An Analysis of Frequency of Hands-On Experience and Science Achievement

Patricia M. Stohr-Hunt

Department of Education, University of Richmond, Richmond, Virginia 23173

Abstract

A variance analysis of the relation between the amount of time students spent experiencing hands-on science and science achievement was performed. Data collected by the National Education Longitudinal Study of 1988 on a nationally representative sample of eighth-grade students were analyzed. Student achievement in science was measured by a cognitive test battery developed by the Educational Testing Service. Information regarding the frequency of hands-on experience was collected through a self-administered teacher questionnaire, which included a series of questions specific to the science curriculum. From the analysis it was concluded that significant differences existed across the hands-on frequency variable with respect to science achievement. Specifically, students who engaged in hands-on activities every day or once a week scored significantly higher on a standardized test of science achievement than students who engaged in hands-on activities once a month, less than once a month, or never.

The successful flight of the Soviet Sputnik in 1957 gave renewed impetus to the curriculum reform movement begun toward the end of World War II (Klopfer & Champagne, 1990). Critics of the US educational system demanded a shift in the goals of education, particularly in the areas of mathematics and science. In response to this shift, science curriculum development boomed during the next 20 years. Many new programs, referred to as "alphabet" science programs, thrived through financing by federal agencies, such as the National Science Foundation, and private foundations (Kyle, Shymansky, & Alport, 1982). Although these programs differed in their organization and style, each placed a great deal of stress on, and value in, hands-on activities that were oriented towards discovery learning.

One of these elementary programs, Science—A Process Approach (SAPA), was designed to emphasize the laboratory method of instruction, and focused on ways of developing basic skills in the processes of science. These processes included observing, classifying, measuring, predicting, and other important skills needed to conduct science investigations. Developed by the American Association for the Advancement of Science's (AAAS) Commission on Science Education, the Commission believed that science was best taught through inquiry, and that an understanding of the scientific approach to gaining knowledge of the world was a fundamental part of a child's education (Lee, 1967). In a policy statement pertaining to elementary science, the Commission wrote,

> Science is more than a body of facts, a collection of principles, and a set of machines for measurement; it is a structured and directed way of asking and answering questions. It is

no mean pedagogical feat to teach a child the facts of science and technology; it is a pedagogical triumph to teach him these facts in their relation to the procedures of scientific inquiry (Kessen, 1964, p. 5).

As SAPA and other similar programs emphasized process skills, it then followed that there was a corresponding de-emphasis on content. Students were rarely required to learn and remember particular facts or principles about the phenomena they were exploring because the developers believed that these concepts would be incidentally learned. However, this incidental learning was not the primary objective of the program. It was the development of general intellectual skills (Gagne, 1967) and the mastery of scientific processes (Washton, 1974) that were of prime importance. Considering the various stages of cognitive development described by Piaget, SAPA encouraged the development of skills, beginning with highly specific and concrete forms, and increasing in generality as students worked through a series of tasks (Dunfee, 1967). This structured sequence of tasks was extremely important for attaining the goals of the program, because even though individual activity goals were modest, SAPA's goal of intellectual development was expected as a result of the cumulative effects of the orderly progression of learning activities (Gagne, 1967).

Given SAPA's lack of emphasis on science content, how could its developers have expected the implementation of such a program to improve student achievement in science? Similar in theory to other activity-based programs (e.g., SCIS, ESS) that emerged during this period, the developers of SAPA believed that by teaching students skills for thinking and acting in science they would not only improve their achievement and understanding of science, but would also learn valuable skills that could generalize to new subjects and situations (Gagne, 1967).

Today, most of these activity-based programs have disappeared as intact programs from the science curriculum (Shymansky, Kyle, & Alport, 1982), yet the current reform movement in science instruction espouses a curriculum that uses "manipulatives" and engages students in discovery and problem-solving activities. In an era when international studies have reported that US school children consistently rank below their international counterparts in science achievement (Jacobson, Doran, Humrich, & Kanis, 1988; Lawson, 1990), the question remains, how effective are hands-on programs in improving student achievement in science, assuming that hands-on programs are being implemented?

In 1975, Wideen compared the effects of SAPA and a traditional textbook program of instruction on science processes and achievement. Results from 25 classrooms indicated that students in the SAPA program scored higher on science achievement and process tests than the students in the traditional program. Another SAPA study conducted 1 year later produced conflicting results, indicating that students in the SAPA program did not differ significantly from those students who were taught using a traditional textbook program (Davis, Raymond, Mac-Rawls, & Jordan, 1976).

