JOURNAL OF RESEARCH IN SCIENCE TEACHING VOL. 40, NO. 9, PP. 922–938 (2003)

Integrating Pharmacology Topics in High School Biology and Chemistry Classes Improves Performance

Rochelle D. Schwartz-Bloom,[1] Myra J. Halpin[2]

[1]*Department of Pharmacology and Cancer Biology, Duke University Medical Center, Box 3813, Durham, North Carolina 27710*

[2]*North Carolina School of Science and Mathematics, Club Blvd., Durham, North Carolina 27706*

Received 23 February 2002; Accepted 19 December 2002

Abstract: Although numerous programs have been developed for Grade Kindergarten through 12 science education, evaluation has been difficult owing to the inherent problems conducting controlled experiments in the typical classroom. Using a rigorous experimental design, we developed and tested a novel program containing a series of pharmacology modules (e.g., drug abuse) to help high school students learn basic principles in biology and chemistry. High school biology and chemistry teachers were recruited for the study and they attended a 1-week workshop to learn how to integrate pharmacology into their teaching. Working with university pharmacology faculty, they also developed classroom activities. The following year, teachers field-tested the pharmacology modules in their classrooms. Students in classrooms using the pharmacology topics scored significantly higher on a multiple choice test of basic biology and chemistry concepts compared with controls. Very large effect sizes (up to 1.27 standard deviations) were obtained when teachers used as many as four modules. In addition, biology students increased performance on chemistry questions and chemistry students increased performance on biology questions. Substantial gains in achievement may be made when high school students are taught science using topics that are interesting and relevant to their own lives. © 2003 Wiley Periodicals, Inc. J Res Sci Teach 40: 922–938, 2003

There has been a great deal of interest in improving science education at the precollege level, especially in the United States, where students rank low in science and math achievement

Contract grant sponsor: National Institute on Drug Abuse/National Institutes of Health; Contract grant number: SEDAPA DA10904.

Correspondence to: R.D. Schwartz-Bloom; E-mail: schwa001@duke.edu

DOI 10.1002/tea.10116

Published online in Wiley InterScience (www.interscience.wiley.com).

compared with their peers in other nations (National Center for Education Statistics, 1998). Over the past 2 decades, various curricular programs have been developed to improve science education, but most efforts in this area have produced only minimal positive gains in student performance (Beardsley, 1992; Bredderman, 1983; Ognes, 1991; Shymansky, Kyle, & Alport, 1983; Stohr-Hunt, 1996; Weinstein, Boulanger, & Walberg, 1982). In addition, rigorous evaluation of curricular programs has been limited because of difficulties in conducting well-controlled studies within the standard classroom.

Science educators have identified numerous factors that may contribute to poor science achievement, including, but not limited to, inadequate teacher training, students' views of science as boring and too hard, an overwhelming vocabulary, and limited hands-on activities (Willett, Yamashita, & Anderson, 1983; Yiping, 1996). Consequently, several initiatives are under way to address these problems, providing specific strategies to improve science education [American Association for the Advancement of Science (AAAS), 1993; National Science Education Standards (NSES), 1996; Glenn Commission, 2000]. For example, the National Research Council's *National Science Education Standards* (NSES, 1996) include guidelines to foster inquiry-based learning, problem solving, integration of subjects, use of real-world relevance, and teacher professional development. Recent studies indicate that instructional practices (i.e., laboratory inquiry, reduced teacher-centered instruction, critical thinking, professional development) meeting *National Science Education Standards* (NSES, 1996) have a positive influence on student achievement in middle and high school science, but the gain in achievement is small and inconsistent among different demographic groups (Kahle, Meece, & Scantlebury, 2000; Von Secker & Lissitz, 1999).

Research in developmental and cognitive psychology theories of constructivism (i.e., learners build on prior ideas to formulate their own understanding of phenomena) supports the idea that information embedded within meaningful contexts and applications fosters learning (Brooks & Brooks, 1993). Furthermore, teaching practices and topics that arouse student interest can help motivate students to learn and increase achievement (Sandoval, 1995). Although there is a paucity of research on the specific topics of interest to students, topics considered to be inherently interesting (especially to students) include death, danger, disease, injury, sex, and romance (Shank, 1979). In this regard, it is our belief that high school students are interested in how drugs affect their bodies. Thus, we hypothesized that science instruction in the context of drug-related topics (i.e., pharmacology) might improve student performance on tests of standard high school biology and chemistry concepts. To test this idea, we developed a set of teaching modules to be used in high school biology and chemistry courses that comprise pharmacology topics related generally to drug abuse. After a year of field-testing of the modules, we tested students for knowledge of standard biology and chemistry content (however, motivation to learn was not assessed). The modules address several of the *Benchmarks for Science Literacy* (AAAS, 1993) and the *National Science Education Standards* (NSES, 1996), including teacher training and participation in materials development (in a partnership with university faculty), inquiry-based learning, real-world relevance, integration of subject matter, and novel assessment strategies. We used a randomized wait-listed control design to test the efficacy of our program on student content knowledge in biology and chemistry.

Methodology

This study [the Pharmacology Education Partnership (PEP)] had four components: material development, teacher training, field-testing, and student evaluation. Each of these components is described below.

