Impact of the Science and Technology for Children Curriculum in the Oshkosh Area School District

Contact Author:

Mark Joseph Lattery
Department of Physics and Astronomy
College of Letters and Science
800 Algoma Boulevard
University of Wisconsin Oshkosh
Oshkosh, Wisconsin 54901

Phone: (920) 424-7105
FAX: (920) 424-1328
Email: lattery@uwosh.edu

Co-Authors:

Dr. John Lemberger
Department of Curriculum and Instruction
College of Education and Human Services
800 Algoma Boulevard
University of Wisconsin Oshkosh
Oshkosh, Wisconsin 54901

Phone: (920) 424-7244
Email: jlemberg@uwosh.edu

Dr. Barbara Herzog
Assistant Superintendent of Instruction
Central Administration
Oshkosh Area School District
215 South Eagle Street
Oshkosh, Wisconsin 54902

Phone: (920) 424-0296
Email: barher@oshkosh.k12.wi.us

Manuscript No.: 2864s

Impact of the Science and Technology for Children Curriculum in the Oshkosh Area School District

Mark Lattery
University of Wisconsin Oshkosh

Barbara Herzog
Oshkosh Area School District

John Lemberger
University of Wisconsin Oshkosh

This study examines the instructional impact of National Science Resources Center's Science & Technology for Children curriculum in the Oshkosh Area School District. Specifically, the instructional effectiveness of four physical-science units for grades 1-4 were investigated. Students were pre- and posttested using a multiple-choice exam containing items adapted from the Third International Mathematics and Science Study, National Association of Educational Progress, TerraNova, and other widely recognized sources. Results were compared with existing instructional materials. Extensive resources for data interpretation include audio taped lessons, classroom observations, interviews with students and teachers, student-attitude surveys, and observations of teacher training. Results suggest that the adoption of this curriculum among experienced teachers in the district will provide little or no immediate gains on student achievement, and potentially a slight decrease in student attitudes toward science.

Background

School districts across the United States are under increasing pressure to perform well on state, national, and international standardized exams. At the same time, education researchers have all but abandoned pencil-and-paper methods as a means to assess hands-on science learning (Resnick & Resnick, 1992; Wiggins, 1989). The position of this paper is that traditional pencil-and-paper tests (including multiple-choice items and short answer questions) are useful for assessing student achievement on content objectives and even process skills (Feinberg, 1990; Meng & Doran, 1993). Evidence that hands-on science programs lead to better performance on traditional tests (for which these programs are not tuned) strengthens the case for science education reform when combined with other authentic measures of learning, such as performance assessment (Gabel, 1994, chapter 14, and references therein).

The purpose of this study is to evaluate the National Science Resource Center's (1997) *Science and Technology for Children* (*STC*) curriculum in the Oshkosh Area School District. While extensive support materials and teacher training for the *STC* curriculum have existed in the region for over 10 years, studies of learning outcomes have only recently emerged (Saint Norbert College Survey Center, 2001). A study of *STC* effectiveness was also conducted in New England (Lesley College Program Evaluation and Research Group, 1992). This study provides a qualitative analysis only.

To estimate the potential impact of the *STC* curriculum in the district, we set out to answer the following questions: What gains in student achievement on standardized tests can be expected in the first year or so of implementation? What types of learning can be expected? How might this learning vary from classroom to classroom? What effect will *STC* materials have on student attitudes toward science?

Study Design

School Environment

This study was conducted in the Oshkosh Area School District (OASD) during the fall of 1999. The OASD is known for its skilled and experienced teaching staff: 80% of teachers have 5+ years of experience; 35%+ hold graduate degrees; approximately 40% received specialized training in modern science pedagogy; and over 50% use hands-on science activities as the dominant form of instruction.

Research Participants

The student sample (grades 1-4, *N*=325) represented seven public elementary schools. Class sizes ranged from 15 to 25 students. Most students came from middle-income homes, although the full range of incomes was represented.

Eight experienced teachers (research instructors) were selected to participate in this study (Appendix A). In addition to significant graduate work, six of eight instructors received training in inquiry-based science teaching through the Einstein Project (www.itol.com/einstein) and/or the Activities That Integrate Math and Science (AIMS) program (www.aimsedu.org).

Content Focus

Four physical-science STC units were evaluated in this study: Weather (Grade 1), Balance and Weighing (Grade 2), Motion and Design (Grade 3), and Electric Circuits (Grade 4). Units were selected on the basis of positive field reports and the content expertise of the research group.