Vanek and Montean (1977) compared the effects of another "alphabet" science program, Elementary Science Study (ESS), and a textbook approach on student achievement in science and classification skills. The differences between the treatment (ESS) and control groups were not shown to be statistically significant. On the basis of these results, the authors concluded that neither the method of instruction nor the materials employed made a difference in the science achievement or classification skills of the students (Vanek & Montean, 1977).

In 1983, Bredderman conducted a meta-analysis of research studies involving three elementary activity-based programs: SAPA, ESS, and the Science Curriculum Improvement Study (SCIS). Analysis of the results of 57 studies on nine different measures of student performance,

including science processes and content, indicated that for all measured outcomes, students in activity-based programs scored significantly higher than students in traditional science programs. Using research literature and data aggregation procedures similar to Bredderman's, Shymansky, Kyle, and Alport (1983) conducted a meta-analysis of 27 activity-based science programs, within the elementary, junior high, and high school curricula. Their analysis of the results of more than 100 experimental studies on 18 different measures of student performance, broadly categorized into areas such as achievement, process skills, and analytic skills, showed the greatest gains in achievement and process skill development for students who received instruction from activity-based programs. In 1986, these results were reanalyzed using refined statistical procedures (Shymansky, 1989). Data from the reanalysis showed that students in hands-on programs outperformed their traditional elementary school counterparts by 9 percentile points on a composite performance measure, taking in to account achievement, student perceptions of science, and other factors. In fact, from the data, they concluded that the new elementary science programs were more effective in enhancing student achievement and problem-solving skills than traditional programs (Shymansky, 1989).

The Intermediate Science Curriculum Study (ISCS), a program included in the meta-analysis by Shymansky et al., was examined again by Atash and Dawson (1986). A meta-analysis conducted using the results from 10 doctoral studies identified such outcome variables as science achievement, process skills, and analytical skills. Although the mean effect size across all outcome variables was .09, the mean effect for achievement was $-.45$, indicating that students in the ISCS programs achieved almost one half a standard deviation below students in traditional science programs.

Activity-based science programs have been evaluated by a number of different characteristics. Results of research to determine the effectiveness of activity-based programs by examining a variety of measures has been statistically significant (Bredderman, 1983; Glasson, 1989; Shaw, 1983; Shymansky et al., 1983; Siegel & Raven, 1971; Wideen, 1975). Today, the growing consensus of educators regarding the goals of science education appears to be stressing the education of scientifically literate persons, or those individuals who (a) have a knowledge of science concepts and their applications, (b) use science processes in solving problems and making decisions, (c) understand the nature of science and the scientific enterprise, and (d) understand the interaction between science, technology, and society (Koballa, 1984). If this is a major goal of science education, research must not judge the effectiveness of activity-based programs solely on scores of science achievement, but also in terms of problem-solving skills, process skills, and analytic skills.

Statement of the Problem

Research on the effectiveness of activity-based science programs continues to examine many different measures of student performance, yielding a variety of results. Some research implies that such programs are highly successful, whereas other studies indicate that these methods are no better than traditional methods of instruction. Given these mixed results, this study was designed to examine the relationship between frequency of hands-on experiences and standardized science achievement scores. Specifically, do students who engage in hands-on activities frequently, defined as daily or once a week, score higher on tests of science achievement than those students who engage in hands-on activities infrequently, defined as once a month, less than once a month, or never? This hypothesis was tested empirically using data collected in the National Education Longitudinal Study of 1988 (NELS:88).

Method

Sample

Student participants for NELS:88 were chosen via a two-stage process. During the first stage, random sampling of schools from across the country resulted in 1052 participating schools, composed of 815 public schools and 237 private schools. Through random selection procedures in stage two, a sample of 24,599 eighth-grade students resulted, with approximately 24 students representing each sampled school (National Center for Educational Statistics, 1989). For a thorough description of the NELS:88 sample design, refer to *NELS:88 Base Year Sample Design Report.*

Data Collection

A cognitive test battery developed by the Educational Testing Service (ETS) was used to measure student achievement. The science component of this battery consisted of 25 multiple-choice items designed to assess science knowledge and scientific reasoning ability (National Center for Educational Statistics, 1989). ETS attempted to select items that were relevant to a typical eighth-grade science curriculum, but did not require "a great deal of isolated factual knowledge." Rather, the emphasis was to be on understanding concepts and the measurement of problem-solving skills (National Center for Education Statistics, 1991). Breakdown of test items into three categories called recall, comprehension, and higher (all levels above comprehension), a scheme loosely based on Bloom's taxonomy, revealed that more than half of the items were above recall. Further breakdown of items into subject areas revealed that almost half of the items required some knowledge of earth science, whereas the remaining items required either some knowledge of life science or physical science. Internal consistency measures based on coefficient alpha resulted in a reliability of .75 with a standard error of measurement of 2.91. Though the science component of the battery was less reliable than the other components, ETS attributed this modest reliability to the fact that some science factors were often not separable from mathematics factors and/or reading comprehension factors. For a detailed description of the test battery, refer to *Psychometric Report for the NELS:88 Base Year Test Battery.*