Development of Teaching Modules

Initially, we developed four teaching modules that applied basic principles of biology and chemistry to specific drug-related topics (Table 1). Topics were chosen with the expectation that students would identify with the subject matter based on personal experience, popular culture, or information from the news media. Three of the topics covered the biological and chemical aspects of abused drugs, and the fourth covered the biology and chemistry of nerve gas. The basic principles of biology and chemistry contained in each topic addressed specific National Science Education science content standards. These include: Content Standard A—structure of atoms, structure and properties of matter, and chemical reactions; Content Standard C—the cell, matter, energy and organization of living systems, and the behavior of organisms; and Content Standard F—personal and community health. Each module was designed to contain a teacher section and a student section. The modules were organized to include the following: a list of learning objectives; a student handout that posed questions about the topic to encourage inquiry-based investigation; a teacher guide with content, illustrations, and answers to the questions; a glossary; supplemental student activities (these were generated by teachers at the workshop described below); and a list of pertinent resources. The four teaching modules were bound in a notebook presented to teachers during a workshop given at Duke University (see next section). The modules can be accessed after June, 2004 and downloaded at the PEP website: http://www.thepepproject.net. They are now joined by an additional two modules (Modules 5 and 6) that were not part of this study.

Teacher Recruitment and Training

Subsequent to the development of the teaching modules, we recruited high school teachers of biology and chemistry throughout the United States to participate in the study. Announcements for a request for applications to the PEP project were included in the National Science Teachers Association newsletter and also at the North Carolina Science Teachers Association annual meeting. Pairs of biology and chemistry teachers from the same school were encouraged to apply. Fifty experienced teachers (25 biology and 25 chemistry, including 8 teacher pairs) were chosen by the authors to participate in the study, based primarily on (a) at least 5 years' teaching experience, (b) creative answers to essay questions within the application, and (c) prior experience in science program development. The demographics of the schools and the teachers who participated in the study are listed in Table 2.

Table 1
Pharmacology modules developed for the PEP project

Module Title	Module Content
Acids, Bases, and Cocaine Addicts	Acid–base chemistry, molecular structure, circulatory system, membrane transport, cocaine formulations, addiction biology
Drug Testing: A Hair-Brained Idea	Acid–base chemistry, molecular structure, cellular structure, anatomy, biology and chemistry of hair, nicotine, cocaine, heroin, racial ethics
How Drugs Kill Neurons: It's Radical!	Oxidation–reduction, oxygen radicals, neuron structure, neurochemistry, cell death, methamphetamine, neurodegenerative diseases
Military Pharmacology: It Takes Nerves	Covalent bonding, enzyme action, autonomic nervous system, physiology, behavior of gases, chemical warfare, Middle East and Japan current events/history

Table 2
Demographics of schools and teachers participating in the PEP project

Category	Number of Teachers
Subject	
Biology	18
Chemistry	18
Biology and Chemistry	11
School type	
Public	40
Parochial	5
Military	2
School minority population	
<10%	9
10–39%	17
40–79%	13
>80%	8
School locale	
Urban	23[a]
Suburban	13[a]
Rural	13[a]
US location	
Northeast	5
Southeast	22
Midwest	14
West	6

[a]Some schools have students from all three locales, so the teachers are listed more than once.

After the teachers accepted the offer to participate in the study, they were assigned randomly[1] to an experimental or control group. We used a wait-listed control design. In the first year, the experimental teachers ($n = 25$) attended a 1-week workshop at Duke University to learn basic pharmacology principles and help design classroom activities to accompany the teaching modules. The control group ($n = 22$; 3 teachers declined to matriculate into the study at the beginning of Year 1 because of changes in their teaching assignments) attended the same workshop 1 year later (see below). The workshop was 5 days in length and consisted of (a) interactive lectures from Pharmacology faculty on basic principles of pharmacology and drug abuse and their relationship to biology and chemistry concepts, (b) postlecture small-group discussions between teachers and Duke Pharmacology faculty, (c) small-group meetings for teachers to develop supplemental classroom activities for each of the four teaching modules, and (d) one afternoon of a hands-on lab experience to carry out one of the supplemental classroom activities. Some of the supplemental or class activities developed by the teachers were hands-on lab activities and some were not. The teachers developed activities that would reinforce the module content or provide some assessment of student learning of basic concepts contained in the modules. One activity for each teaching module was included in the teaching manual (these were tested by the authors to ensure that they worked) and the remainder of the activities was included in a separate section of the notebook for reference.

To assess the effectiveness of the workshop, a pretest/posttest test was administered. The test consisted of 20 true–false biology-, chemistry-, and pharmacology-related questions (see Appendix A for sample questions) constructed by the authors and it was administered to the teachers at the beginning of the workshop (pretest). The same test was administered to the teachers at the end of the workshop (posttest) and again 1 year later (without prior notification).

During the first year, control teachers did not participate in the workshop or have access to the materials developed. However, they did participate in the end of course testing (see below). The following year, they attended the workshop (i.e., wait-listed controls), receiving identical treatment compared with the teachers in the experimental group the year before.

Field-Test

During the year after the workshop, teachers in the experimental group field-tested the modules in their classrooms.[2] (In the first year, teachers in the wait-listed control group taught their classes as they did normally; they did not field-test until after they attended the workshop in the second year.) The classes consisted of biology and chemistry at both the lower and upper levels (see below). Despite instructions to teachers to implement materials in a specific manner, many teachers tend to modify the use of instructional materials to fit their own teaching style (James & Francq, 1988). Thus, there was no prescribed procedure for using the modules in the teaching plan. At the workshop, teachers discussed different ways in which they could incorporate the modules into their own teaching plans to either replace or supplement material they normally present in their courses (the modules were not designed to be standalone units of instruction.) These included: (a) Give students the student handout (this is inquiry-based) and have the students bring the answers to the questions back to the class for discussion; teachers use the teacher's guide to facilitate the discussion; (b) include the content from the teacher's guide to supplement the normal lecture content (either during a specific lecture, over a week, or even throughout the course); (c) include some or all of the content from the teacher's guide to replace the normal lecture content; and (d) Include the supplemental class activities whenever possible. Regardless of the implementation, the teachers were instructed to use as many modules as possible. A post-hoc tally of the modules used during the field-testing revealed that Modules 1, 2, 3, and 4 were used in 38, 22, 17, and 23 classrooms, respectively.