Teacher Training

Prior to the study (summer of 1999), research instructors attended a 1-day workshop through the Einstein Project of Green Bay, Wisconsin. The workshop provided the same type and level of instruction that would be received by OASD teachers during the first year of *STC* implementation. Research instructors each received basic instruction in the *STC* unit at their grade level. Teacher observations of this training are given in Lattery, Lemberger, & Herzog (2001, Appendix 2).

Instruction

During the school year, instructors taught science twice per day: once using the *STC* materials (treatment group), and again using existing OASD materials (control group). Both treatment and control groups received about 50 minutes of instruction per day for 8 weeks. To control for individual student differences, students were pre- and posttested and learning gains were compared. To estimate possible class-personality effects, two teachers at each grade level participated in the study.

Curricula

The *STC* units have strong intentions toward the inquiry approach to science instruction (NSRC, 1997). Lesson activities require students to work in groups to collect and analyze data. *STC* objectives are to (a) foster positive attitudes toward science, (b) introduce science concepts, and (c) expose students directly to the scientific thinking process through hands-on activities.

Prior to the school year, research instructors developed 8 weeks of control unit instruction in grade-level pairs. All development was based on the *OASD Curriculum Guide* (OASD, 1998) and the results of a district-wide teaching survey of current science teaching practice (Lattery et al., 2001, Research Documents). The control units represent a mixture of traditional and hands-on approaches based on the Merrill Science textbook; these units cover the same general objectives as the *STC* units. For a complete description of the control unit materials, see (Lattery, et. al, 2001, Appendix 3).

Research Instruments

Quantitative

Experimental and control groups were pre- and posttested using a multiple-choice test (Lattery & Lemberger, 2001a). Resources for test construction included the Third International Mathematics and Science Study (Schmit, McKnight, & Raizen, 1997), National Association of Educational Progress (Blumberg, Epstein, MacDonald, & Mullis, 1986), TerraNova (CTB/McGraw-Hill, 1996), and other widely recognized sources. A content-specific test was used for each grade level. The selection of items was determined by the authors and guided by the National Science Education Standards document (National Research Council, 1996) without prior knowledge of the specific content of the *STC* units or the alternative control-group units.

To address the diverse concerns of local stakeholders, five main categories were used to study student-learning outcomes: (a) attitudes, (b) recall knowledge, (c) conceptual reasoning (e.g., concept discrimination), (d) experimental methods, and (e) mathematical application. Student achievement is based on performance in categories 2-5. Sample test items are given in Appendix B.

All items were validated locally by the authors. Validity studies were conducted with children in small-group interviews. Each interview was approximately 20-40 minutes in length. Test items were revised and re-piloted as many as five times before final adoption. As needed, items were adapted to enhance item validity. Test items were further validated through consultation with subject-matter experts, science education specialists, and elementary teachers. Several of these consultants are known internationally for their work in science education and assessment.

Test reliability was determined using the test-retest method. Students at each grade level took the test two times: once at the end of the unit and again several months later. Between the test administrations, students were exposed to the existing OASD curriculum (not the content of the unit). The Pearson product-moment correlation between scores ranged from 0.6 to 0.8, depending on the test (grade level). Differential developmental effects between test administrations make these estimates conservative. Implications for the statistical power of this study are discussed later in this article.

The research instructors helped administer the pretest. However, the instructors did not view the test prior to the development of the control group lessons. During instruction, the instructors were directed to implement the control group lessons as planned; i.e., with no midcourse corrections. Taped classroom instruction revealed little or no deviation from the control lesson plans.

Qualitative

Classroom instruction was monitored continuously with a portable microphone (clipped to the research instructor's lapel) and tape recorder. The purpose of tape recording is twofold: (a) to ensure that control and treatment groups received the same amount of instruction, and (b) to determine the relative quality of teacher-student dialogue. Instructional episodes are classified into following six categories: requests for students to (a) volunteer their own personal beliefs about the topic; (b) recite some aspect of scientific understanding (i.e., define a term or state a principle); (c) read about the scientific topic; (d) reason scientifically from facts or data; (e) participate in or learn about experimental design; and (f) apply mathematics. Both episode frequency and content were used to interpret student test results.

