An Analysis of Student Learning: Using a Standard-Based Earth Science Curriculum in the U.S.

Do-Yong Park*

Science Education, Department of Curriculum and Instruction, Illinois State University,
IL 61790-5330, USA

Abstract: The purpose of this study was to investigate the effectiveness of EarthComm implementation in the U.S. high schools in terms of demographic background including school size, urban/rural area, and teachers' teaching experiences. In addition, this study examined impact of students' higher-order thinking skills by using the visions of National Science Education Standards. Two modular of the EarthComm curriculum were used for this purpose with thirty one teachers and around thousand students involved across four states. Findings were that EarthComm did not significantly impact student achievement differentially in schools of varying sizes and school location, i.e., urban and rural areas. The years of teaching experiences did not impact student achievement scores for Module I but did significantly impact for Module II. It is noted that the two results seemingly conflict with each other similar to other research findings (Ferguson, 1998; Yager et al., 1988). Student higher-order thinking skills, on the other hand, were significantly improved as a result of studying with EarthComm. Implications were discussed at the end of the paper.

Keywords: EarthComm, National Science Education Standards, Curriculum, Higher-order Thinking Skills, Textbook

Introduction

Of the various major curriculum reform projects ongoing in the U.S., National Science Education Standards (NSES) (NRC, 1996) was singled out in terms of its influence on school science curriculum. With NSES's visions infused, a series of new curriculum have been developed for the last decade. The central concept of this curriculum reform in the U.S. was 'Sciences in the Community (COM).' After Chemistry in the Community (ChemCom) was published, which did not have NSES's visions, Biology in the Community (BioCom) in 1996, Active Physics in 2000, Earth Systems Science in the Community (EarthComm) in 2000, and Active Chemistry of which curriculum development started in 2002 (Park, 2006). These curricula all exemplify the visions of National Science Education Standards in its philosophy, teaching, learning, professional development, assessment, and content. Currently the COM curricula

became a model for other science curriculum in the U.S. and around the world.

Earth System Science in the Community (EarthComm) was developed by the American Geological Institute (AGI). It is a high school earth science curriculum seeking to promote scientific literacy among all levels of students in the nation and to produce a citizen who understands the "big ideas" about our Earth. The need for reform of Earth science education in the U.S. is imminent. Unlike Asian countries including Korea, only less than 10% of the U.S high schools provide earth science courses. Exline (1998) pointed out several factors that contribute to this situation through an extensive survey: 1) lack of certified professional teachers 2) many states lacking teacher certification standards 3) lack of appropriate professional development programs for earth science teachers 4) inappropriate earth science instructional resources.

The NSES call for the teaching of earth science at all grade levels. To implement the NSES standards in Earth science, its curriculum, assessment, and content are to change to accommodate the visions of it. EarthComm specifically focused on these requirements for change. Consistent with the aforementioned current

*Corresponding author: dypark@ilstu.edu
Tel: 309-438-5050
Fax: 309-438-8659

reform efforts, EarthComm curriculum focused on making science local, relevant, and personal to the lives of learners. EarthComm curriculum does not cover as many topics as the traditional high school earth science textbook (Park, 2005). Instead, it focuses on key concepts, conceptual understandings, appreciation of the earth systems, critical thinking skills and abilities that all students can utilize to make wise decisions in their life. Park's study (2005) indicated that Korean Earth science textbook was characterized as being traditional in terms of topics to cover, laboratory activity, and questioning style that are key components of NSES. Therefore, the vision of EarthComm is to promote and improve the teaching and learning of Earth science in all high schools not only in the U.S. but also the worldwide.

Research Questions

When a new curriculum is developed, people in the field often come to a wonder if the new curriculum could be generalized in a large scale in terms of its use with multiple factors or variables. This study chose three generic variables to see if they impact on students learning that is measured by student achievement scores when a new Earth Science curriculum is implemented in classrooms: (1) school community, (2) school size, and (3) years of teaching experiences.

Thus, this study is interested in:

1. How do various groups of students who were involved with the EarthComm achieve on two modules related to 1) school size, 2) urban/rural, and 3) extent of previous teaching experience of the teachers?

In addition, since the EarthComm program and the *National Science Education Standards* are closely related in terms of fundamental goals and philosophy. So the development of EarthComm heavily relies on NSES. Specifically, the NSES emphasizes inquiry in science activities to promote students' higher order thinking skills. So this study explores the following question:

2. How does EarthComm impact student success in developing higher-order thinking skills?

The EarthComm program was designed specifically to address the National Science Education Standards. Although variability in teaching approaches and content coverage in Earth science would be expected from state to state, most states have based their science standards upon the NSES. Thus, results gleaned from the analysis of EarthComm with respect to the Standards shall shed implications for Earth science education in each state.

Research Context

Teaching Activity in Modules I and II

Basically, all the chapters of EarthComm are structured with a variety of inquiry-based activities. Module I contains three chapters and 17 activities about *Earth's Dynamic Geosphere* whereas Module II has three chapters and 18 activities about *Understanding Your Environment*. Modules I and II were field tested for this study. All the activities are designed with the 5-E learning cycle model in which inquiry skills are promoted.