Measures

Approximately 2 weeks before the end of the course, both experimental and wait-listed control teachers were sent a short multiple choice test and Scantron score sheets. They administered the test unannounced to their students at the conclusion of their courses (end of course test). Use of an unannounced test minimized issues of studying and memorizing to explain improved performance. The tests were constructed by the authors with input from chemistry and biology teachers at the North Carolina School for Science and Mathematics. The test was composed of 20 questions (9 biology and 11 chemistry) (basic knowledge) similar to those found in first-year biology and chemistry textbooks. Multiple choice questions assessed student knowledge of facts and concepts in biology and chemistry as well as reasoning skills, according to the framework provided by the 1996 National Assessment of Educational Progress science test (O'Sullivan, Reese, & Mazzeo, 1997). Validation of the content relevance and appropriate difficulty of the questions was made by a separate group of biology and chemistry teachers at the North Carolina School for Science and Math.[3] Basic knowledge questions addressed knowledge and skills contained in the NSES Science Content Standards A—structure of atoms, structure and properties of matter, and chemical reactions; C—the cell, matter, energy and organization of living systems, and the behavior of organisms; and F—personal and community health. Following the 20 basic knowledge questions were 8 questions specific for new knowledge about drugs (advanced knowledge). The advanced knowledge questions probed knowledge about drugs obtained from the four teaching modules, i.e., information not provided at the high school level in most curricula.

Although the assessment items measured multiple areas of knowledge, the reliability of the assessment items was acceptable, with a Cronbach alpha coefficient of .74. Sample questions from the basic and advanced knowledge sections are included in Appendix B. Overall, scores from 4038 students were obtained for analysis; scores from 163 students could not be obtained owing to marking errors on the score sheets.

We hypothesized that several factors may affect students' scores on the end of course tests. Therefore, we obtained demographic information from the students regarding (a) their year in high school (student year, i.e., 9th through 12th grade), (b) course length (i.e., one-semester block or traditional year), (c) course type (i.e., biology or chemistry), and (d) course level (i.e., first year or second year/AP). Also, each teacher indicated the number of modules used (0, 1, 2, 3, or 4) for each class tested. Because both experimental and control teachers administered the tests at the end of Year 1, test scores could be compared (Table 3). Once control teachers attended the workshop and field-tested the modules the following year, test scores were obtained from their new batch of students. This second group of teachers provided the opportunity to use a paired comparison of student scores, with teachers serving as their own controls (Table 4).

Analytical Model

Percent correct scores on the tests were obtained from students in control and experimental teachers' classes (from both years) and the data were subjected to a hierarchical linear model (HLM) for analysis (Bryk & Raudenbush, 1992). We used a two-level HLM to analyze the data. This analysis takes into account the nested nature of the data, which include test scores from individual students within different classes. The first level models the relationship between student characteristics (i.e., student year—9th/10th and 11th/12th grades) and student scores on the end

Table 3
Results of HLM (Year 1 data) to identify student and classroom effects that predict basic knowledge and advanced knowledge test scores

	Basic Knowledge Scores	Advanced Knowledge Scores
Fixed effect		
Level 1		
Intercept	29.71 (1.07)	28.80 (1.41)
Student grade	−1.31 (0.99)	−2.70* (1.10)
Level 2		
Course length	−1.37 (2.67)	1.74 (3.08)
Course type	0.75 (2.14)	−2.36 (2.79)
Course level	10.12*** (2.20)	7.69* (3.25)
Teacher treatment	5.63** (2.05)	11.80*** (2.96)
Random effect		
Level 1 variance	136.27	278.21
Level 2 variance	77.41***	134.03***
Level 2 variance accounted for	0.19	0.20

Note. A positive value for the beta coefficient of each predictor indicates the following: Student grade—11th/12th graders scored higher than 9th/10th graders. Course length—students in a traditional school year course scored higher than students in a semester block course. Course type—students in chemistry classes scored higher than students in biology classes. Course level—students in a second-year or AP course scored higher than students in a first-year course. Teacher treatment—students in classes whose teacher attended the workshop scored higher than students in classes whose teacher did not attend the workshop. A negative value for the beta coefficient indicates the reverse. Standard errors are shown in parentheses.
*$p < .05$, **$p < .01$, ***$p < .001$.

Table 4
Results of HLM (controls in Year 1 and experimentals in Year 2) to identify student and classroom effects that predict basic knowledge and advanced knowledge test scores

	Basic Knowledge Scores	Advanced Knowledge Scores
Fixed effect		
Level 1		
Intercept	32.42 (0.88)	28.01 (1.16)
Student grade	−0.56 (0.85)	−0.80 (0.94)
Level 2		
Course length	2.76 (2.19)	3.17 (2.62)
Course type	5.05* (1.67)	0.04 (2.78)
Course level	13.16** (2.00)	7.41* (2.54)
Teacher treatment	9.09** (1.74)	9.91** (2.26)
Random effect		
Level 1 variance	146.90	271.45
Level 2 variance	52.88**	85.69**
Level 2 variance accounted for	0.55	0.27

Note. Beta coefficients indicate the percentage point gains between groups as described in Table 3. Standard errors are shown in parentheses.
*$p < .01$, **$p < .001$.

of course tests. Level 2 equations model the relationship between classroom factors and student scores. Three classroom factors were controlled for in each analysis: course type (biology versus chemistry), course length (one semester block schedule versus traditional yearlong course), and course level (first year versus second year/AP).

