshmen, and choice of major (see Tables I and II). These differences could influence the pre- and posttest results. To overcome these limitations, the data from the Treatment and Comparison groups were analyzed to

[6]There were an estimated 200 *unmatched* test results from both the Treatment and Comparison groups that could not be identified over the 3-year period (e.g. missing date of birth [used in matching] or dates of birth that could not be matched; pretests, but no posttests, etc.).

TABLE I: Number of courses and students (with matched pre- and posttests) involved in the *Weather in a Tank* project over 3 y.

| Groups | | Courses and Students | | | | Total No. of Classes | Total No. of Students |
|---|---|---|---|---|---|---|---|
| | | No. of Intro Courses | No. of Students in Intro Courses | No. of Intermed/ Advanced Courses | No. of Students in Intermed/ Advanced Courses | | |
| Treatment | Lab | 0 | 0 | 5 | 44 | 5 | 44 |
| | Lecture | 7 | 240 | 14 | 174 | 21 | 414 |
| Comparison | Lecture (only) | 7 | 345 | 6 | 111 | 13 | 456 |

account for group differences. These results are described later in this article.

## RESULTS OF THE EVALUATION

### Formative Evaluation

The information collected and analyzed from instructor logs, collaborator meetings, surveys, and site visits provided evidence and examples of instructors' use of experiments and the overall reactions of teachers and students to the project. The results of these analyses are described in this section.

### Instructors' Use of the Experiments

The instructors were free to use any of the experiments fully supported on the Web site in their courses, but they generally used those experiments that addressed either meteorology or oceanic phenomena, depending on the content of their courses and their own specialty (instructors in this study were more often atmospheric scientists than oceanographers). Seven of the experiments supported on the Web site address atmospheres and oceans, five address only oceans, and two address only the atmosphere (see Fig. 7). Instructors reported using four experiments frequently: "Dye Stirring," "Fronts," "General Circulation," and "Ekman Layers." The popularity of these experiments lies in their ease of use, relevancy to the level and content of the courses involved in the project, their ability to isolate a single phenomenon, and their stunning aesthetic value and visual effects. Usage also reflects the fact that four of the experiments used less often and listed to the right on Figure 7 only became available toward the end of the project. Indeed, as the project progressed, ocean-related experiments were increasingly used to illustrate the close connection between atmospheric and oceanic dynamics in classes.

One of the great success stories of the project was the innovation brought to the project by the students themselves. For example, students at the University of Massachusetts–Dartmouth suggested a clever use of computer fans in the "Ekman Pumping and Suction" experiment (see Beesley et al., 2008). The experiment is easy to set up, works every time and proved very popular, illustrating patterns of upwelling and downwelling induced by the action of the wind. This is a fundamental concept of ocean dynamics with great relevance to environmental issues, such as "The Great Pacific Garbage Patch" (Silverman, 2007, as reported in Beesley et al., 2008).

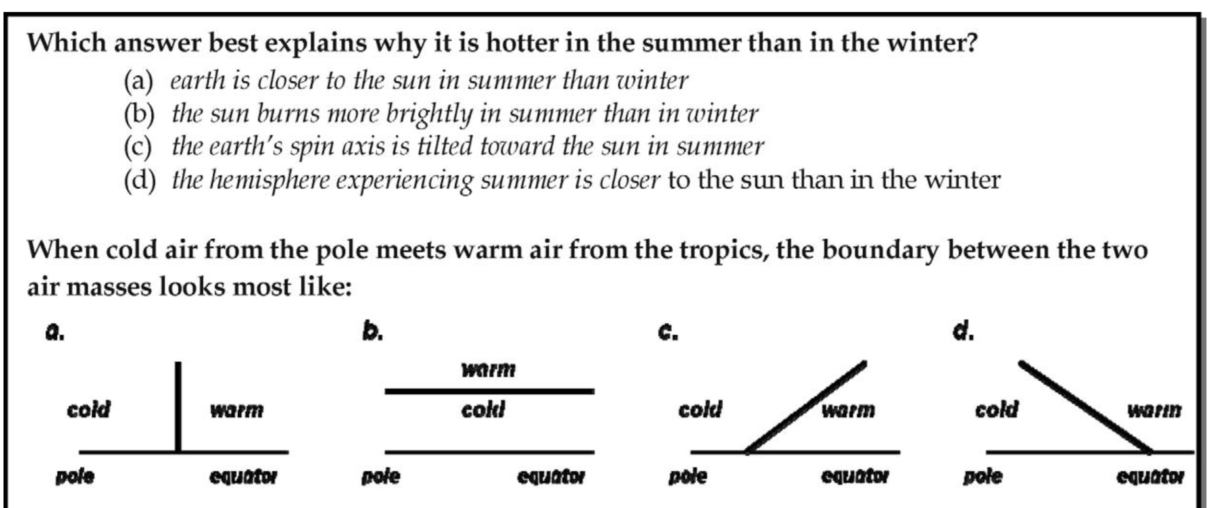


FIGURE 6: Two sample questions from the pre- and posttest. The full test is a 27-item, multiple-choice test covering general content related to atmospheres, oceans, and climate sciences, such as the importance of the earth's rotation on atmospheric circulation, the underlying cause of seasons, and reasons for typical wind speeds. This test was designed specifically for the *Weather in a Tank* project, but with an eye for wider, more general use. It is available on request from Lodovica Illari.