Activity begins with *Chapter Challenge* (Engage) where students read and discuss a scenario that presents a community-based issue to solve through Earth Science and Inquiry. Teachers allow students to share their current thinking openly without a closure. Second is *Think About It* (Engage) where students answer open-ended questions that set the context for an activity and provide the teacher with a pre-assessment of students' ideas. Teachers allow students to share their ideas openly. Third, *Investigate* (Explore) in which students collaborate on an inquiry activity that requires hands-on work, literature or web research, or fieldwork. Fourth is *Reflecting upon the Activity and the Challenge* (Explain). In this stage, students read a brief summary of the main ideas explored in the investigation and their relationship to the chapter challenge. Fifth, *Digging Deeper* (Explain) lets students read text, illustrations, and photographs that explain concepts explored in the investigation.

Terms are defined and clarified here. Teachers provide further information and clarification of concepts through lecture, slides, videos, or laser disk presentations. Sixth is *Understanding and Applying What You Have Learned* (Elaborate). Students respond to questions that check their understanding of key principles and concepts (learning goals) for the activity. Teachers review student responses and use the questions to further probe and hone understanding of key learning goals. Seventh is *Preparing for the Chapter Challenge* (Elaborate/Evaluate) in which students put their investigative results into the context of the challenge by preparing or organizing their work as it relates to their final products. Teachers review student performance in terms of its consistency with criteria. Eighth is *Inquiring Further* (Elaborate/Evaluate). Students are presented with options for deepening their understanding within the activity whereas teachers promote and encourage further inquiry.

At the end of the activity, students present their conclusion to the Chapter Challenge and teachers use the assessment criteria to assess the extent to which student work demonstrates mastery of concepts and skills. With laboratory activities and questioning strategies stressed in every class of EarthComm, it is notable that the NSES visions remain as a main structure of lessons (excerpt from Park, et al., 2005).

EarthComm Teacher Enhancement Workshop

All the teachers who taught Modules I and II for this study attended a week long training workshop. The workshop was designed and conducted for the field test teachers to become immersed in the design characteristics and innovative features of the EarthComm curriculum. The workshop covered many topics, issues, and concerns related to effective science teaching and learning, curriculum design and field testing. The workshop started with a general overview of the workshop and the EarthComm curriculum.

The first session included an overview of the National Science Education Standards. The remaining workshop sessions each had particular pedagogical focus. Each of the pedagogical focuses was linked to an

instructional component of the EarthComm curriculum. This linkage between pedagogy and curricula served to not only provide participants with the theory or rationale behind the design and intent of the curricula, it also allowed for the participants to glean an overview of the curricula's instructional model was provided. The model is based on sound constructivist based ideas of content delivery. This model was discussed in terms of: the theoretical underpinnings of constructivism: the relationship between the constructivist philosophy and hands-on instruction and: the relationship to the Teaching Standards in the National Science Education Standards. This discussion was carried out using the curriculum as the vehicle to deliver the pedagogical content - in this particular case, aspects of Module I, Volcanoes and Your Community. In other words, the discussion took place using concrete examples of activities, within a chapter of EarthComm, that were based on the model. Pedagogical content was delivered using the EarthComm curricula as the instructional vehicle.

Other sessions were devoted to topics such as thematic instruction, teaching to multiple learning style, cooperative learning and alternative assessment. Aligned with these topics were instructional issues specific to the EarthComm curricula. For example, teachers practiced with the delivery of numerous activities. Teachers who were selected to field test certain units worked collaboratively and immersed themselves into the activity from both a teaching and learning perspective. Teachers were allowed time to work as teams in gleaning an overview of particular chapters.

Then they performed some of those activities and presented summaries of chapters and activities to the entire group. To promote the use of local resources, activities were performed that served to utilize local resources. For example, Chapter 2 has seismographs activities that were discussed in conjunction with a local seismograph laboratory. A field trip to explore local geological formations aided in the discussion of the content numerous chapters devoted to geological land forms. All the while, as teachers participated in

these activities, ties to the EarthComm curriculum were emphasized and discussed in terms of how the teachers use a variety of resources to support the implementation of the curriculum. A week-long workshop ended by providing lines of communication including Webboard, emails, and online web homepage to ensure a continuous line of communication and feedback between all parties including participants and project personnel (excerpt from Park, et al., 2005).

Methodology

Sample Selection: Teachers and Students

A group of EarthComm teachers reported their student achievement scores from Module I (N = 24) and Module II (N = 16). Of the group, nine teachers (N = 9) were overlapped in both Modules I and II. Therefore, there were thirty one teachers all together who participated in the study. These thirty one teachers had various teaching backgrounds, i.e., Earth science, physics, and biology. Their teaching experiences varied from 2 years to 33. The sample group of teachers was employed in different communities across the United States, including both rural and urban areas. About 1,000 students

participated in the study in 24 classrooms for Module I. The class size ranged from 12 to 120 students. There were 40 students per class on average for Module I. On the other hand, the Module II involved 428 students from 16 classrooms. The range for class size was from 9 to 86 students. The average number of students was 27 per class for Module II. The overall pool of students came from a variety of backgrounds in terms of socioeconomic status, school size, and school location. Although the study sample did not completely represent all major geographic areas of the United States, they were widespread enough to signify an extensive range of students.

Tables 1 and 2 show the number of the sample teachers and students who participated in this investigation. As reported in Table 1, data gathered for Module I were obtained from new teachers who have 2-5 years of teaching experiences. Also, a large portion of the field-test teachers for Module I worked in urban or rural settings.