To identify student and classroom factors that could predict student scores, HLM analyses were conducted on data from three subgroup combinations. The first HLM was limited to Year 1 students' scores obtained from classrooms whose teachers either attended or did not attend (wait-listed controls) the workshop. The second HLM included data from classrooms whose teachers were wait-listed the first year but participated in the workshop during the second year. In these two HLMs, the main independent variable of interest was teacher treatment, i.e., whether the teacher attended the PEP workshop. The third HLM included all data from experimental teachers (both years) to determine whether the number of modules (0, 1, 2, 3, or 4) used by a teacher was a significant predictor of student scores. In each model, there was no variation between classrooms in the effect of the Level 1 predictor, student year, so the random effect was not included in the Level 2 regression equation. In addition, all independent variables were grand-mean centered.

For the HLM analysis of each subgroup, we tested four models: (a) an unspecified model that compared individual scores with class scores; (b) a model that added the Level 1 control of student grade; (c) a model that added Level 2 controls of course length, type, and year; and (d) a full model that added the Level 2 variable of teacher treatment (or in some cases, number of modules); (see Tables 5 and 7). For the full model, the effect size was calculated for selected significant predictors of student scores. The effect size of an independent variable was determined using the following equation: Regression coefficient for a predictor/Standard deviation (SD) of student scores.

Results

Student Assessment

To determine which factors were significant predictors of the student scores on the end of course tests, we performed three separate HLM analyses. In each analysis, Level 1 included

Table 5
Results of HLM (experimentals only) to identify student and classroom effects that predict basic knowledge and advanced knowledge test scores

	Basic Knowledge	Advanced Knowledge
Fixed effect		
Level 1		
Intercept	34.62 (0.98)	33.44 (1.14)
Student grade	−1.15 (0.96)	−2.36 (1.33)
Level 2		
Course length	1.37 (2.54)	3.16 (2.69)
Course type	4.33* (2.00)	0.73 (2.53)
Course level	10.62*** (2.65)	6.01* (2.75)
1 module	6.24 (3.39)	−0.17 (3.04)
2 modules	6.08** (2.32)	6.31* (3.04)
3 modules	11.97*** (3.23)	10.84* (4.30)
4 modules	19.54*** (3.21)	27.62*** (3.77)
Random effect		
Level 1 variance	146.86	290.08
Level 2 variance	69.24***	83.98***
Level 2 variance accounted for	0.52	0.61

Note. Beta coefficients indicate the percentage point gains between groups as described in Table 3. Standard errors are shown in parentheses.
*$p < .05$, **$p < .01$, ***$p < .001$.

characteristics of the students and Level 2 included characteristics of the classrooms. The first HLM was used to identify whether the teacher treatment (attending the workshop; experimentals) was a significant predictor of student scores regardless of the number of modules used. Table 3 summarizes HLM results (full model only) for the basic knowledge and advanced knowledge questions. The HLM analysis indicates that the Level 1 factor, student year (i.e., 9th/10th and 11th/12th grades), was not a significant predictor of the individual students' scores for the basic knowledge questions. At the classroom level, Level 2 control variables (course length, course type, and course level) accounted for 12% of the between-class variance for basic knowledge questions and 0% of the between-class variance for advanced knowledge questions. Of these variables, only course level was a significant predictor of student scores; students in advanced courses scored on average about 10 percentage points and 7 percentage points higher than students in first-year courses on basic knowledge and advanced knowledge questions, respectively. Teacher treatment (attendance at the workshop) was a significant predictor of students' scores for both basic and advanced knowledge questions ($p < .001$ each). Students scored approximately 6 percentage and 12 percentage points higher on basic knowledge and advanced knowledge questions, respectively, when teachers attended the workshop. The effect sizes were 0.38 and 0.58 standard deviations, respectively (Table 6).

The second HLM was performed to rule out possible positive teacher effects (i.e., that the increase in student scores could reflect better teachers in the experimental group) despite random assignment. The analysis included data from classrooms whose teachers were in the wait-listed control group in the first year and in the experimental group in the second year. The results from this analysis are similar to those obtained in the first HLM analysis (Table 4). At the classroom level, Level 2 control variables (course length, course type, and course level) accounted for 38% of the between-class variance for basic knowledge questions and 8% of the between-class variance for advanced knowledge questions. Of these variables, course type and level were significant

Table 6
Summary of effect sizes for selected significant predictors identified in Tables 3–5

Level 2 Predictor		Effect Size	
		Basic Knowledge	Advanced Knowledge
Teacher treatment—Year 1 data (Table 3)		0.38	0.58
Teacher treatment–Year 1 & 2 data (Table 4)		0.58	0.50
Two modules	Year 1 and 2 data (Table 5)	0.38	0.22
Three modules		0.74	0.50
Four modules		1.21	1.27

predictors of student scores. Students in chemistry courses scored on average about 5 percentage points higher than biology students on basic knowledge questions, and students in the advanced courses scored on average about 13 percentage points higher than students in first-year courses on basic knowledge and advanced knowledge questions, respectively. For advanced knowledge questions, students in the advanced courses scored on average about 7 percentage points higher than students in the first-year courses. At the classroom level, teacher treatment was a significant predictor of students' scores on both basic and advanced knowledge. Students scored approximately 9 percentage points higher on basic knowledge questions when their teachers had attended the workshop and they scored approximately 10 percentage points higher on advanced knowledge questions, when teachers attended the workshop. Interestingly, the effect size for the basic knowledge questions was 0.58 standard deviations, considerably larger than the effect size obtained in the first HLM (Table 6). For advanced knowledge questions, the effect size was 0.50 standard deviations. Thus, it is unlikely that the increased scores in both knowledge areas (shown in Table 3) were due to better teachers in the experimental group.