TABLE II: Subgroups represented in the Treatment and Comparison groups (2007–2009).

| Subgroups | Introductory Courses | | Intermediate/Advanced Courses | | |
|---|---|---|---|---|---|
| | Comparison (n=345) | Treatment (n=240) | Comparison (n=111) | Treatment (n=218) | Treatment (lab) (n=44) |
| Male | 33% | 57% | 71% | 71% | 70% |
| SAT/ACT Math | 562 | 630 | 602 | 606 | 626 |
| SAT/ACT Verbal | 562 | 613 | 565 | 561 | 597 |
| Freshman | 44% | 11% | 2% | 1% | 2% |
| Sci & Sci. Ed. Majors | 21% | 61% | 91% | 97% | 98% |

## Instructor Perceptions of the Value of the Experiments in Teaching

Despite any challenges collaborators faced in implementing the experiments, such as inadequate lighting for display and the occasional need to refit equipment, instructors overwhelmingly reported that the benefits were well worth the effort they put into the planning, set up, and research prior to using the experiments. From the instructors' perspective, the level of student engagement and interest in the experiments had a direct impact on students' ability to grasp difficult concepts, visualize outcomes, and engage in further independent inquiry. Many commented that the experiments contributed to a livelier class discussion and more interaction between themselves and the students, interaction that is often lacking in a lecture format classroom. The following comments from instructors illustrate the point that experiments and curricular materials had an impact on student learning and the quality of the teaching experience:

- **Dye Stirring,** in which beautiful interleaving patterns are created by stirring colored dyes into a rotating fluid: *Lots of interest from this group of more advanced students. Half of them had taken atmospheric circulation classes at an advanced level, but still were very surprised by the dye stirring experiment. I think [that before the experiment] they hadn't really grasped the ideas of solid*

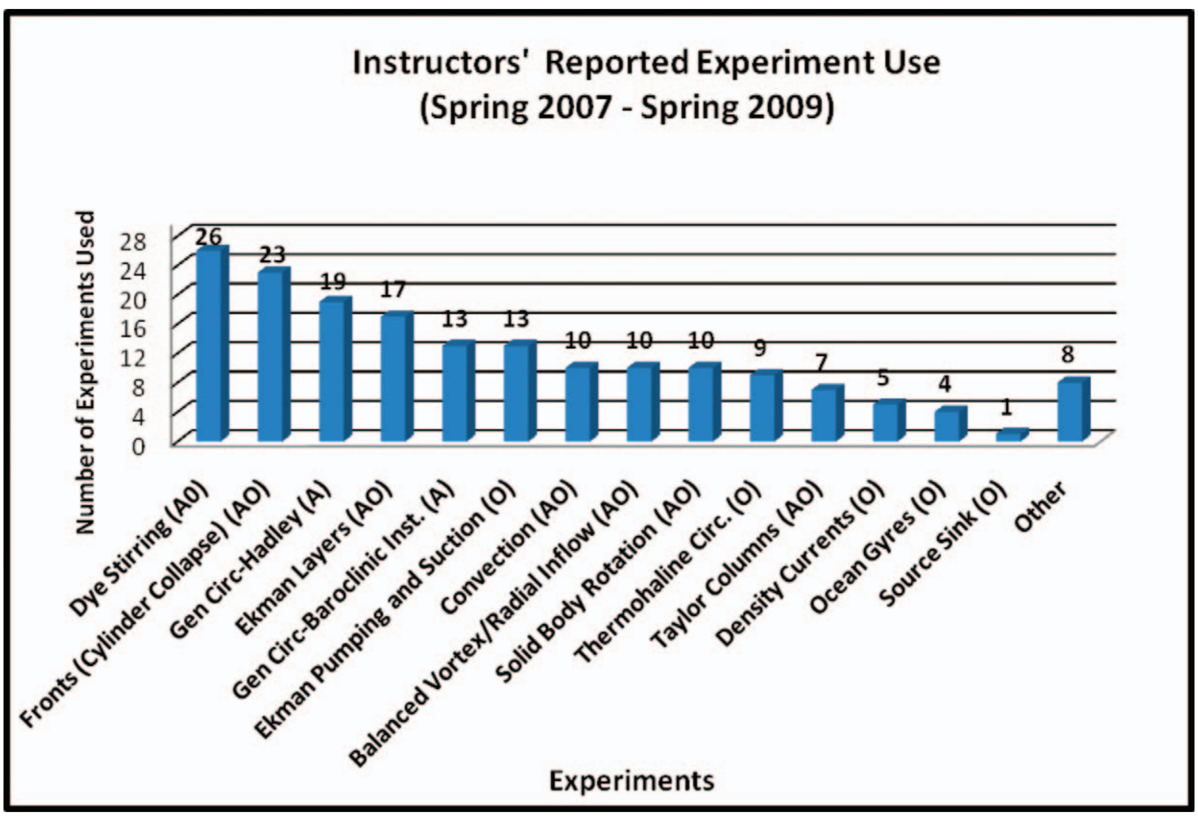


FIGURE 7: Frequency of experiment use by instructors, spring 2007–Spring 2009. "A" indicates experiments used in atmosphere classes; "O" represents classes in oceanography, and "AO" indicates atmosphere and/or oceanography courses. A full description of these and other experiments can be found on the project website.