As indicated in Table 1, the students who participated in the pretest did not match the students in posttest for Module I. However, the students for Module II were matched (see Table 2). Thus, the means of student scores were used to conduct a statistical analysis

Table 1. Demographic Information Concerning Teacher Participants Involved with Module I

Region	Number of Teachers	Teaching Experience (yrs)			Community of School			Number of Students	
		1-5	6-15	16+	U ^a	S ^b	R ^c	Pretest	Posttest
State A	5	3	0	2	0	4	1	335	300
State B	9	4	1	4	7	0	2	381	343
State C	8	2	4	2	2	1	5	191	170
State D	2	2	0	0	1	0	1	41	40
Total	24	11	5	8	10	5	9	948	853

^aU indicates Urban, ^bS Suburban, and ^cR Rural.

Table 2. Demographic Information Concerning Teacher Participants Involved with Module II

Region	Number of Teachers	Teaching Experience (yrs)			Community of School			Number of Students	
		1-5	6-15	16+	U ^a	S ^b	R ^c	Pretest	Posttest
State A	5	1	2	2	1	2	2	160	160
State B	3	0	1	2	1	0	2	94	94
State C	6	1	3	2	3	2	1	124	124
State D	2	1	0	1	1	0	1	50	50
Total	16	3	6	7	6	4	6	428	428

^aU stands for Urban, ^bS for Suburban, and ^cR for Rural.

rather than using raw scores. This step made it possible to accomplish the intended statistical analysis. However, the power of the test was weakened by using means rather than raw scores for each student (Hinkle et al., 1979).

Instrumentation

Student Achievement Test

Tests with items developed separately for Modules I and II were designed to study improvement of student performance. American Geological Institute staff and an independent evaluator developed the test items specifically to evaluate the EarthComm program. Both tests for Modules I and II included 23 multiple choice questions that promote applying what the students learned to their local community's needs that characterized the NSES recommendations (relevance, systems, community, and inquiry). This test was administered to students by classroom teachers before and after the implementation of both EarthComm modules in each classroom. The reliability (KR-20) for Module I is $r = 0.55$ and the reliability (KR-20) for Module II is $r = 0.52$. Samples of the question for

Module test are presented below. The questions' foci in Module I and II tests center on relevance, community, and inquiry application.

Question Sample 1:

If you were planning to develop land along an active fault zone, which type of development would be LEAST affected by the geologic hazards of the site?

- parkland for a golf course
- single-family homes
- an apartment or condominium complex
- a mall shopping center

Question Sample 2:

The map below shows the earthquake risk for the United States. Which of the following explains why areas in Missouri and South Carolina are at such high risk?

There are very frequent, large earthquakes at these locations.

Both Missouri and South Carolina lie along boundaries of crustal plates.

There has been earthquake activity recorded at these locations in the past.

No earthquakes have occurred at these locations but statistically they are due one.

Ebel's Relevance Guide

Ebel developed a classification scheme for taxonomy of educational achievement (Ebel and Frisbie, 1991). This scheme was employed to classify test questions into six categories with a focus on higher-order thinking skills. Significant changes between pre- and post- test scores with respect to their higher-order skills were also determined. The six categories of Ebel's Relevance Guide (1956) are Content Detail, Vocabulary, Factual, Generalization, Explanation, and Application.

Data Collection

Multiple Variables and Student Achievement Scores

With two tests items for Modules I and II, EarthComm Field Testing Feedback Survey that included demographic information about both schools and participants was sent out to the participating teachers and students before the test of two modules. Information of the multiple variables is three fold: school size, school location (rural, urban, and suburban), and teacher teaching experiences. After the completion of each module, the Feedback Survey was filled out and returned with pre- and post- test scores from each of the participating teachers.

EarthComm and Higher Order Thinking Skills

Data regarding the relationship between EarthComm and higher-order thinking skills were collected from two tests through using Ebel's relevance guide (Ebel, 1956; Ebel and Frisbie, 1991). Two different kinds of tests were given to students for EarthComm Module I and Module II. Test items were classified according to Ebel's classification scheme by two raters. Two raters, one with a Ph.D. in measurement and the other a doctoral student in science education, were trained for 15 hours concerning Ebel's classification scheme until 90% of agreement was attained with sample items

suggested by Ebel (1956). Two raters independently analyzed 23 test-items for both Modules I and II. The agreement was 86.9% for Module I and 91.3% for Module II. Higher-order thinking levels parallel Ebel's generalization, explanation, and application levels while lower-order thinking skills are associated with content detail, vocabulary, and factual.

All the items were found as high-level questions according to the Ebel's classification scheme. The ratings include: 11 items generalization, 3 explanation, and 9 application for the Module I test; 10 items generalization, 8 explanation, and 5 application in the Module II test. Using Ebel's taxonomy, all items were classified as higher-order thinking levels for all items for both Modules I and II.

Method of Data Analysis

For the analysis of student performance, the pre- and post- test answer sheets of students were scanned through a Scantron machine and the outcomes recorded in a text file. Using the SPSS statistical package, the pre and post- test outcomes were analyzed regarding the socioeconomic status, school size, community type (i.e., urban/rural, percentage of minority status, and teaching experience of teachers. Student scores were analyzed for Module I and Module II, respectively. The results of Module I and Module II were compared in terms of the factor that impacted students' achievement scores. A two-way ANOVA using each component of the above factors was used to determine statistically significant changes between pre- and post- test scores between the two groups of students.