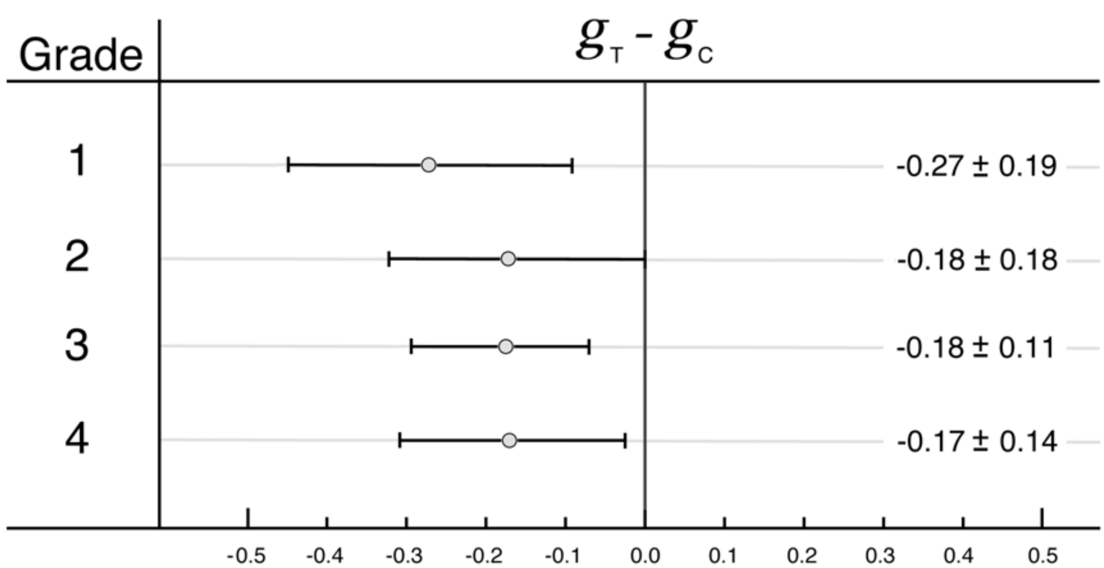
Findings

Improvements in student achievement and attitudes are quantified using a normalized learning gain (g), defined as the ratio of the student gain (score improvement) to maximum possible gain. To assess the relative effectiveness of the *STC* materials, gains for the treatment group (g_T) and the control group (g_C) are compared. The significance of results is evaluated using a z-test for two population means (α=0.90), and conducted by Monte-Carlo calculation to avoid assumptions of normality (Kanji, 1999).

Student Attitudes

Overall, student attitudes declined or stayed the same over the 8 weeks of instruction, with the most pronounced effect at Grade 1 (g_C = -0.18 ± 0.15, g_T = -0.45 ± 0.11). The gain differential (Δg = -0.27) for Grade 1 is represented by the upper data point in Figure 1. Note that the data point is left of the centerline (Δg = 0.0). This

Figure 1. *Student attitude results.*

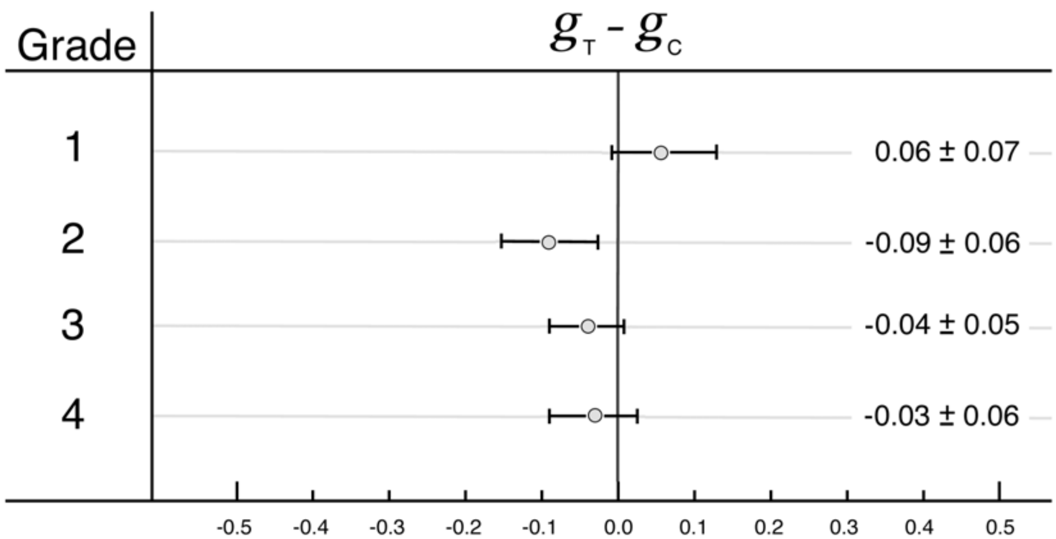


indicates that the control group outperformed the treatment group. The decline in student attitudes is systematically larger for the treatment group across all four grade levels.