**TABLE III: Raw score pre- and posttest performance of study groups. For each group a standard deviation and standard error of the mean is calculated for the 27-item test. Gain is the difference between post- and pretest scores.**

| Level | Group | Pre-Test | | | Post-Test | | | N | Gain |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | SD | SE | Mean | SD | SE | | |
| Introductory | Comparison/Lecture | 10.79 | 3.07 | 0.17 | 12.30 | 3.64 | 0.20 | 345 | 1.51 |
| | Treatment/Lecture | 14.45 | 4.30 | 0.28 | 16.37 | 4.88 | 0.31 | 240 | 1.92 |
| Intermediate/Advanced | Comparison/Lecture | 17.06 | 3.38 | 0.32 | 17.86 | 3.38 | 0.32 | 111 | 0.80 |
| | Treatment/Lecture | 17.18 | 3.71 | 0.28 | 18.46 | 3.48 | 0.26 | 174 | 1.28 |
| | Treatment/Lab | 15.82 | 4.05 | 0.61 | 18.59 | 3.44 | 0.52 | 44 | 2.77 |

*body rotation and the balance of forces in the rotating frame.*

- **Taylor Columns,** in which the rigidity imparted to a fluid by rotation is investigated by studying flow over a submerged object**:** *Prior to the demonstration, students were introduced to the concept of Taylor column and ''stiffness'' of rotating flow. Even with this prior knowledge, most students were impressed with what they observed. Their first reaction was how pretty the pattern was, and then somewhat surprised that the rotating fluid was indeed stiff! It may be my wishful thinking, but I believe that this first experiment helped some of the students appreciate the theory.*
- **Cylinder Collapse,** in which fronts are created in a rotating fluid by bringing together two bodies of water of differing densities**:** *To describe the separate effects of rotation and density difference, the students were asked to make predictions. Quite a few were not clear about the effect of rotation on the shape of a front. The demonstration greatly helped them to visualize how fronts adjust to a cone shape under the effect of rotation.*

## Student Perceptions of the Value of the Experiments

Forty-two percent (42%) of the students enrolled in courses across the six universities during the third year of the project responded to the online evaluation survey. Like the instructors, the student reactions were overwhelming positive about the *Weather in a Tank* experiments and the project's impact on classroom climate and their own learning. The following sample of open-ended comments from students illustrates the point that experiments intensified their interest in the subject matter, deepened their understanding of phenomena, and motivated them to conduct further inquiry or apply mathematical formulas in solving conceptual problems.

- *The experiments clarified misconceptions; SEEING phenomena was far more real than ''proving'' with equations.*
- *I had a misconception regarding the balance of forces in a hurricane. The experiment, Balanced Vortex, and the following Rossby number computation from the experiment helped me understand it.*
- *In the case of the convection experiment, I felt that the experiment greatly enhanced my understanding of the concept. I had learned about convection in other classes, but I wasn't entirely clear how it applied in the atmosphere. The experiment was a clear enough and visual enough analog to clarify the concepts of convection and potential temperature.*
- *The experiments cleared up any issues with fronts and balanced flow. Being able to visualize these concepts in three dimensions as they occur was the largest contributing factor to my understanding.*
- *The experiments, Solid Body Rotation, Balanced Motion, and Fronts helped me understand the physics behind the mathematical derivations that seemed quite obscure.*

Evidence of student content gain is presented in our summative evaluation, which we describe now.

## Summative Evaluation

As discussed in details below, results of the MIT *Weather in a Tank* pre- and posttest analysis of undergraduate student data after three years of experimentation and five iterations of the project (spring and fall, 2007; spring and fall of 2008; and spring 2009) indicate that the experiments had a positive impact on student learning (see Table III). As mentioned earlier, over the course of the three years, 915 matched pre/posttest scores were collected and analyzed from undergraduate students, 458 and 456 respectively in the Treatment and Comparison groups. In our analysis we attempted to address the following questions:

- What are the effects of *Weather in a Tank* experiments on undergraduate student achievement gains in knowledge of meteorology and climatology concepts?
- Are the *Weather in a Tank* experiments more effective for some subgroups of students than others (e.g., science vs. nonscience major, students in introductory or advanced courses)?

## Instrument Characteristics

All matched pre- and posttest results were analyzed to determine item characteristics. For each of the 27 test items both difficulty and discrimination are calculated. Difficulty refers to the proportion of students who answered the question correctly. Item difficulty ranged from a low of 0.10 to a high of 0.88 with a mean of 0.56. Item discrimination is a measure of the degree to which success on a particular item correlates with success on the overall test. Also known as the point-biserial (Sutton, 1977; Hopkins et al., 1990), this measure can range from −1.00 to +1.00, with higher values considered an indicator of item quality. For our test, items range from 0.09 to 0.59, with a mean value of 0.39.