Concerning the assessment of higher-order thinking skills, Ebel's relevance guide was used to classify test items developed and used for EarthComm evaluation. The questions classified as requiring higher-order thinking skills were grouped and analyzed to examine the effect of the EarthComm implementation. The student scores for pre- and post- test were grouped accordingly and analyzed by the paired t-test to find statistically significant effects. This procedure was followed for Module I and Module II respectively. The

results of pre- and post- test scores for Module I and Module II were compared with each other to check similarities and differences in improvement concerning the inclusion of higher-order thinking skills.

Results

The data gathered for this study were analyzed to evaluate student achievement from several perspectives, including socio-economic status of the community where students live and use of higher-order thinking skills in the materials. The results obtained from answering the research questions follow:

Description of Module I and Module II Tests

Tests for Modules I and II consisted of 23 items respectively which were prepared by the AGI curriculum authors. For Module I, twenty-four individual teachers and over 900 students participated

in the investigation. For the test for Module I, student mean scores for the pretests ranged from 6.00 to 13.08 while mean scores for the posttests were spread widely from 8.00 to 17.05 (see Table 3). Test item analysis for all students involved indicated that the mean score for the pretest was 8.99 with the reliability (KR-20) of $r=0.55$. The mean difficulty of the pretest was 39.09 with a low mean discrimination index (0.30). This means that 39% of all students on average could be expected to get correct answers when they completed. The discrimination is the correlation of the item score with the total test score. A value of mean discrimination index 0.30 for the test means that there is a 0.30 correlation between scoring well on the given item and scoring well on the test as a whole. On the other hand, the mean score for the posttests was 11.70. Item analysis statistics indicated that the reliability of the posttest was $r=0.73$ with 50.91 for the mean difficulty, and 0.38 for the mean discrimination.

Table 3. Student Data for Teachers Who Participated in EarthComm Module I

Teacher	Student Pretest			Student Posttest			t
	N	Mean	SD	N	Mean	SD	
T1	39	9.14	2.71	37	11.06	3.45	2.173*
T2	87	9.49	2.78	65	15.00	1.97	11.683*
T3	35	11.54	2.76	33	14.00	2.22	4.259*
T4	95	10.76	2.74	94	17.04	3.04	15.069*
T5	29	11.28	3.51	28	12.64	2.70	1.676
T6	21	9.57	3.08	21	17.05	2.97	8.630*
T7	10	8.10	3.56	11	11.09	2.27	3.161*
T8	26	11.58	4.31	25	15.84	3.48	3.995*
T9	41	8.00	3.04	35	9.69	3.26	2.029
T10	76	7.86	2.44	70	9.46	3.43	3.395*
T11	98	8.34	2.50	95	9.96	2.98	3.864*
T12	35	10.49	3.70	35	12.11	3.08	1.806
T13	42	8.24	3.37	27	10.26	4.09	2.551*
T14	15	10.07	3.75	17	13.53	2.68	2.747*
T15	86	8.52	3.09	79	9.72	2.98	2.198*
T16	24	8.58	4.05	24	10.71	4.45	1.700
T17	19	8.89	2.95	17	10.94	2.65	1.837
T18	14	7.43	1.59	15	8.93	3.43	1.470
T19	51	9.99	3.39	35	12.80	3.33	3.470*
T20	12	6.00	1.87	12	9.25	4.44	2.288*
T21	29	8.90	3.02	21	12.86	2.73	4.413*
T22	12	13.08	3.48	9	14.11	5.26	0.267
T23	24	6.83	2.87	21	8.00	3.10	1.326
T24	28	7.82	2.16	27	8.37	2.64	0.746

*Significant at $p < 0.05$.

Table 4. Student Data for Teachers Who Participated in EarthComm Module II

Teachers	Student Pretest			Student Posttest			t
	N	Mean	SD	N	Mean	SD	
T1	9	7.89	2.71	9	6.00	2.60	-1.434
T2	86	7.31	2.77	86	17.70	3.48	21.749*
T3	34	10.09	2.67	34	11.47	3.28	1.929
T4	64	6.39	2.52	64	7.11	3.14	0.750
T5	17	9.06	2.63	17	11.29	2.62	3.351*
T6	13	6.31	2.53	13	16.54	3.84	10.493*
T7	18	5.39	2.25	18	7.39	3.58	2.034
T8	22	8.23	2.67	22	12.50	3.29	6.587*
T9	38	6.45	2.81	38	7.08	3.23	0.996
T25	22	8.32	2.23	22	11.36	2.30	4.949*
T26	9	8.78	1.56	9	9.22	3.11	0.294
T27	12	7.29	2.36	12	9.00	3.65	1.395
T28	18	6.39	2.52	18	7.11	3.14	0.750
T29	33	5.30	1.86	33	5.76	1.75	0.965
T30	22	6.59	1.87	22	8.86	2.98	3.321*
T31	11	5.55	2.16	11	7.73	2.28	2.782*

*Significant at $p < 0.05$.

For Module II, sixteen individual teachers and over 500 students were involved in the study. For the Module II test, student mean scores for the pretests ranged from 5.30 to 10.09 while mean scores for the posttests were widely spread from 5.76 to 17.70 (see Table 4). Test item analysis for all students involved indicated that the mean score for the pretests was 8.01 with the reliability (KR-20) of $r = 0.52$, 34.74 for mean difficulty, and 0.29 for mean discrimination while mean posttest score was 11.20 with 0.85 for reliability, 48.65 for mean difficulty, and 0.47 for mean discrimination. The mean score of the posttests rose by three points from the mean score of the pretests with a high reliability for the test ($r = 0.85$).