Student Achievement

For student achievement, mild gains (*g* = 0.15-0.25) were observed for both control and treatment groups. These gains varied by instructor, classroom, and item category, with no simple trends. For example, Mrs. Birch's Grade 2 classrooms (control and

Figure 2. *Student achievement results.*



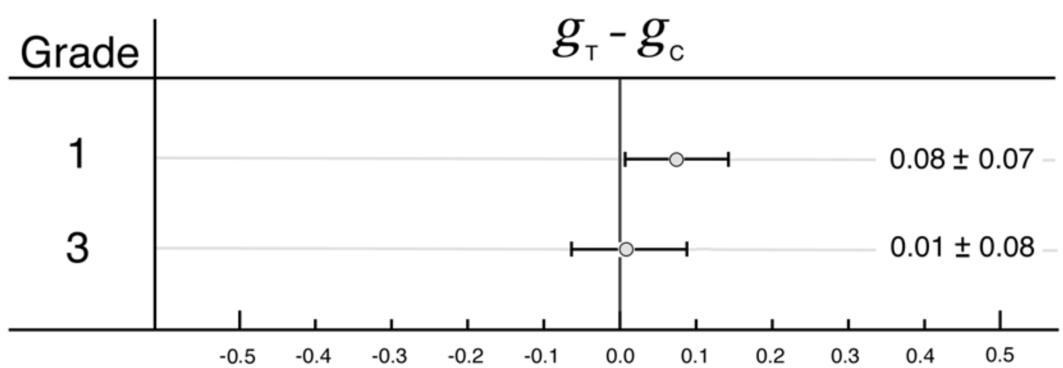
treatment) achieved impressive gains in all five learning-outcome categories (*g* = 0.3-0.8), while her grade-level partner obtained much smaller gains (*g* = -0.1-0.2).

In grades 1, 3, and 4, no statistically significant differences were observed between control and treatment groups (Figure 2). In Grade 2, the control group outperformed the treatment group ($z = 1.50$, $p < 0.10$). This result is driven by performance on recall knowledge, experimental methods, and mathematical-application.

Due to small classroom sizes ($N = 15$ to 25 students) and modest test reliability, the statistical power of this comparison is not great ($1-\beta \approx 0.3-0.6$). Nevertheless, the instruments described are sensitive to student-achievement differences of approximately $|\Delta g| \geq 0.1$. Certainly, smaller gains make the justification of a new curriculum difficult.

To establish a sufficient condition to reject the null hypothesis (no difference), additional test items, tailored specifically to the content and representations of the *STC* units, were appended to the Grade 1 and 3 tests (Lattery & Lemberger, 2001a). Even on these items, no significant differences between the control and treatment groups were observed (Figure 3).

Figure 3. *Student achievement results for items tailored to the STC units.*



Classroom Observations

Case studies for three research instructors (both control and treatment classrooms) are briefly summarized below. The different styles and strengths of these instructors serve to illustrate a range of possible instructional dynamics. Well over 200 teacher-student interactions (episodes) were analyzed.

In Mrs. Oak's Grade 1 classrooms, about one fourth of episodes involved listening or reading (Table 1); that is, both classroom were relatively "active." In the control classroom, a smaller fraction of episodes were dedicated to expressing beliefs (0.13 in the control group, and 0.24 in the treatment group) and a much larger fraction to reciting knowledge (0.31 in the control group, and 0.04 in the treatment group). However, the treatment group performed better than the control group on the recall portion of the test (g_T = 0.6 and g_C = 0.2). The fraction of episodes in other areas were comparable.

According to Mrs. Oak, the *STC* (treatment) unit is more instructionally effective than the control unit by virtue of its "hands-on activities." She praised the *STC* (treatment) units for requiring "little listening and sitting" (the data suggests her classrooms spent the same fraction of time listening), and she criticized the control group instruction for being overly committed to direct instruction. However, the control group performed better than the treatment group on reasoning items (g_T = 0.1 and g_C = 0.4).

In Mrs. Pine's Grade 3 classrooms, the fraction of episodes in several categories were comparable. However, the relative number of episodes dedicated to reading and listening was much greater for the control group (0.43 in the control group, and 0.17 in the treatment group). Experimental design was emphasized in the treatment units, but not at all for the control units. Mrs. Pine stated that the control unit did not get children out of their seats to move around enough. However, she believed that the control unit did a better job in the area of scientific reasoning. This claim is supported by student performance on the experimental methods section on the test, in which the gain difference (in favor of the control unit) was about 0.20.

According to Mrs. Pine, students were unable to learn through inquiry alone, and relied upon her for basic information. However, Mrs. Pine noted that the *STC* unit opened up more science questions for the children. She stated that the treatment unit would be more appropriate for Grade 5, though the first several lessons (building vehicles) were workable.