Once we determined that student scores could be predicted by teacher treatment, we subjected all of the data obtained from experimental teachers (both years) to an HLM to determine whether

Table 7
Results of HLM to identify student and classroom effects that predict basic knowledge test scores obtained from only biology students or only chemistry students

	Biology Students		Chemistry Students	
	Biology Questions	Chemistry Questions	Biology Questions	Chemistry Questions
Fixed effect				
Level 1				
Intercept	42.36 (2.41)	25.72 (1.35)	32.22 (2.01)	31.41 (2.40)
Student grade	−4.62 (2.93)	1.58 (1.96)	1.47 (1.23)	−1.97 (1.10)
Level 2				
Course length	−6.52 (7.30)	−5.94 (5.08)	8.58* (3.91)	4.00 (4.96)
Course level	19.06** (6.10)	11.60** (3.50)	6.07 (4.64)	15.21* (5.68)
Modules	22.65** (6.75)	12.31* (4.76)	18.55*** (4.32)	12.77* (4.99)
Random effect				
Level 1 variance	255.40	158.87	236.84	173.89
Level 2 variance	146.34***	44.76***	181.66***	259.19***
Level 2 variance accounted for	0.50	0.46	0.38	0.26

For this HLM, the Modules variable included two groups, 0 or 1 module compared with 2, 3, or 4 modules. Beta coefficients indicate the percentage point gains between groups as described in Table 3. Standard errors are shown in parentheses.
*$p < .05$, **$p < .01$, ***$p < .001$.

the number of modules used could predict student scores. Table 5 lists the results of this HLM. The HLM analysis indicates that the student year (i.e., 9th/10th and 11th/12th grades) was not a significant predictor of the individual students' scores for either basic knowledge or advanced knowledge questions. Level 2 control variables accounted for 26% and 14% of between-class variance, respectively. Of these variables, students in chemistry classes scored about 4 percentage points higher than biology students on basic knowledge questions. In addition, students in advanced classes scored on average about 11 percentage points and 6 percentage points higher than students in first-year courses on basic and advanced knowledge questions, respectively. Use of the modules was also a significant predictor of students' scores for both basic and advanced knowledge questions. On average, students scored significantly higher on both sets of questions when at least two modules were used. For basic knowledge questions, students scored approximately 6, 12, and 20 percentage points higher when 2, 3, or 4 modules were used, respectively, compared with 0 modules. Very large effect sizes were obtained when teachers used more than two modules in their classrooms (up to 1.21 standard deviations) (Table 6). Similarly, for advanced knowledge questions, students scored approximately 6, 11, and 28 percentage points higher when 2, 3, or 4 modules were used, respectively, compared with 0 modules. Again, very large effect sizes were obtained when teachers used more than two modules (up to 1.27 standard deviations).

Mean student scores for both basic knowledge and advanced knowledge questions are shown in Figure 1, grouped by number of modules used. As can be seen, end of course tests revealed relatively low scores in all groups; there may be several factors responsible for this. First, the tests included a mixture of biology and chemistry questions; most biology students (first year) have not had chemistry yet. This is supported by the data provided in Table 7; biology students scored about 10 percentage points higher on the biology questions than on the chemistry questions (see below). Second, the multiple choice test included a ''don't know'' choice (0% chance of a right answer), and students may have chosen this option rather than guessing (25% chance of a right answer). Third, in most cases, students did not study for this test because they did not know ahead of time that they would be tested on the modules. Regardless of the low baseline scores, the HLMs indicated that the use of at least two modules was a significant predictor of higher students scores on basic knowledge and the advanced knowledge questions.

Two other questions were interesting in this study. First, could the use of modules containing both biology and chemistry principles (in a pharmacology-oriented context) help biology students learn biology better and chemistry students learn chemistry better? Second, could the use of the pharmacology modules help biology students learn chemistry better and chemistry students learn biology better? To address these questions, we performed four separate analyses (using HLM but without the independent variable of course) of the biology or chemistry basic knowledge scores from biology or chemistry students. For these analyses, the data for the variable Module were grouped as follows: 0 or 1 module versus $\geq$2 modules. The previous HLM (Table 6) provided the basis for grouping these data (one module was not a significant predictor of student scores), thereby providing a sufficient cell size for analysis of the smaller data sets. HLM results (full model only) for each of the four analyses are shown in Table 7. For each comparison, use of two or more modules was a significant predictor of student scores on basic knowledge questions. For example, biology students using $>$2 modules scored approximately 23 percentage points higher on biology questions than students using $<$2 modules (effect size of 0.22 standard deviations); the same students scored about 12 percentage points higher on chemistry questions (effect size of 0.78 standard deviations). Chemistry students using $>$2 modules scored approximately 13 percentage points higher on chemistry questions than students using $<$2 modules (effect size of 0.58 standard deviations); the same students scored about 19 percentage points higher on biology questions (effect size of 0.82 standard deviations).