Effectiveness of EarthComm Related to Various Demographic Variables

The effectiveness of the EarthComm program was investigated in terms of statistically significant changes in student achievement across demographic variables, including size of school, urban and rural settings, and the experience level of teachers. Student achievement scores in pre- and post- test were used as dependent variables. Generic information from participating teachers concerning the aforementioned factors was

obtained for Module I and II through the EarthComm Field Testing Instruments used to collect the teachers' feedback concerning the field-testing of EarthComm materials.

Module I. School size was quite diverse. The number of students enrolled ranged from 350 to 2900. Only five schools (20.8%) had more than 2000 students while 4 schools (16.7%) had fewer than 500 students enrolled. Most of the students who participated in Module I lived in urban settings (41.7%) and in rural areas (37.5%). Fewer students (12.5%) went to schools in suburban regions. The years of teaching experience varied considerably, ranging from 2 to 33 years. However, most were the new teachers who had fewer than 5 years of teaching experience (45.8%). Nearly 21% Module I teachers had more than 20 years experience teaching science.

Module II. School size ranged from 169 to 2300 for Module II. Five schools (31.3%) for Module II enrolled fewer than 500 students while 2 schools (12.5%) enrolled over than 200 students. Schools were mostly located in urban areas (37.5%) and rural areas (18.8%). Teachers varied in terms of number of years of teaching experience. The percentage of teachers who had fewer than 5 years of all science teaching

experience was 18.8% for Module II. The percentage of teachers with more than 20 years experience teaching science was 25.0% for Module II.

Large vs. Small Schools

Large schools were defined as schools which enrolled more than 2000 students and schools with fewer than 500 students were labeled as small. Of the 24 teachers for Module I, four teachers were in large schools (T10, T13, T17, and T25) while 5 teachers taught in small schools (T4, T5, T12, T14, and T21). Table 9 illustrates the results of two-way ANOVA for repeated measures, which tested the impact of school size on student achievement after the implementation of the EarthComm Module I. As reported in Table 9, the difference in student achievement between the large schools and the small schools was not significant at the level of 0.05. Neither is there significant interaction.

The difference between pre- and post- test scores within groups was highly significant ($p < 0.001$). The mean of student scores from the large schools was 9.14 ($SD = 1.64$) for the pretests and 11.63 ($SD = 2.87$) for the posttests while the mean for students from the small schools was 10.30 ($SD = 0.90$) for the pretests and 13.64 ($SD = 1.97$) for the posttests.

These results indicate that student achievement was not significantly affected by the EarthComm implementation in terms of the size of school although student mean scores were significantly increased for students in both groups.

Of the 16 teachers who field-tested Module II, two teachers (T8 and T26) were from large schools while 5 teachers (T4, T5, T25, T27, and T28) taught in small schools. Table 10 indicates the results of two-way ANOVA for repeated measures which tested the impact of school size on student achievement for Module II. As indicated in Table 10 for Module II data, no significant changes in student achievement were found in terms of school size except that the difference between pre and posttest was statistically significant ($p < 0.05$).

This outcome is the same as the results reported for Module I. The mean for student scores for large schools was 8.52 ($SD = 2.79$) for pretests and 11.42 ($SD = 3.34$) for posttests while the mean for students from small schools was 8.70 ($SD = 3.21$) for pretests and 11.87 ($SD = 4.42$) for posttests.

In summary, the results for implementing Modules I and II indicate that student achievement was not affected by the size of schools when EarthComm was taught.

Table 9. Two-way ANOVA for repeated measures concerning the impact of school size on student achievement scores for Module I

Source	SS	df	MS	F	P
Between	11.16	1	11.16	1.87	0.21
Error (B)	41.86	7	5.98		
Within	37.62	1	37.62	25.99	0.00*
Error (W)	10.13	7	1.45		
Interaction	0.81	1	0.81	0.56	0.48

*Significant at $p < 0.05$.

Table 10. Two-way ANOVA for repeated measures concerning the impact of school size on student achievement scores for Module II

Source	SS	df	MS	F	P
Between	4.91	1	4.91	0.36	0.55
Error (B)	2161.69	157	13.77		
Within	460.52	1	460.52	50.21	0.00*
Interaction	0.90	1	0.90	0.06	0.80
Error (W)	2195.46	157	13.98		

*Significant at $p < 0.05$

Table 11. Two-way ANOVA for repeated measures concerning the impact of school setting on student achievement scores for Module I

Source	SS	df	MS	F	P
Between	1.61	1	1.61	0.20	0.66
Error (B)	140.15	17	8.24		
Within	60.56	1	60.56	39.64	0.00*
Interaction	2.58	1	2.58	1.69	0.21
Error (W)	25.97	17	1.53		

*Significant at $p < 0.05$ **Table 12.** Two-way ANOVA for repeated measures concerning the impact of school setting on student achievement scores for Module II

Source	SS	df	MS	F	P
Between	54.07	1	54.07	2.86	0.09
Error (B)	4865.01	257	19.29		
Within	295.27	1	295.27	23.24	0.00*
Interaction	333.36	1	333.36	26.24	0.00*
Error (W)	3265.29	257	12.71		

*Significant at $p < 0.05$

Urban vs. Rural Communities

For Module I, ten teachers (T4, T6, T9, T13, T15, T16, T17, T18, T19, and T20) were identified as teaching in urban schools and 9 teachers (T1, T5, T7, T12, T14, T21, T22, T23, and T24) were found to teach in rural schools. Table 11 indicates the results of two-way ANOVA for repeated measures which tested the impact of school type on student achievement. As presented in Table 11, no significant changes in student achievement scores were found in terms of school type. But the difference between pre and posttest was statistically significant ($p < 0.05$). The mean for all students from urban schools was 8.59 ($SD = 1.33$) for pretests and 11.64 ($SD = 3.05$) for posttests while the mean score for all students from rural schools was 9.52 ($SD = 1.93$) for pretests and 11.53 ($SD = 2.15$) for posttests.