In Mrs. Maple's Grade 4 classrooms, the fraction of episodes involving expressing beliefs and designing experiments was higher for the treatment group, but episodes requiring students to read, listen, and reason scientifically were more frequent for the control group. Other areas were comparable. In the control classroom, the basic instructional pattern was to read about the topic and then perform an experiment—a strongly deductive approach. Mrs. Maple stated that the *STC* unit did not provide enough opportunities for students to read, and when opportunities were provided, they were not written at the level of the children (Grade 4). Mrs. Maple felt the control group integrated reading well and included "more meat, more facts to learn." She believed the *STC* unit could do a better job of showing children how the topic fits into a "bigger picture." Mrs. Maple's general impression was that the control group unit was more effective. It is difficult to evaluate this claim using test results, given the statistics of her classroom alone; however, student achievement gains were: $g_T = 0.1 \pm 0.3$ and $g_C = 0.2 \pm 0.3$.

Table 1.
Fraction of Teacher Requests During Control (C) and Treatment (T) Group Instruction (Total Number of Episodes in Parentheses).

	Mrs. Oak Grade 1		**Mrs. Pine Grade 3**		**Mrs. Maple Grade 4**	
Request type	**C (39)**	**T (50)**	**C (51)**	**T (29)**	**C (27)**	**T (21)**
Express beliefs	.13	.24	.18	.24	.11	.24
Read/listen	.23	.26	.43	.17	.33	.19
Recite knowledge	.31	.04	.12	.10	.26	.33
Reason scientifically	.23	.32	.28	.24	.26	.10
Design an experiment	0	0	0	.21	0	.14
Apply mathematics	.10	.14	0	.03	.04	0

In summary, several notable differences were observed between the experimental groups:

1. Treatment group instruction elicited initial student beliefs more often.
2. Control group instruction placed more emphasis on reading and listening skills.
3. Treatment groups were exposed to more instruction in experimental design.
4. Control group instruction placed a greater emphasis on recall knowledge (though not evident from the case studies described).

Emphases on scientific reasoning and mathematical applications varied by subject (grade level) and instructor, with no simple trends.

Discussion

Neither the *STC* units nor the existing OASD materials stood out clearly as superior to the other. This result has to two possible interpretations. First, there was no difference between the units, or the difference was buried within uncertainties of the measurement. Second, the *STC* units were superior to the control units, but the experimental methods and test instruments fail to reveal this difference.

Suppose that the *STC* units *were* superior to the control units. What experimental factors could degrade this difference?

1. *The research instructors were biased.* In this study, the research instructors designed their own control units. One may, therefore, question whether the research instructors took more "ownership" over these units and were more enthusiastic in the control classrooms.

To account for instructor bias, research instructors were carefully instructed (during pre-instructional interviews and in a letter to all research instructors) to remain impartial toward the *STC* and control units, in both subjective terms (projected level of enthusiasm) and objective terms (time of instruction). For accountability, classroom instruction was taped and several on-site visits were conducted. No bias was observed.

2. *The research instructors were underqualified for the study.* The study marked the first or second experience with the STC units for the research instructors. One may question whether the research instructors had adequate training and experience in the underlying methodology of the *STC* units.

However, the teachers selected for this study were very experienced and not at all naïve about modern hands-on approaches to science instruction (Appendix A). The research instructors represented some of the best teachers in the district, in a region known for its high-quality public schools. Onsite classroom observations and tape data indicate that the research instructors understood the unit content well and, with respect to teaching methodology, adequately fulfilled the intentions of the *STC* developers as documented in the *Teacher's Guide* and summer teacher training. The instructors spent considerable energy to prepare for each *STC* lesson, paying careful attention to lesson details and the inquiry-oriented spirit of the materials.

3. *The research instructors were overqualified for the study.* Given the training and background of the research instructors, one may question whether the control group units were taught too well and, therefore, did not reflect typical teaching in the district.

This objection is addressed by the following argument: Modern instructional strategies (like those implied by the *STC* units) place heavy demands on the teacher in terms of context knowledge (one cannot simply lean on the text book), facilitating collaborative small-group activities, eliciting and addressing students' conceptual difficulties, encouraging open-ended exploration, and so on. Hence, novice teachers, still refining basic classroom-management skills, may actually do *better* with more traditional approaches (like those implied by the control units)—at least at first. With proper training and experience, modern approaches to science instruction (e.g., those associated with "constructivism") eventually outrun traditional approaches (Gabel, 1994). Therefore, this study, conducted with highly qualified teachers, increases the likelihood of measuring a difference in favor of the *STC* units, if one exists. (In any case, one is impressed that the instructors were able to develop control units in less than two short weeks that achieved the same gains as the heavily supported *STC* units; this suggests the need to incorporate experienced teachers more fully in local course development.)