**TABLE IV: Results from the general linear model for different groups. These two models predict student posttest scores for each of the five groups while controlling for initial pretest score, different student majors, gender, math and verbal SAT or ACT scores, and student year in college.**

| Variables | Degrees of Freedom | Introductory Model Probability | Intermediate/Advanced Model Probability |
|---|---|---|---|
| Constant | 1 | ≤0.0001 | ≤0.0001 |
| Pre-Test Score | 1 | ≤0.0001 | ≤0.0001 |
| Major | 4 Intro, 6 Int/Adv | 0.0001 | 0.0110 |
| Gender | 1 | 0.0190 | 0.0054 |
| Math SAT/ACT | 1 | 0.0077 | 0.0147 |
| Verbal SAT/ACT | 1 | 0.0482 | 0.0472 |
| Year in College | 4 | 0.1863 | 0.0009 |
| Group | 1 Intro, 2 Int/Adv | 0.0472 | 0.0055 |
| Total | $R^2$ | 17.9% | 41.3% |

Items of moderate difficulty generally have the highest discrimination, while those that are very difficult or very easy have lower discriminations.[7] Fig. 8[8] shows the difficulty and discrimination of each of the items in the test. Items with the highest discrimination can be thought of as the best for splitting the subjects into two groups: those with low knowledge and those with high knowledge. For example, item 30 has a discrimination of 0.59; it alone accounts for 35% ($0.59^2$) of the variance in test scores. Item 30 and 31, taken together, explain 50% of the variance in total test scores. Hence, these two items measure a kind of knowledge that probably underlies or at least is a prerequisite for doing well on the entire test. The items with the lowest discrimination (i.e., 9, 12, 29) have the smallest correlation with the total test score and are on the extremes of item difficulty.

An overall measure of test reliability is the Kuder-Richardson–20 statistic (KR–20), which measures the internal consistency or reliability of a test (Kuder and Richardson, 1937; Anderson et al., 2002). KR–20 tells how well individual items correlate with the overall test score on a scale from 0 to 1. A high KR–20 is characteristic of a test with high internal consistency, in that all items are aligned with a common underlying construct. The KR–20 for this test is 0.77, reasonable for a specialized instrument designed to measure a variety of concepts.

## Student Learning Outcomes

Using this set of items as both a pre- and posttest, one can compare performance of the three Treatment groups and two Comparison groups in the study. For each, a mean test score is calculated, along with the standard deviation and standard error of the mean for both pre- and posttests (see Table III). Standard deviations for each test administration show the range of student scores. The standard deviation of the pretest score is also used in educational research as a standard by which the size of any gain observed can be compared between studies. Standard errors for each mean indicate the confidence range of each mean and are shown in Table III. The largest absolute gains were in the lab group. The smallest absolute gains appear in the intermediate/advanced Comparison group. The tendency overall was for the Treatment groups in the introductory and intermediate/advanced courses to gain more than students in the Comparison groups in both conditions.

One potential problem with simply reporting pre-and posttest scores in this simple fashion is that that our study does not employ an experimental methodology in which subjects are assigned randomly to each group. As discussed earlier, such randomized studies are relatively rare in education. Instead, methods are commonly used that help to account for any substantive differences between treatment and comparison groups. Self-selection by students for inclusion in courses is evident in our data, particularly in their pretest performance. While pretest mean scores for intermediate/advanced groups are very similar for Comparison (lecture) and Treatment (lecture) groups, 17.06 and 17.18 respectively; pretest mean scores are quite different for introductory groups (10.79 for Comparison, and 14.45 for Treatment). This indicates that there are substantial differences between students in the Comparison and Treatment groups in initial knowledge. For this reason it is wise to examine the systematic differences between the backgrounds of students in each group and then to account for these variations in a general linear model. Differences in student background, by group, have been summarized earlier in Table II and include gender, standardized test scores, year in college, and choice of major, in addition to pretest score.

We constructed two general linear models to account for differences in student background. General linear models are similar to regression models, but can also include categorical, as well as continuous independent variables, to predict a single dependent variable. Such models are the mainstay of social science and educational research. We use this model to examine the relationship between student background, pretest score, inclusion in Comparison or

[7] The relationship of difficulty and discrimination is simply a product of the mathematical computation of each. To be a "perfect" discriminator, the top half of students answer correctly and the bottom half of students answer incorrectly.

[8] Item difficulty in this figure is more accurately defined as proportion of students answering correctly.