Of the 16 teachers for Module II, six teachers (T4, T6, T9, T26, T29, and T31) were employed in urban schools while 3 teachers (T25, T27, and T28) taught in rural schools. Table 12 displays the results of two-way ANOVA for repeated measures which tested the impact of location of school on student achievement. As shown in Table 12, the difference between pre- and post- test was statistically significant ($p < 0.05$)

and the interaction was highly significant ($p < 0.001$). However, no significant changes for student scores were found in terms of urban vs. rural schools.

The mean for student scores from urban schools was 7.24 ($SD = 3.04$) for pretests and 10.50 ($SD = 5.28$) for posttests, which was statistically significant ($t = 7.69$). The mean for student scores from rural schools was 8.24 ($SD = 3.32$) for pretests and 8.14 ($SD = 3.22$) for posttests, which was not statistically significant ($t = -0.23$).

In summary, the outcomes from the analysis of Modules I and II data indicate that student achievement was not associated with the type of community in which the school was located.

Highly Experienced Teachers with Over 20 Years of Experience vs. Teachers with Fewer Than 5 years of Teaching Experience

Data concerning teacher experience were divided into two groups: One group included teachers with more than 20 years of teaching experience and the other group included teachers with fewer than 5 years of teaching experience.

For Module I, five teachers (T2, T15, T12, T13, and T9) were identified as teachers who had over 20

years of teaching experiences and 11 teachers (T10, T17, T18, T19, T5, T20, T24, T7, T16, and T11) were found to be teachers who had fewer than 5 years of teaching experience. Table 13 shows the results of a two-way ANOVA for repeated measures which tested the impact of teacher experience on student achievement. As shown in Table 13, the difference between student scores in classes taught by highly experienced teachers and student scores in classrooms taught by less experienced teachers was not significant. But the difference between pre- and post-test in the two groups was statistically significant ($p < 0.05$). No significant interaction was found for Module I. The mean score for students in classrooms taught by teachers with over 20 years in teaching experience was 8.95 (SD = 1.03) for pretests and 11.36 (SD = 2.26) for posttests while the mean for classes taught by teachers with fewer than 5 years of teaching experience was 8.84 (SD = 1.15) for pretests and 10.80 (SD = 1.53) for posttests.

For Module II, four teachers (T2, T9, T28, and T29) were identified as teachers with over 20 years of teaching experience and 3 teachers (T1, T5, and T7) had fewer than 5 years of teaching experience. Table

14 displays the results of two-way ANOVA for repeated measures which tested the impact of teacher experience on student achievement. As indicated in Table 14, the difference between the two groups was highly significant ($p < 0.001$) and the difference between pre- and post- test scores was also highly significant ($p < 0.001$) for Module II. In addition, the interaction was highly significant ($p < 0.001$).

The mean for achievement scores for all students in classrooms taught by experienced teachers with over 20 years in teaching was 6.82 (SD = 2.60) for pretests and 13.32 (SD = 6.51) for posttests, which was highly significant ($t = 11.78$). The mean score for classes taught by teachers with fewer than 5 years of teaching experiences was 7.32 (SD = 2.96) for pretests and 8.61 (SD = 3.71) for posttests, which did not differ significantly ($t = 2.19$).

In summary, the difference in student achievement scores between the two groups was not statistically significant for Module I but significant between the two groups of teachers for Module II. The differences found in statistical significance for student achievement between the two groups of teachers for both Modules I and II show different patterns.

Table 13. Two-Way ANOVA for Repeated Measures for the Test for the Impact of Teacher Experience on Student Achievement Scores for Module I

Source	SS	df	MS	F	P
Between	0.77	1	0.77	0.20	0.66
Error (B)	52.35	14	3.74		
Within	32.74	1	32.74	50.81	0.00*
Interaction	0.35	1	0.35	0.54	0.47
Error (W)	9.02	14	0.64		

*Significant at $p < 0.05$

Table 14. Two-Way ANOVA for Repeated Measures for the Test for the Impact of Teacher Experience on Student Achievement Scores for Module II

Source	SS	df	MS	F	P
Between	294.70	1	294.70	11.74	0.00*
Error (B)	4492.58	179	25.10		
Within	1012.87	1	1012.87	57.24	0.00*
Interaction	451.69	1	451.69	25.52	0.00*
Error (W)	3167.70	179	17.70		

*Significant at $p < 0.05$

Table 15. Differences for Higher-order thinking skills required for use of Modules I and II.

Module	Pretest Mean	Posttest Mean	SD	df	t-test
Module I	9.18	11.85	1.74	23	7.48*
Module II	8.03	11.20	6.12	428	11.67*

*Significant at $p < 0.01$.

Higher-Order Thinking Skills

One of the goals in science teaching is to provide students with the opportunity to think (NRC, 1996). When students are engaged in inquiry-based learning activities and are asked high level questions, then they are expected to use higher-order thinking skills.