Results from a self administered teacher questionnaire provided information regarding the frequency of hands-on experience. Teachers were asked "How often do students conduct science experiments in class?" To ensure that teachers did not confuse student hands-on experience with teacher experience, teachers were also asked "How often do you demonstrate a science experiment in class?" (National Center for Education Statistics, 1989) Designed primarily to supplement student data by collecting information pertaining to specific eighth-grade students and the courses they were enrolled in, each student's data were supplemented by information collected from teachers responsible for teaching the student in two of the four test subjects. Therefore, these teachers were not randomly selected, but rather were first paired according to the subjects they taught, with one pair comprised of math and science teachers, and the other of English and social studies teachers. One teacher from each pair was then selected to respond to the questionnaire. For a thorough description of the NELS:88 teacher sample, refer to *NELS:88 Base Year Sample Design Report* (National Center for Educational Statistics, 1989). Teacher and student data are located on computer tape and can be accessed through the Quantitative Analysis Lab.

Procedures

Two separate analyses were computed using SPSS X: One using the school as the unit of analysis and the other using the student as the unit of analysis. An analysis of variance (ANOVA) of science achievement by frequency of hands-on experience was conducted for each unit of analysis. For the first analysis, student and teacher data were aggregated to the school level by calculating mean scores for frequency of hands-on experience and science achievement. These data were then matched on the basis of school identification numbers. For the second analysis, teacher data were matched to student data on the basis of student identification numbers. Once matched, all students missing teacher information regarding hands-on experience were excluded.

Results

The data were analyzed in two parts. At the school level, the one-way (five level) ANOVA for science achievement yielded a significant main effect [$F(4, 478) = 6.4, p < .001$]. Given this significant finding, the analysis proceeded to the student level. At the student level, the one-way (five level) ANOVA for science achievement yielded a significant main effect [$F(4, 10,313) = 48.2, p < .001$]. Differences in the relative values of F at the school and student level can be attributed to a substantial increase in variance as analysis at the student level includes almost three times as many test scores as the analysis at the school level.

Descriptive data for science achievement at the schools and student levels are presented in Tables 1 and 2, respectively. At both the school and student levels, there appears to be homogeneity of variance for all levels of the independent variable as the standard deviations for all five groups are comparable. At the student level, inspection of the means reveals that, for all means but one, the means of each group fell outside of the adjacent group's confidence interval. A t test for differences among means yielded a critical difference of .68, indicating that students who experienced hands-on activities frequently had significantly ($p < .001$) higher scores of science achievement than those students who experienced hands-on science infrequently (for a definition of the terms frequently and infrequently, refer to the statement of the problem).

Table 1
*Means, SD, and Confidence Intervals for Science Achievement
for Frequency of Hands-on Experience at the School Level*

Frequency of hands-on experience	*n*	$\bar{X}$	SD	95% Confidence intervals	
Every day	73	52.6	4.7	51.5	53.7
Once a week	226	50.5	5.1	49.8	51.1
Once a month	107	49.6	5.1	48.6	50.6
<Once a month	56	48.4	5.6	46.9	49.9
Never	17	49.7	4.8	47.3	52.2
Totals	479	50.3	5.2		

Table 2
Means, SD, and Confidence Intervals for Science Achievement for Frequency of Hands-on Experience at the Student Level

Frequency of hands-on experience	n	$\bar{X}$	SD	95% Confidence intervals	
Every day	1299	52.8	10.1	52.3	53.4
Once a week	4892	51.3	10.3	51.1	51.6
Once a month	2042	49.7	9.7	49.3	50.2
<Once a month	1373	49.4	10.1	48.9	49.9
Never	712	47.5	9.7	46.8	48.2
Totals	10318	50.7	10.2		

Limitations

In examining these data and results, we must consider carefully the limitations of this study. First, although the data collected by NELS:88 were nationally representative of students and schools, they were not representative of teachers. A peculiar sample of teachers resulted from the fact that many teachers completing surveys were responsible for teaching science to more than one student in the sample. By matching teachers to students, an assumption of ANOVA that data were sampled independently was violated. Compounding this limitation was the problem of specifically matching teachers to students on the basis of class. Although teachers filled out survey information for every class they taught, it was not possible to determine exactly which class a particular student was in. This forced the assumption that for each teacher there was little variance across classes taught in the frequency of activities scheduled. In other words, it was assumed that teachers basically employed the same methods in all of the classes they taught. Even though the *F* ratio is robust enough to withstand these violations, especially given the large size of the sample, data were aggregated to the school level to avoid these violations. When the *F* ratio yielded a significant finding at this level, the analysis proceeded at the student level to determine if this too was significant.