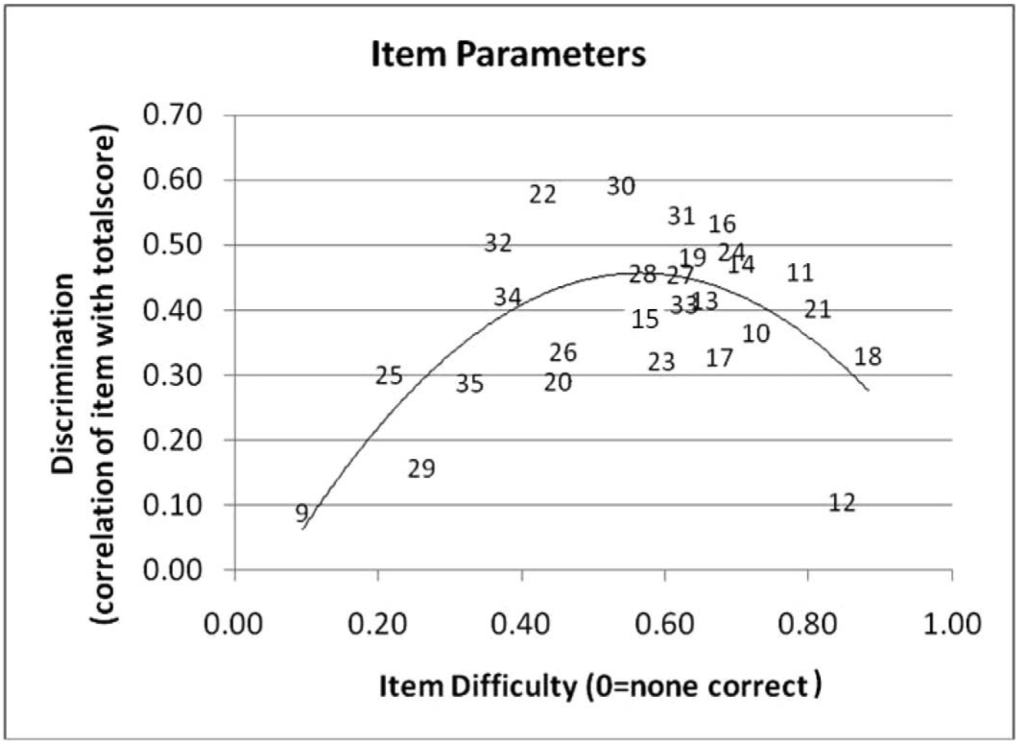


**FIGURE 8: Item difficulty vs. discrimination for the combined data set. Note that the item numbers (running from 9 to 35) correspond to the placement of items after the demographic items. The test displays a range of difficulties and corresponding discriminations.**

Treatment group, and posttest score. One model predicts gains for the introductory courses and a second for intermediate/advanced courses. One can think of these models as a post hoc method for making the compared groups as similar as possible and then comparing the results of different exposures to *Weather in a Tank* usage. While not a substitute for randomized assignment to groups, this process can be thought of as accounting for alternative explanations of differences in gain, based on the independent variables included. The probability calculated is that of randomly selected data points for each independent variable (instead of measured data) having a similar effect on posttest scores. A probability of $p \leq 0.05$ is considered acceptable for samples of this size.

In the introductory model, pretest scores, choice of major, gender, and SAT/ACT score are all significant predictors of student gains, while year in college is not (Table IV). The difference in groups, whether students are in the introductory Comparison or Treatment classes, is significant. This model explains 17.9% of the variance in student gain. This means that 82.1% of the variation in student posttest scores is not explained by the variables included. For the intermediate/advanced model, all of the background variables are significant, including year in college. Differences between groups are significant. This model explains 41.3% of the variance in posttest scores.

The modeled gains for each group are shown in Table V and graphed in Figure 9. These represent the gains expected after accounting for differences between groups at each level. It can be interpreted as if the groups were similar in pretest scores, distribution of majors, gender, SAT/ACT scores and year in college, but different in the use of *Weather in a Tank* demonstrations or labs. The gains observed in educational experiments are commonly measured in units of standard deviation of the pretest mean, known as effect sizes (ES; see Guzzetti et al., 1992; Cohen, 1998). These allow comparison to the results of other educational experiments. Gains of 0.25 or less are generally considered small effects and 0.75 and greater as large. Effect sizes are seen to be small to moderate in size.

The results in which we are most interested are the differences in gains for each group after accounting for students' background. Significant differences between group gains are indicative of classrooms using *Weather in a Tank* demonstrations or labs being more effective than those that did not use them. The statistical test for these differences used is the Scheffe Post Hoc Test. The differences between gains are listed in Table VI. They are consistent with the appearance of the error bars in Figure 9. The effect sizes for two introductory groups are seen to be significant (at the $p \leq 0.05$ level) since the error bars do not overlap (the Sheffe Test has $p = 0.05$). Differences between the effect sizes of Comparison and Treatment groups at the intermediate/advanced level are not significant (at the $p \leq 0.05$ level) since the error bars do overlap (the Sheffe Test has $p = 0.60$). Differences between the effect sizes for the two intermediate/advanced lecture groups and the lab group are significant.

Student posttest scores for all groups increased when compared to pretest scores. The gains for introductory level courses, which were all lecture-based, were larger than the gains for the intermediate/advanced lecture-based courses.

TABLE V: Modeled gains for each group. The expected gains for each group at each level are calculated from the general linear model, which controls for differences in student characteristics. These modeled values are somewhat different from those in Table III. Effect sizes, gain in units of standard deviation, are also reported.

| Level | Groups | Gain | SE | ES | SE ES |
|---|---|---|---|---|---|
| Introductory | Comparison (Lecture) | 1.45 | 0.18 | 0.39 | 0.05 |
| | Treatment (Lecture) | 2.17 | 0.18 | 0.59 | 0.05 |
| Intermediate/Advanced | Comparison (Lecture) | 0.88 | 0.23 | 0.24 | 0.06 |
| | Treatment (Lecture) | 1.22 | 0.21 | 0.33 | 0.06 |
| | Treatment (Lab) | 2.46 | 0.30 | 0.68 | 0.08 |

Gains for the intermediate/advanced lab-based course were the highest. Differences between Comparison and Treatment groups are only significant (at the $p \leq 0.05$ level) between introductory Comparison and Treatment groups, and between the Treatment lab group and other intermediate/advanced lecture-based courses. There is no significant difference between the gains of Comparison and Treatment students in the intermediate/advanced classes that were only lecture-based. These results support the conclusion that the inclusion of *Weather in a Tank* experiments worked particularly well to increase learning for students in the introductory Treatment group and the intermediate/advanced Treatment lab group. The largest difference in gain was in the Treatment lab versus Comparison (lecture) and Treatment (lecture) groups.

In summary, our results suggest that classroom demonstrations using *Weather in Tank* curricular materials are particularly effective in introductory level courses. For intermediate/advanced level courses, students appear to have benefited more from conducting these experiments in a laboratory setting rather than from classroom demonstrations.