Data were obtained and analyzed to determine significant changes in student achievement associated with higher-order thinking skills prior to and after the implementation of EarthComm in schools. Table 15 displays the results of the statistical analysis of differences for higher-order thinking skills found in Modules I and II.

The differences between pre- and post- test found both in Modules I and II are significant ($p < 0.01$). The low degree of freedom ($df = 23$) for Module I is associated with the data justification for the purpose of statistical analysis. This result indicates that students' higher-order thinking skills improved when using the EarthComm curriculum. But the differences between higher-order thinking skills taught in Modules I and II were not large enough to determine differences in the contribution of the two modules for cultivating high-order thinking skills.

Discussion

Effectiveness of EarthComm Related to Demographic Variables

The effectiveness of the EarthComm curriculum was studied in terms of statistically significant changes in student achievement when some demographic variables were considered. One unexpected outcome was found in Table 4. One of the data sets revealed student performance opposite what was expected, i.e., students scored higher on the pretest than they did on the posttest (see T1). Although the cause of this result

needs to be further investigated, this was likely because of the student demographics involved for T1. The students were from an inner city setting; students had many kinds of troubled backgrounds (social, emotional, and/or legal problems); they experienced difficulty in accomplishing the small group work (Personal communications with the project manager who made frequent visits to each school, 2000); most teachers who worked with students in an inner city tended to mention that students' low achievement resulted from their troubled socioeconomic circumstances. It is also possible that students did not take the test seriously because they knew that it did not count for a grade in the class. Many possible reasons for this outcome can be suggested but further research is needed to establish the actual result.

Large Schools vs. Small Schools

Student achievement scores for both Modules I and II were not significantly affected by the EarthComm program in terms of school size. There were no statistically significant differences in student achievement between large schools and small schools (see Tables 9 and 10). Only changes in student gains between pre- and post- tests within group were significant in Module I ($t = 5.36$) and Module II ($t = 7.46$). This result implies that teaching EarthComm will result in the same impact on student learning regardless of the size of schools.

Urban vs. Rural Communities

This study did not produce any significant differences in student achievement for both Modules I and II between schools located in urban and rural areas. In other words, student achievement was not significantly affected by the type of community from which students came (see Tables 11 and 12). Only changes

in student gains between pre- and post- tests within group were significant for Module I ($t = 6.26$) and Module II ($t = 6.34$). This result is desirable when the NSES visions are considered. EarthComm emphasizes the use of community-related issues and problems in which students are invited to understand concepts and application of knowledge about local resources. This feature is consistent with the NSES visions, "A community approach enhances learning: it helps to advance understanding, expand students' capabilities for investigation, enrich the questions that guide inquiry, and aid students in giving meaning to experiences (NRC, 1996, pp.46)."

Highly Experienced Teachers with Over 20 Years of Experience vs. Teachers with Fewer Than 5 Years of Teaching Experience

There was interest in this study to determine whether or not teaching experience impacted student achievement when EarthComm was taught. The percentage of teachers who had fewer than 5 years of science teaching experience was 45.8% for Module I and 18.8% for Module II. On the other hand, the percentage of teachers who experienced more than 20 years of science teaching was 20.8% for Module I and 25.0% for Module II. Thus, the distribution of years of teacher experience is typical normal except for Module I for when almost half of the group members were new teachers who possessed fewer than 5 years of experience. The difference in student achievement scores between the two groups was not statistically significant for Module I test while it was significant for Module II. The two results for the two modules conflict with each other (see Tables 13 and 14). The results of Module I are consistent with the studies of Ferguson (1998). Levels of teacher experience and preparation contribute to student gains across numerous school districts (Ferguson, 1998). Therefore, the more years of teaching experience and more education teachers have, the greater qualifications they will acquire. This will ultimately increase their teaching knowledge to improve their practices. This result contrasts with the outcome from the study of

Yager et al. (1988) where teaching experience was not found to vary among least and most effective teachers. The results from use of EarthComm Module II correspond with this regard. A statistically significant difference in student achievement was not found between teachers who have more than 20 years of experience and teachers having fewer than 5 years experience. This finding implies that, given strong beliefs in the vision of NSES, teachers at all levels of teacher experience can effectively implement reform curricula and result in improved student learning.

Higher-order thinking skills

Higher-order thinking skills are promoted by facilitating activities and questions (NRC, 1996; Ajeyalemi, 1993; Hurd, 1992; Aikenhead, 1988). The results of this study indicate that student higher-order thinking skills were significantly improved in EarthComm classes. The differences in student scores between pre- and post- tests regarding higher-order thinking skills are highly significant ($p < 0.001$) (see Table 15). Students should be given opportunities to apply what they learn in everyday life and develop higher-order thinking skills (NRC, 1996; Weiss, 1993). Within the EarthComm curriculum, students are frequently asked questions across a range of cognitive levels which require that students explain their thinking and learning. Tobin and Fraser (1990) assert that asking questions and probing student responses for clarification and elaboration are important actions to stimulate thinking and to nurture thinking skills. EarthComm is successful in teaching students and enhancing their learning of higher-level questioning. Although a quantitative analysis of student higher thinking skills shows a statistical significance, it is yet to be confirmed with student observations and interviews about their higher order thinking skills about specific science tasks.

Conclusions

EarthComm did not significantly impact student achievement differentially in schools of varying sizes.