4. *The tests were not fair*. Perhaps the test instruments did not properly or fairly survey the instructional content of the units; e.g., the test items favored control-unit content, examples, and wording.

Three steps were taken to reduce this possibility. First, test construction was guided by the NSES document (National Research Council, 1996) without knowledge of the specific content of the *STC* or control group units. A posthoc analysis of test content showed no bias toward the control units. Second, items were selected to measure the most essential concepts of the content domain (e.g., “Does box *A* balance box *B?*” and “Which circuit lights the bulb?”). Third, in subject matter areas given to a wider variety of questions (weather, Grade 1), or not commonly taught in elementary school (motion and design, Grade 3), *STC*-tailored items were included and studied independently. An analysis of these items did not alter the conclusions of the study.

Conclusions

Having addressed these objections, we conclude that the adoption of the *STC* curriculum in the OASD will result in little or no gains on standard pencil-and-paper assessments (e.g., standardized tests) and potentially a slight decrease in student attitudes toward science. While improved teacher training may lead to higher learning gains, it is unclear that such training would lead to a significant gain differential.

The above conclusion is based on a study of more than 300 students, 16 classrooms, four grade levels, and four physical science units. The effect of repeated exposure to *STC* units across multiple grade levels, with novice teachers, or in conjunction with a more extensive teacher-training program, is not addressed.

Educational decision-makers should be guarded toward the wholesale adoption of any new elementary science program, especially if the primary goal is improved performance on standardized tests (Lattery & Lemberger, 2001b). However, one should not conclude from this study that, “traditional methods are just as good a modern methods." Unlike commercial educational comparisons, the control units described in this article are not a proverbial straw man.

Future investigations should pursue the following research questions: What if this study were performed with life science topics instead of physical science topics? What factors determine declines in student attitudes? Would a positive gain differential, in favor of the *STC* units, be observed in the second or third year of implementation?

References

Lattery, M., & Lemberger, J. (2001a). *Fox Valley Einstein Assessment Project: Research instruments for grades 1-4*. Oshkosh, WI: University of Wisconsin Oshkosh, Forrest Polk Library (SC No. 49).

Lattery, M., & Lemberger, J. (2001b). *Fox Valley Einstein Assessment Project (oral report).* April meeting of the OASD School Board. Send transcript requests to Oshkosh Area School District, Central Administration, 215 South Eagle Street, Oshkosh, WI, 54902.

Lattery, M., Lemberger, J., & Herzog, B. (2001). *Fox Valley Einstein Assessment Project (written report)* [online]. Available: http:// www.phys.uwosh.edu/lattery/ fveap/fveap.html.

Blumberg, F., Epstein, M., MacDonald, W., & Mullis, I. (1986). *A pilot study of higher order thinking skills assessment techniques in science and mathematics.* Princeton, NJ: National Assessment of Educational Progress.

CTB/McGraw-Hill (1996). *The Wisconsin state assessment (TerraNova).* California: Author.

Feinberg, L. (1990). Multiple-choice and its critics: Are the alternatives any better? *Commentaries from the College Board*, 3-15.

Gabel, D., ed. (1994). *Handbook of research on science teaching and learning.* Macmillan: New York.

Kanji, G.K. (1999). *100 statistical tests*. London: Sage.

Lesley College Program Evaluation and Research Group (1992). *Lesley report*. Cambridge, MA: Author.

Meng, E., & Doran, R. (1993). *Improving instruction and learning through evaluation.* Columbus, OH: ERIC Clearinghouse From Science, Mathematics, and Environment Education (ERIC Document Reproduction Service No. ED 359 066).

National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.

National Science Resources Center. (1997). *The science and technology for children (STC) curriculum*. Burlington, NC: Carolina Biological Supply Company.

Oshkosh Area School District. (1998). *The Oshkosh Area School District (OASD) curriculum guide*. University of Wisconsin Oshkosh Polk Library (reserve).

Resnick, L.B., & Resnick, D.P. (1992). Assessing the thinking curriculum: New tools for education reform. In B.R. Gifford & M.C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement and instruction* (pp. 37-75). Boston: Kluwer.

Saint Norbert College Survey Center. (2001). *Cornerstone study* [online]. Available: http://Program website: www.einsteinproject.org/cornerstone/index.shtml.

Schmidt, W., McKnight, C., & Raizen, S. (1997). *A splintered vision: An investigation of U.S. science and mathematics education*. Boston: Kluwer.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, *70*, 703-713.

Acknowledgements

This work was made possible by Grant FDR722 from the University of Wisconsin Oshkosh. We extend special thanks to Dr. Peter Hewson (University of Wisconsin Madison), Dr. Francis Lawrenz (University of Minnesota), and Dr. Larry Lowery (University of California Berkeley) for valuable suggestions and comments.