## CONCLUSIONS AND IMPLICATIONS FOR INSTRUCTION AND LEARNING

Through an extensive formative and summative evaluation of the *Weather in Tank* project over 3 y of the project we found that the experiments and curriculum had the following impact:

### Successful Introduction of New Experiments in Atmospheric and Oceanic Science Courses

One of the most important outcomes of the *Weather in a Tank* project has been the successful introduction of a set of proven experiments and demonstrations to a field that has traditionally been dominated by a lecture-based instructional platform, numerical models, descriptions, and Web-based simulations. As demonstrated through instructor logs,

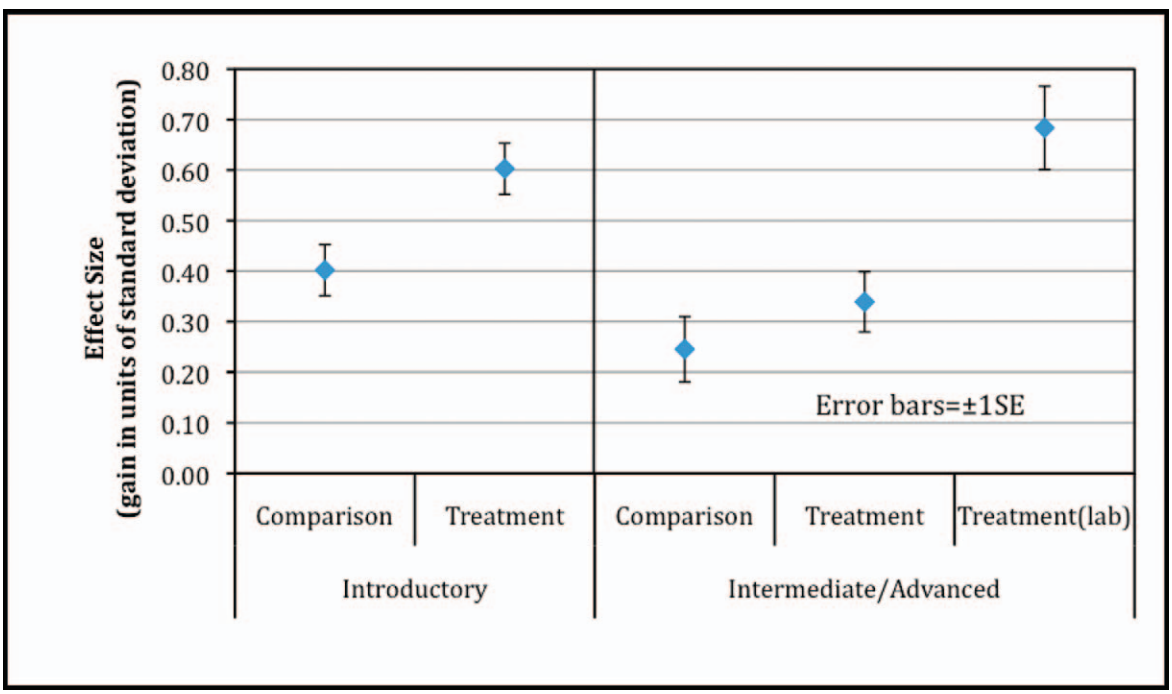


FIGURE 9: Gains for introductory and intermediate/advanced groups. Significant effect sizes are seen for all groups since each is more than 2 SE from 0.00. Treatment groups in the introductory and intermediate/advanced courses gain more than students in the Comparison groups: gains are significant for introductory groups and intermediate lab groups.

**TABLE VI: Sheffe tests for significant differences in gains. These tests support interpretation of significance from the overlap of error bars in Figure 9.**

| Level | Group | Difference in Gain | SE | Prob |
|---|---|---|---|---|
| Introductory | Treatment (Lecture)-Comparison (Lecture) | 0.72 | 0.36 | 0.05 |
| Intermediate/Advanced | Treatment (Lecture)-Comparison (Lecture) | 0.34 | 0.33 | 0.60 |
| | Treatment (Lab)-Comparison (Lecture) | 1.58 | 0.49 | 0.01 |
| | Treatment (Lab)-Treatment (Lecture) | 1.24 | 0.46 | 0.03 |

instructor feedback on yearly surveys, student surveys, and in personal conversations during site visits and at meetings, this kind of interactive curriculum offered instructors an opportunity to increase their repertoire of teaching strategies and fully explore the multiple concrete phenomena at play in the content of atmospheric and oceanic courses. As advocated by the National Research Council (2000), this kind of experimentation leads to changes in the classroom atmosphere where curiosity is fostered and students are engaged and motivated to study scientific concepts and theories.

### Expanded Student Cognitive Models

The *Weather in a Tank* curriculum makes use of strategies to explore students' initial conceptions before instruction allowing them to link conceptual understanding to observable events and build cognitive models. Using the *Weather in a Tank* experiments and curriculum allowed instructors to isolate variables of interest in a more controlled environment to help students connect their conceptual understanding to the physical world. As suggested by authors such as such as Sokoloff and Thornton (1997) and Roebber (2005), exploring variables in this controlled environment under the direction of the teacher provides a context for students to test out hypotheses and see evidence that either supports or refutes them. Given the importance of GFD in global climate change, the understanding facilitated by *Weather in a Tank* is critically important to today's undergraduates, assisting them in creating a two-way bridge between abstract thought and concrete, large-scale problems.