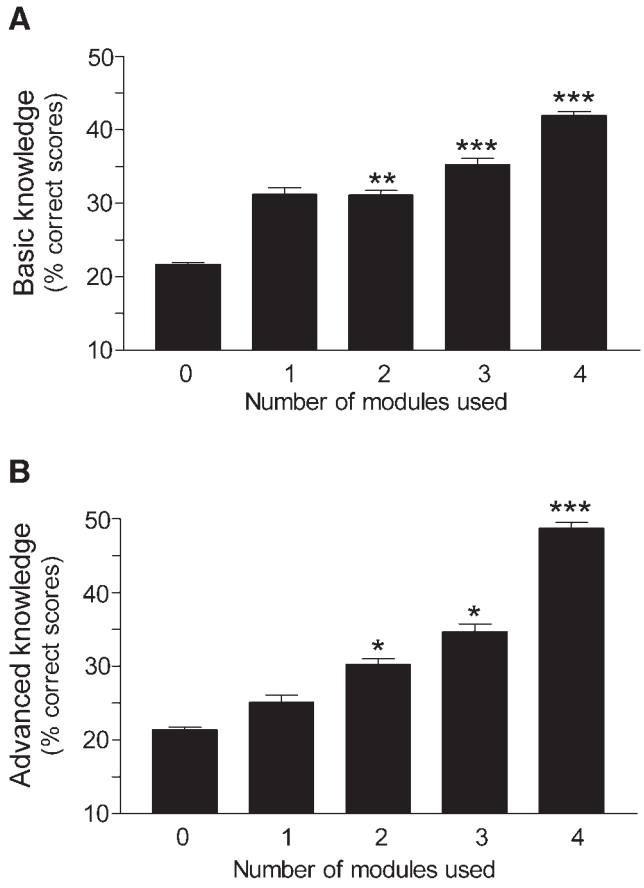

Figure 1. Performance of all students on questions of basic knowledge (A) and advanced knowledge (B), using 0, 1, 2, 3, or 4 modules during the course. Data for the 0 modules group includes scores from controls as well as experimental teachers who used 0 modules. Data are the mean ± SEM scores from all students (biology and chemistry at both levels). $*p < .05$, $**p < .01$, $***p < .001$ versus zero modules (HLM analysis). The raw scores ranged from 0 to 18 correct and 0 to 8 correct for basic and advanced knowledge, respectively. (A) $n = 2416, 343, 536, 351$, and 662 for 0, 1, 2, 3 or 4 modules, respectively. (B) $n = 2361, 326, 530, 348$, and 658, respectively.

Teacher Assessment

The training and professional development of the teachers were important components of this study. The workshop provided teachers with basic pharmacology and principles of drug abuse and it gave them the opportunity to develop student activities in the teaching modules. To assess the effect of the workshop on teacher knowledge of basic biology, chemistry, and pharmacology principles, teachers were administered pre- and posttests without prior warning (see Appendix A). The 1-week posttest indicated a significant gain in knowledge ($p < .001$) of these areas at the end of the training workshop (Figure 2). To assess the impact of the workshop long-term, teachers were administered a posttest again, 1 year later. Teachers maintained their knowledge gain ($p < .001$) for at least 1 year after they completed the workshop.

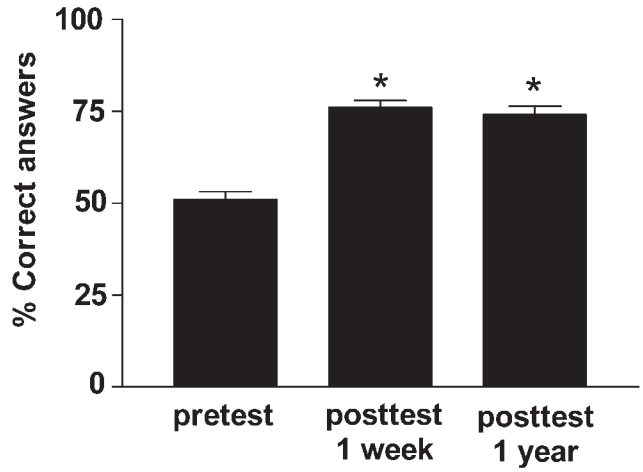

Figure 2. Long-term knowledge retention by teachers attending the Pharmacology training workshop. Data are the mean ± SEM percent correct scores from a 20-item test of biology, chemistry, and pharmacology principles taken the first day of the workshop (pretest), the last day of the workshop (1 week posttest), and 1 year later (1 year posttest). *$p < .001$, repeated measures ANOVA, and Tukey multiple comparisons test ($n = 31$).

Discussion

In this study, we used a randomized wait-listed control design to test the hypothesis that the use of pharmacology topics in high school biology and chemistry classes improves student performance on tests of basic biology and chemistry concepts. Also, we hypothesized that the use of pharmacology topics would increase high school student knowledge about the effects of drugs (advanced knowledge). In both cases, student achievement on the end of course tests demonstrated a dose-responsive relationship; as the number of modules used increased, the students' performance increased further, with gains of up to 28 percentage points when four modules were used compared to no modules. For both basic and advanced knowledge, effect sizes were large when more than two modules were used. The substantial gains in biology classes on chemistry questions, and vice versa, highlight the usefulness of pharmacology topics, which have an inherently integrative nature, in science education. The dose–response design is difficult to perform in educational testing, yet it can be a powerful assessment instrument. For example, an analysis of data collected from the National Education Longitudinal Study of 1988 (NELS, 1988) revealed that the frequency of hands-on activities (i.e., never, monthly, weekly, daily) was associated significantly with improved science achievement (Stohr-Hunt, 1996). In that study, a maximum gain of only 5 percentage points was obtained for the most frequent activities.

Several other features of this study are notable in educational testing. The first is the random assignment of teachers to the control and experimental groups. Random assignment increases the likelihood that the two groups are inherently similar and that differences in outcome are caused by the treatment being tested and not by teacher differences. Second, the study design also allowed wait-listed control teachers to participate as the experimental group the following year, helping to rule out teacher effects on the positive outcome. Third, the implementation of the modules did not conform to a specified format; the ways in which teachers incorporated the modules into their teaching were unique to each teacher. This design was used to mimic real-life classroom conditions in which teachers tend to modify the use of instructional materials to fit their own

teaching style (James & Francq, 1988). However, the lack of a prescribed implementation format may also be considered a limitation in this study because it prevented us from testing or identifying those implementation styles that may have worked better than others. The use of a classroom instructional setting rather than the laboratory setting in educational testing can yield different outcomes. This can be illustrated in studies of computer-assisted instruction in high school science classes. Meta-analyses reveal that computer-based instruction in science education is associated with small increases in student achievement (Willett et al., 1983). However, in a study of high school chemistry students assigned randomly to control and treatment groups, a widely accepted computer-assisted instruction package did not contribute to enhanced achievement (Wainwright, 1989). In the latter study, the instructional setting rather than the controlled laboratory setting was used, perhaps contributing to the lack of significant achievement. In contrast, we obtained substantial gains in student performance despite testing our program in standard classroom environments, where there was variation in the implementation of the modules.