Appendix A: Research Instructors[*]

Mrs. Oak (Grade 1, Instructor A). Mrs. Oak has 7 years of teaching experience. She has a B.S. in elementary education and has a history minor. In addition to science methods classes as a student, she participated in two 1-week science education workshops locally. This was Mrs. Oak's second year using the STC Weather unit.

Mrs. Aspen (Grade 1, Instructor B). Mrs. Aspen has 5 years of teaching, two in the second grade and the remainder in first grade. She has a B.S. in elementary education and exceptional education with a minor in learning disabilities. She completed *STC* training in Solids and Liquids and AIMS training.

Mrs. Birch (Grade 2, Instructor A). Mrs. Birch has 23 years of teaching experience: 13 in the Oshkosh district, and 10 of those in the second grade. She has a B.S. in elementary education with 30 graduate credits. Her minor is in psychology. Her science background consists of previous *STC* training in the Butterfly unit. She is a WEST award winner and also the chair of OASD elementary science committee.

Mrs. Willow (Grade 2, Instructor B). Mrs. Willow has 5 years of teaching experience. She holds a B.S. in elementary education with a minor in language arts. Her science background consists of two college biology courses and high school biology and chemistry.

Mrs. Pine (Grade 3, Iinstructor A). Mrs. Pine has 12 years of teaching experience. She holds an M.S. in elementary education and in education administration. Her science background consists of previous *STC* training in the Butterfly unit, and 1.5 years of WASDI (Wisconsin Academy Staff Development Initiative) training in science and technology and math. Mrs. Pine has strong interests in scientific inquiry and cooperative learning.

Mrs. Elm (Grade 3, Instructor B). Mrs. Elm has 15 years of teaching experience. She is certified for grades 1-6 in elementary education and grades K-8 in learning disabilities. She has completed 31 graduate credits in science, health and economics. Her other science background consists of participation in the summer science workshop at University of Wisconsin Oshkosh (Operation Chemistry) and AIMS training. Mrs. Elm stresses both factual knowledge and hands-on experiences.

Mrs. Maple (Grade 4, Instructor A). Mrs. Maple has 31 years of teaching experience. She holds a B.S. in elementary education and is a science minor. She has training in AIMS and other local teacher-training programs. Mrs. Maple has a strong interest in science content.

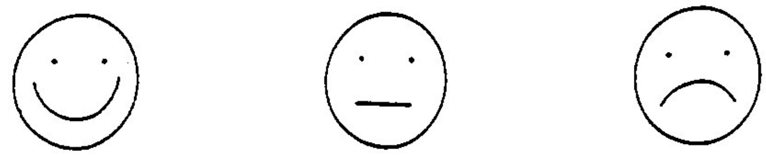
[*] Instructor names have been changed to maintain anonymity.

Mrs. Palm (Grade 4, Instructor B). Mrs. Palm has 8 years of teaching experience. She holds a B.S. in elementary education, with 20 graduate credits and a minor in English. She had two science courses in college: biology and general science. Mrs. Palm has a strong preference for activity-based teaching.

Appendix B: Sample Test Items

6. My best friends love science.

Yes Maybe No



7. Most kids in my class think science is fun.

Yes Maybe No

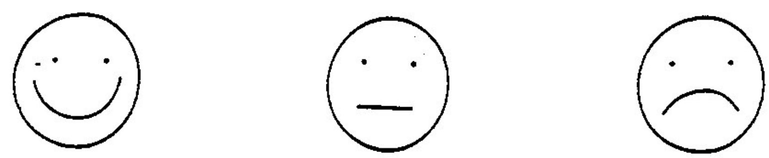


Figure 1. *Student attitude items.*

	Pre-Test	Post-Test
Control	88%	86%
Treatment	75%	97%

Part II

DIRECTIONS: After the teacher reads each question, circle the best answer.

1. Five degrees Fahrenheit (5 °F) is a measure of

temperature *weight* *speed*

2. Which is made mostly of water?

sun *snow* *wind*

Figure 2. *Weather items (recall knowledge). The insert shows the percentage of correct student responses for Item 2.*

	Pre-Test	Post-Test
Control	55%	70%
Treatment	65%	85%

4. Bob puts a thermometer in ***hot*** water.

Then, Bob puts a thermometer in *cold* water.