### Increased Student Learning Outcomes

Test scores can often be a very problematic measure of student learning, making it difficult to define the impact of the experiments separate from the instructor and their approach to implementation, instruction, and follow-up. What test scores do offer is important trend data that allows us to draw conclusions about student learning outcomes for those who were exposed to the experiments compared to those who received *no* instruction with the *Weather in a Tank* experiments. Encouragingly, we found that, over a wide range of teaching environments involving over 900 undergraduate students, in all cases the Treatment group *consistently scored higher* than the Comparison group. These statistics were significant in the following contexts: introductory courses (Treatment vs. Comparison), $p = 0.05$; intermediate/advanced courses (Treatment Lab vs. Comparison Lecture), $p = 0.01$; and intermediate/advanced courses (Treatment Lab vs. Treatment Lecture), $p = 0.03$. These findings add to the body of empirical evidence called for by the National Research Council (Fox and Hackerman, 2003; Nelson, et al., 2010) suggesting that experiments and laboratories do indeed aid students in conceptualizing phenomena and understanding content.

These results are based on a full implementation of the *Weather in a Tank* model in atmospheric and oceanic science courses that included training and monitoring of instructors, Web-based support for utilizing experiments, and consistent and robust testing of student outcomes.

Based on our experience we conclude with some general advice for those who would like to adapt this project for use in their classes. The benefits of the laboratory experiments are maximized when they are used as an integral part of the class. When instructors have an opportunity to use the demonstrations to explain phenomena that are already part of the course content and not a course add-on, instruction is more effective and students are more highly engaged. Instructors repeatedly stressed the fact that introducing experiments into the course requires background preparation and in most cases, rehearsal. To be effective and to maximize results with students, instructors need to feel comfortable with the rotating tank equipment, the experimental procedures and expected outcomes, and to be prepared to make clear connections between the experiments and the phenomena being studied in the course. Learning outcomes for students are enhanced when the instructors allow time for students to make predictions based on prior knowledge, fully observe (and, where feasible, assist with the experiment), discuss the demonstrated phenomena, demonstrate increased understanding of content, and engage in further inquiry or replicate or develop related experiments.

Finally, it is pleasing to report on one unexpected benefit of the *Weather in a Tank* project. The laboratory experiments were found to be invaluable in "outreach" activities such as presentations to visiting school groups, prospective students, open houses, or in informal learning venues such as science fairs and museums.

### Acknowledgments

We would like to thank the students and professors of the collaborating universities for their interest in and support of the *Weather in a Tank* project, and the National Science Foundation (NSF), whose support made this project possible.

## REFERENCES

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

Aron, R., Francek, M., Nelson, B. and Bisard, W. 1994. Atmospheric misconceptions. *The Science Teacher*, 61:30–33.

Barowy, W., and Lochhead, J. 1980. Abstract reasoning in rotational physics, Available at http://www.lesley.edu/faculty/wbarowy/pub/AbstractReasoning.pdf.

Beesley, D., Olejarz, J., Tandon, A., and Marshall, J. 2008. A laboratory demonstration of Coriolis effects on wind-driven ocean circulation. *Oceanography*, 21:71–76.

Cohen, J. 1998. Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale. NJ: Lawrence Erlbaum Associates.

Crouch, C., Fagen, A., Callan, J. and Mazur, E. 2004. Classroom demonstrations: Learning tools or entertainment? *American Journal of Physics*, 72:835–838.

Dori, Y.J., and Belcher, J. 2005. How does technology-enabled active learning affect undergraduate students' understanding of scientific concepts? *The Journal of the Learning Sciences*, 14:243–279.

Fox, M.A., and Hackerman, N., eds. 2003. Evaluating and improving undergraduate teaching in science, technology, engineering, and mathematics. Washington, DC: National Research Council, Center for Education.

Gentner, D., Loewenstein, J., and Thompson, L. 2003. Learning and transfer: A general role for analogical reasoning. *Journal of Educational Psychology*, 95:393–405.

Gobert, J. 2005. The effects of different learning tasks on model-building in plate tectonics. *Journal of Geoscience Education*, 53:444–455.

Gunstone, R. 1971. Circular motion: Some pre-instruction alternative frameworks. *Research in Science Education*, 14:125–136.

Guzzetti, B., Snyder, T., and Glass, G. 1992. Promoting conceptual change in science: Can texts be used effectively? *Journal of Reading*, 35:642–649.

Gyllenhaal, E.D., and Perry, D.L. 1998. Doing something about the weather: Summative evaluation of Science Museum of Minnesota's Atmospheric Explorations computer interactives. *Current Trends in Audience Research and Evaluation,* 11:25–35.

Gynnild, V., Myrhaug, D., and Pettersen, B. 2007. Introducing innovative approaches to learning in fluid mechanics: A case study. *European Journal of Engineering Education*, 32:503–516.

Hammer, D. 1996. Misconceptions or p-prims: How may alternative perspectives of cognitive structure influence instructional perceptions and intentions? *Journal of the Learning Sciences*, 5:97–127.

Harrington, J. 2008. Misconceptions: Barriers to improved climate literacy. *Physical Geography*, 29:575–584.

Hay, K.E., Marlino, M., and Holschuh, D.R. 2000. The virtual exploratorium: Foundational research and theory on the integration of 5-D modeling and visualization in undergraduate geoscience education. *In* Fishman, B.J., and Samuel F. O'Connor-Divelbiss, S.F., eds., Proceedings of the International Conference of the Learning Sciences: Facing the Challenges of Complex Real-world Settings. Ann Arbor: University of Michigan, p. 214–217.