Another limitation of this study is that it cannot distinguish whether the use of the pharmacology topics to provide basic biology and chemistry concepts or the repetition of biology and chemistry concepts within the modules (independent of the pharmacology topic) is responsible for the increased student scores. Although we cannot state with certainty whether the use of the pharmacology topics or the repetition is responsible for the improvement, at the least, the pharmacology topics can provide the teacher with an interesting mechanism to develop repetition without boring students. One might question whether the use of these modules adds to the burden of an already crowded curriculum. For a few teachers, this may have been the case. However, in our post-hoc discussions with teachers, the most common reason for running out of time was the loss of teaching days owing to snow. Teachers who were able to incorporate the modules into their curricula did not indicate that doing so was a burden. In fact, almost half of the teachers in the study were able to use all four modules.

A key feature of this project was the teacher training at the workshop. The teacher training addressed several elements of the *National Science Education Standards* (NSES, 1996). First, in a partnership with university scientists, teachers learned new information (i.e., basic pharmacology principles) to help students integrate biology and chemistry concepts. The long-term knowledge retention by the teachers exemplifies the positive impact of the workshop. The *National Science Education Standards* include the integration of subjects for teaching science, but currently, teachers may not have the tools to achieve this goal. Traditional secondary teachers rarely have the opportunity to experience a curriculum that explores connections and interrelationships among disciplines (Mosenthal & Ball, 1992). Although we did not test for the effect of integrating disciplines on student achievement, it is clear that the use of pharmacology topics, which are inherently integrative in nature, was associated with improved student performance in both biology and chemistry disciplines. This use of connectivity of disciplines and the societal implications is called for in redefining science education today (Bardeen & Lederman, 1998). Second, the teachers participated in the development of the teaching modules. Small-group discussions with university scientists fostered a creative atmosphere for the teachers to develop student activities for each module, tailored to their target audience. Active participation of teachers in the development of materials may have provided them with the individual ownership critical to motivating teachers to make changes in their teaching (Givens, 2000). Interestingly, in a study of 3500 high school science teachers, the amount of professional development and the content preparation had a strong influence on science teaching practices that would be expected to increase student achievement (Supovitz & Turner, 2000). In our study, the professional development afforded by the 1-week workshop (and working with university faculty) was associated with increased student scores. However, students in classrooms of teachers who

attended the workshop but did not field-test modules did not score higher on the end of course tests than students of control teachers who did not attend the workshop. Thus, it appears that both knowledge obtained in the workshop and implementation of the modules were necessary to improve student performance. Although it is possible that use of the modules, per se, without the benefit of teacher training may enhance student performance, we think this is unlikely.

Last, and perhaps most important, the real-world relevance of the content in the modules may have been a major factor in the successful outcome of this study (a separate study with more controlled conditions would be required to prove such an association). Topics on drugs and drug abuse are probably highly relevant and meaningful to high school students. If such topics can help capture student interest in science, other features of science education reform may be more effective. One of the ultimate goals in science education is to help students use science to be critical thinkers and make good decisions in their daily lives (Yiping, 1996). It remains to be determined whether a program such as the one we developed will help teenagers make intelligent decisions about drug use. Nevertheless, the approach we have taken should be applicable to many areas of science that are parts of students' daily lives.

The authors thank all of the teachers in the PEP project for their efforts in making this project a success. They thank the National Science Teachers Association and the North Carolina Science Teachers Association for providing assistance in the teacher recruitment. In addition, they thank Eric Banilower for statistical help and critical reading of the manuscript, and Dr. Paul Bloom for providing helpful comments during the preparation of the manuscript.

Notes

[1]Random assignment to the two groups was carried out with a coin toss. Once 25 teachers were reached in one group, the remainder was placed in the other.

[2]One experimental teacher dropped out during the first year owing to a serious illness.

[3]A panel of 5 additional teachers reviewed the basic knowledge questions for content relevance in biology and chemistry and appropriate difficulty. Of the 20 questions, 4 or 5 teachers rated 19 questions relevant and appropriate. One of the 20 questions was rated relevant and appropriate by 3 of the 5 teachers.

Appendix A

Sample Questions from Teachers' Pre/Post Workshop Test	*True or False*
The right heart receives oxygenated blood.	False
A base is a compound that donates an H+	False
When iron is oxidized, it loses an electron to oxygen.	True
Drug actions are terminated only by removal of the drug from the body.	False
Drugs can remain in the body for weeks after administration ceases	True
Lipids contain long chains of $–CH_2–$ groups; this makes them nonpolar.	True
Alcohol is removed from the body in the urine.	False
All blood vessels contain smooth muscle except capillaries.	True
Inhaled compounds travel through the general circulation before reaching the heart.	False
Drugs bind to proteins and to DNA	True

Appendix B

Sample Questions from End of Course Test (Basic Knowledge) (Answers Are Underlined)

Biology Questions

10. Which of the following is responsible for the color present in hair and skin?
    A. Keratin
    B. Collagen
    C. Melanin
    D. Cuticle
    E. Don't know

14. The connection formed between two neurons is called the:
    A. Synapse
    B. Node of Ranvier
    C. Dendrite
    D. Myelin
    E. Don't know