Bob mixes some hot water and cold water into a new cup. Which shows what the thermometer might read? *(Circle one)*

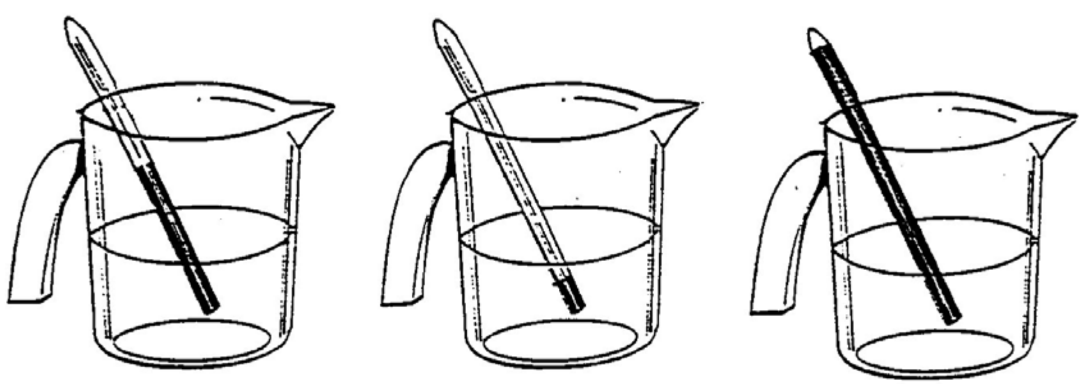
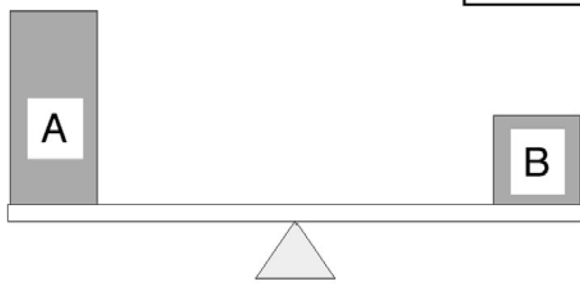


Figure 3. *Weather item (conceptual reasoning). The insert shows the percentage of correct student responses.*

3. This picture shows Box A *balancing* Box B.

	Pre-Test	Post-Test
Control	35%	40%
Treatment	40%	60%



Which is *true*? *(Circle one)*

a) Box A is heavier than Box B.

b) Box B is heavier than Box A.

c) Box A is the same weight as Box B.

Figure 4. *Balance & weighing item (conceptual reasoning). The insert shows the percentage of correct student responses.*

	Pre-Test	Post-Test
Control	40%	65%
Treatment	55%	58%

DIRECTIONS: Use these words and the table to answer the next questions.

Paul and Bria are born on Sunday. This table shows how much they *weigh* Monday, Tuesday, and Wednesday.

Baby Weight

Day	*Paul*	*Bria*
Monday	6 pounds	4 pounds
Tuesday	5 pounds	5 pounds
Wednesday	5 pounds	6 pounds

2. What is Bria's weight on Wednesday?

4 pounds *5 pounds* *6 pounds*

3. After Monday, Paul gets

heavier *lighter* *colder*

Figure 5. *Balance and weighing items (mathematical application). The insert shows the percentage of correct student responses for Item 3.*

2. Which electric circuit will cause the bulb to light?

	Pre-Test	Post-Test
Control	40%	75%
Treatment	22%	60%

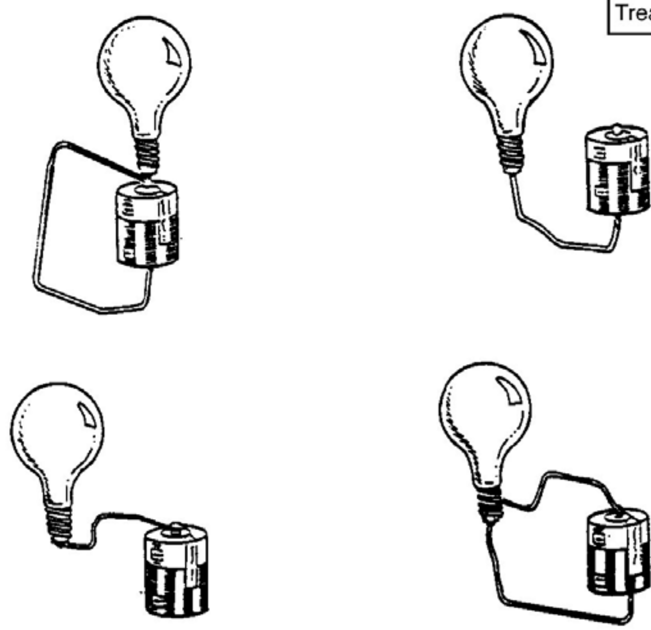


Figure 6. *Electric circuits item (conceptual reasoning). The insert shows the percentage of correct student responses.*