There was no statistically significant difference in student achievement between large schools and small schools. EarthComm did not produce any significant differences in student learning in schools in urban and rural areas. In other words, student achievement was not significantly affected by the type of community in which the school is located. Considering the implementation of EarthComm in Korean context, these results may or may not be the same as what would be found in Korea because of the idiosyncratically different educational settings including the ethnicity of students, rural and suburban location, and school size. For instance, an area called an 'inner city' in the U.S. constitutes a complicated meaning of students' racial background. Yet, it may be similar if we look at the educational settings in terms of social, emotional, and legal problems when implementing a new curriculum. Thus, the culture of educational context must be carefully examined and considered when implementing the EarthComm.

Regarding the year of teaching experiences, the differences in student achievement scores between the highly experienced teachers group (over 20 years of teaching experience) and new teachers group (fewer than 5 years of teaching experience) were not found to be statistically significant for the Module I test but significant for Module II. It is noted that the two results seemingly conflict with each other similar to other research findings (Ferguson, 1998; Yager et al., 1988).

Student higher-order thinking skills, on the other hand, were significantly improved as a result of studying with EarthComm. EarthComm classes in Modules I and II were taught with activities in each chapter as designed with Learning Cycle, inquiry-based 5E's model (Engagement, Exploration, Explanation, Elaboration, and Evaluation).

These findings of EarthComm provide a niche for earth science education at the high school level since student achievement scores with EarthComm did not differ from school to school in different locations, from teacher to teacher with different teaching experiences. Especially it is noted that students

achieved their higher order thinking skills when it is taught in ways that the NSES envisions.

The findings of this study imply that curriculum developers and teachers and administrators are to take into account issues and factors, i.e., school location, school size, teachers' teaching experiences, and content that impact student achievement scores when they select for use of the Earth science curriculum.

Limitation of The Study

The analysis of data in this study was limited by the availability of the teacher and student data from the field test sites. Not all of the EarthComm field-test teachers returned student achievement scores and the EarthComm Field Testing Surveys about the implementation of EarthComm. Thus, the comparison and analysis in this investigation was only possible with matched data of teachers' surveys and students' scores. In fact, generalizability was limited by the small sample size of teachers surveyed. Each module was field tested by only 20-30 teachers. Outcomes from a sample of this size can hardly be generalized for the entire population of earth science teachers in the U.S.

References

- Ajeyalemi, D.A., 1993, Teacher Strategies Used by Exemplary STS Teachers. In Yager R.E. (ed.), *What Research Says To the Science Teacher*, Vol. 7: The Science, Technology, Society Movement (49-52). Washington, D.C.: The National Science Teachers Association.
- Aikenhead, G.S., 1988, *Teaching Science Through a Science-Technology-Society-Environment Approach: An Instruction Guide*. SIDRU Research Report No. 12. Saskatchewan, Canada: Institutional Development and Research Unit, Regina University (ERIC Document Reproduction Service No. ED 323 111).
- American Geological Institute., 1999, *Earth System Science in the Community (EarthComm)*, Washington, D.C.
- Ebel, R.L., 1956, *An Outline for a Course in the Construction and Use of Classroom Tests*. The State University of Iowa, Iowa City, IA.
- Ebel, R.L. and Frisbie, D.A., 1991, *Essentials of Educational Measurement*, 5th ed. Englewood Cliffs, Prentice Hall, Inc., NJ, USA.

- Exline, J.D., 1998, The National Survey with Respect to Earth Science Education at State, District, and School Levels. Paper presented at the American Geological Institute Conference, Toronto, Canada.
- Ferguson, R.F., 1998, Can Schools Narrow the Black-White Test Score Gap? In Jencks, C. and Phillips, M. (eds.). *The Black-White Test Score Gap* (318-374). Washington, D.C.: Brookings Institute Press.
- Hinkle, D.E., Wiersma, W., and Jurs, S.G., 1979, *Applied statistics for the behavioral sciences*. Rand McNally College Publishing Co., Chicago, USA.
- Hurd, P.D., 1992, *Achieving Reforms in Science Education*. Stanford, CA: Stanford University.
- National Research Council., 1996, *National Science Education Standards*. Washington DC: Author.
- Park, D., 2006, Curriculum Reform Movement of Science Education in the U.S.: A Case of Earth Science Curriculum. *Journal of the Korean Earth Science Society*, 27 (7), 730-744.
- Park, D., 2005, Differences between a Standards-Based Curriculum and Traditional Textbooks in High School Earth Science. *Journal of Geoscience Education*, 53 (5), 540-547.
- Park, D., Yager, R., and Smith, M., 2005, Implementing EarthComm: Teacher Professional Development and Its Impact on Student Achievement Scores in a Standards-Based Earth Science Curriculum. *Electronic Journal of Science Education*, 9 (3). [Online]. Available: <http://wolfweb.unr.edu/homepage/crowther/ejse/ejsev9n3.html> [2005, March].
- Tobin, K.G. and Fraser, B.J., 1990, What Does It Mean To Be an Exemplary Science Teacher? *Journal of Research in Science Teaching*, 27 (1), 3-25.
- Weiss, I.R., 1993, Science Teachers Rely on the Textbook. In Yager, R.E. (Ed.) *What Research Says to the Science Teacher*, Volume Seven: The Science, Technology, Society Movement. Washington D.C.: NSTA.
- Yager, R.E., Hidayat, E.M., and Penick J.E., 1988, Features Which Separate Least Effective from Most Effective Science Teachers. *Journal of Research in Science Teaching*, 25 (3), 165-177.

Manuscript received: 2 July 2007

revised Manuscript received: 31 August 2007

Manuscript accepted: 21 September 2007