Henriques, L. 2002. Children's ideas about weather: A review of the literature. *School Science and Mathematics*, 102:202–215.

Hestenes, D., Wells, M., and Swackhamer, G. 1992. Force concept inventory. *Physics Teacher,* 30:141–151.

Hofstein, A., and Lunetta, V.N. 2003. The laboratory in science education: Foundations for the twenty-first century. *Science Education,* 88:28–54.

Hopkins, K., Stanley, J., and Hopkins, B. 1990. Educational and psychological measurement and evaluation, 7th ed. Englewood Cliffs, NJ: Prentice Hall.

Illari, L., Marshall, J., Bannon, P., Botella, J., Clark, R., Haine, T., Kumar, A., Lee, S., Mackin, K.J., McKinley, G.A., Morgan, M., Najjar, R., Sikora, T., and Tandon, A. 2009. Weather in a Tank: Exploiting laboratory experiments in the teaching of meteorology, oceanography, and climate. *Bulletin of the American Meteorological Society*, 90:1619–1632.

Illari, L. 2001. MIT Synoptic Laboratory. Available online at http://paoc.mit.edu/synoptic.

Kahl, J.D.W. 2008. Reflections on a large-lecture, introductory meteorology course: Goals, assessment, and opportunities for improvement. *Bulletin of the American Meteorological Society*, 89:1029–1034.

Ki-moon, B. June 8, 2007. Secretary-General's opening statement at the G8 Press Conference, Heiligendamm, Germany.

Knox, J.A. 2008. Recent and future trends in U.S. undergraduate meteorology enrollments, degree recipients, and employment opportunities. *Bulletin of the American Meteorological Society*, 89:873–883.

Kuder, G.F., and Richardson, M.W. 1937. The theory of the estimation of test reliability, *Psychometrica,* 2:151–160.

Libarkin, J., and Anderson, S. 2005. Assessment of learning in entry-level geoscience courses: Results from the Geoscience Concept Inventory. *Journal of Geoscience Education*, 53:394–401.

Mackin, K. 2008. Evaluation report, year 1: Weather in a Tank: Employing laboratory experiments in the teaching of meteorology, oceanography, and climate. Accessed July 2011 from http://paoc.mit.edu/labguide/assess.html.

Marshall, J., and Plumb, R.A. 2008. Atmosphere, ocean, and climate dynamics: An introductory text. International Geophysical Series, Vol. 93. Elsevier Academic Press.

National Research Council (NRC). 1996. National science education standards. Washington, D.C.: National Academy Press.

National Research Council (NRC). 2000. How people learn: Brain, mind, experience, and school. *In* Bransford, J.D., Brown, A.L., Cocking, R.R., eds., Committee on Developments in Science and Learning. Washington, DC: National Academy Press.

Nelson, B., Aron, R., and Francek, M. 1992. Clarification of selected misconceptions in physical geography. *Journal of Geography*, 91:76–80.

Nelson, K.G., Huysken, K., and Kilibarda, Z. 2010. Assessing the impact of geoscience laboratories on student learning: Who benefits from introductory labs? *Journal of Geoscience Education*, 58:43–50.

Pandya, R.E., Domenico, B., Molina, M., and Murray, D. 2004. 11th AMS Education Symposium. *Bulletin of the American Meteorological Society*, 85:425–430.

Persson, A. 1998. How do we understand the Coriolis force? *Bulletin of the American Meteorological Society*, 79:1373–1385.

Piaget, J., Tomilson, J., and Tomilson, A. 2007. The child's conception of the world: A 20th century classic of child psychology. Lanham, MD: Rowman & Littlefield.

Roebber, P.J. 2005. Bridging the gap between theory and applications: An inquiry into atmospheric science teaching. *Bulletin of the American Meteorological Society*, 86:507–517.

Rollins, M., Denton, J., and Janke, D. 1983. Attainment of selected earth science concepts by Texas high school seniors. *Journal of Educational Research,* 77:81–88.

Roth, W., McRobbie, C., Lucas, K., and Boutonne, S. 1997. Why may students fail to learn from demonstrations? A social practice perspective on learning in physics. *Journal of Research in Science Teaching,* 34:509–533.

Silverman, J. 2007. Why is the world's biggest landfill in the Pacific Ocean? Retrieved from Web site, *How Stuff Works* on September 19, 2007 at http://science.howstuffworks.com/great-pacific-garbage-patch.htm.

Singer, S.R., Hilton, M.L., and Schweingruber, H.A., eds. 2005. Committee on high school science laboratories: Role and vision. Washington, DC: National Research Council.

Sokoloff, D.R., and Thornton, R.K. 1997. Using interactive lecture demonstrations to create an active learning environment. *The Physics Teacher*, 35:340–346.

Stepans, J. 1994. Targeting students' science misconceptions. Riverview, FL: Idea Factory, Inc.

Sutton, R. 1977. Diagnosing prior knowledge levels of first-year university physics students. *Innovations in Education & Training International*, 14:92–95.

Thornton, R.K. 1996. Learning physics concepts in the introductory course, microcomputer-based labs and interactive lecture demonstrations. In Proceedings Conference on the Introductory Physics Course. New York: John Wiley & Sons, p. 69–85.

Turner, J.S. 2000. Development of geophysical fluid dynamics: The influence of laboratory experiments. *Applied Mechanics Reviews*, 53:11–23.