23. Water-soluble vitamins are large molecules that cannot pass through cell membranes by passive diffusion. Which transport process do they use?
    A. Filtration
    B. Facilitated diffusion
    C. Secretion
    D. Active transport
    E. Don't know

Chemistry Questions

8. An acid that does not dissociate completely in water is called:
    B. A strong acid
    C. A weak acid
    D. Ionized
    E. Hydrophobic
    F. Don't know

13. Nerve gas (sarin) is a colorless and odorless gas that is extremely toxic. It is a liquid that vaporizes readily at room temperature and hovers close to the ground. This implies that nerve gas
    A. Has a low vapor pressure
    B. Has a high boiling point
    C. Is less dense than air
    D. Is more dense than air
    E. Don't know

19. An atomic or molecular species in which one or more outer orbital(s) contains an unpaired (lone) electron is:
    A. Very stable
    B. Very unstable (free radical)

C. Similar to oxygen
D. Similar to hydrogen
E. Don't know

21. In the reaction: $Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + H_2O + O_2^{\cdot -}$ iron is:
A. Oxidized
B. Reduced
C. Neutralized
D. Hydrolyzed
E. Don't know

Sample Questions from End of Course Test (Advanced Knowledge)

26. Drugs like nicotine and morphine accumulate in hair because they are:
A. Highly metabolized
B. Poorly metabolized
C. Strong bases and bind to keratin
D. Weak bases and bind to melanin
E. Don't know

27. Methamphetamine (speed or ice) causes destruction of neurons by:
A. Inhibiting cell metabolism
B. Increasing dopamine oxidation
C. Decreasing ATP levels
D. Increasing dopamine hydrolysis
E. Don't know

29. Drugs like nicotine, heroin, and cocaine enter the brain easily because they are:
A. Hydrophilic
B. Lipophilic
C. Highly charged
D. Water soluble
E. Don't know

32. Nerve gas poisoning causes salivation, vomiting, urination, diarrhea, convulsions, and death owing to:
A. Destruction of the central nervous system
B. Inhibition of the parasympathetic nervous system
C. Stimulation of the parasympathetic nervous system
D. Inhibition of the sympathetic nervous system
E. Don't know

References

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

Bardeen, M.G. & Lederman, L.M. (1998). Coherence in science education. Science, 281, 178–179.

Beardsley, T. (1992). Teaching real science. Scientific American, 267, 92–108.

Bredderman, T. (1983). Effects of activity-based elementary science on student outcomes: A quantitative synthesis. Reviews of Educational Research, 53, 499–518.

Brooks, J.G. & Brooks, M.G. (1993). In search of understanding: The case of constructivist classrooms. Alexandria, VA: Association for Supervision and Curriculum Development.

Bryk, A. & Raudenbush, S. (1992). Hierarchical linear models. Thousand Oaks, CA: Sage.

Givens, N. (2000). Curriculum materials as a vehicle for innovation: A case study of the Nuffield Design and Technology Project. Research in Science and Technological Education, 18, 71–83.

Glenn Commission. (2000). Before it's too late: A report to the nation from the National Commission on Mathematics and Science Teaching for the 21st Century. Washington, DC: US Department of Education.

James, R.K. & Francq, E. (1988). Assessing the implementation of a science program. School Science and Mathematics, 88, 149–159.

Kahle, J.B., Meece, J., & Scantlebury, K. (2000). Urban African-American middle school science students: Does standards-based teaching make a difference? Journal of Research in Science Teaching, 1019–1041.

Mosenthal, J.H. & Ball, D.L. (1992). Constructing new forms of teaching: Subject matter knowlegde in inservice teacher education. Journal of Teacher Education, 43, 347–356.

National Center for Educational Statistics. (1998). Pursuing excellence: A study of US twelfth-grade mathematics and science teaching, learning, curriculum, and achievement in international context. Third International Mathematics and Science Survey. Washington, DC: US Government Printing Office.

National Research Council. (1996). National science education standards. Washington, D.: National Academic Press.

Ognes, E. (1991). A review of science education: Past failures, future hopes. American Biology Teacher, 53, 199–203.

O'Sullivan, C.Y., Reese, C.M., & Mazzeo, J. (1997). NAEP 1996 science report card for the nation and the states. Washington, DC: National Center for Education Statistics.

Sandoval, J. (1995). Teaching in subject matter areas: Science. Annual Reviews in Psychology, 46, 355–374.

Shank, R.C. (1979). Interestingness: Controlling inferences. Artificial Intelligence, 12, 273–297.

Shymansky, J.A., Kyle, W.C., & Alport, J.M. (1983). The effects of new science curricula on student performance. Journal of Research in Science Teaching, 20, 387–404.

Stohr-Hunt, P.M. (1996). An analysis of frequency of hands-on experience and science achievement. Journal of Research in Science Teaching, 33, 101–109.

Von Secker, C.E. & Lissitz, R.W. (1999). Estimating the impact of instructional practices on student achievement in science. Journal of Research in Science Teaching, 36, 1110–1126.

Wainwright, C.L. (1989). The effectiveness of a computer-assisted instruction package in high school chemistry. Journal of Research in Science Teaching, 26, 275–290.

Weinstein, R., Boulanger, D., & Walberg, H.J. (1982). Science curriculum effects in high school: A quantitative synthesis. Journal of Research in Science Teaching, 19, 511–522.

Willett, J.B., Yamashita, J.J.M., Anderson, R.D. (1983). A meta-analysis of instructional systems applied in science teaching. Journal of Research in Science Teaching, 20, 405–417.

Yiping, L., Abrami, P.C., Spence, J.C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. Review of Educational Research, 66, 423–458.