Another limitation of this study concerns the nature of self-report data. One must wonder how accurately the reported data reflect what actually happens in the science classrooms of the teachers sampled. Also of concern is the notion of what motivated students to do their best on a test that for them was not matched to their school science curricula, not used for course grades, not required by the school or state, and, therefore, perhaps of little importance to them. Finally, one must consider whether or not science achievement as measured by a paper and pencil test is an appropriate measure of performance for those students engaged in hands-on science programs.

Discussion

To summarize, eighth-grade students who experience hands-on activities either every day or once a week score significantly higher than eighth-grade students who experience hands-on activities once a month, less than once a month, or never. Even though this study was limited, the logic for doing it still exists. Student performance on standardized tests is still viewed as a useful measure of the effectiveness of a program, even when the program bears little resem-

blance to the test used to assess it. Considering the emphasis placed on hands-on science instruction today, one must question whether or not programs employing such methods are assessed very well by paper and pencil instruments. Extending this concern to students taught science through hands-on methods, how accurate a measure of the students' abilities in science are these tests? Realistically, we must move beyond these tests to more authentic measures of student understanding. Because many educators no longer consider paper and pencil, multiple-choice tests as the best measure of students' accomplishments in science (Koballa, 1984), perhaps we should reconsider their use as the only, or predominant, method.

Despite the limitations of this study, the frequency of hands-on experience was strongly related to science achievement even when the test was not focused specifically on science process outcomes. These results do offer support for activity-based programs. However optimistic these results appear, there are serious questions that must be raised in their consideration. Although the NELS:88 data document the frequency of hands-on activities in the classroom, this notion is not defined. If researchers are going to continue to study activity-based science programs, a rigorous definition of hands-on in the context of the total classroom experience is required. What does a hands-on science classroom look like? What kinds of performances are expected by students as a result of this instruction? How should students in these programs be evaluated? How should the programs themselves be evaluated? Answering these pedagogical questions should help teachers better employ and understand hands-on science as a method of choice.

In defining the hands-on science classroom, one must consider the role of the teacher and the students, the time spent on task, and the required products of student performance. It is with certainty that one could describe what the hands-on science classroom does not look like. Imagine a classroom where students regularly spend class periods listening to a lecture, taking notes, and/or reading from a textbook. Teachers employing hands-on science do not eliminate the methods, but rather place them on the periphery of preferred methodologies, incorporating them as needed.

Another way to consider this is to say that hands-on methodology places prime importance on the experiences of the students, whereas the others rely heavily on teacher experience. Ultimately, teachers in hands-on science classrooms should be concerned with actively motivating and involving students in experiences that will in some way extend the students' knowledge and understanding of the science content being studied.

If hands-on experiences are to be beneficial to student understanding, the time spent on task must be meaningful. It would be foolish to state that only activities lasting close to a full class period provide the best learning experiences. The length of a task is not nearly as important as its illustrative clarity, and the skill with which the teacher is able to use the students' experience to help lead them to an understanding of the observed phenomena. The teacher's role in this situation is much more important than that of the traditional view of the teacher as a knowledge-giver, for in a hands-on science classroom, teachers must help students to construct their own knowledge.

Teachers using a hands-on approach must carefully consider how students will be evaluated. Although the acquisition of specific content knowledge must be examined, so must actual student performances. In evaluating both performance and knowledge, teachers can choose from a variety of assessment approaches. Daily performance can be evaluated through skills checklists, direct observations, journal writing, and/or activity reports. Knowledge can be assessed through traditional paper and pencil tests and informal interviews. Performance assessments and student portfolios can be used to evaluate both performance and knowledge. In the hands-on classroom, record keeping and grading will be significantly different than a classroom

taught through traditional methodologies. Teachers must take time to ensure that both students and parents are able to understand, interpret, and realize the value in these forms of assessment.

Lastly, one must consider program evaluation. Merely contending on the basis of theory that hands-on programs improve students achievement is not sufficient. These programs must be evaluated. However, the impact of these programs must be considered not only in terms of student achievement, but also with regard to process and analytical skills, and attitudes. In changing the scope of program evaluations to be more reflective of the total classroom experience, the mode of testing students in these programs must also change. Standardized achievement tests alone do not offer a comprehensive view of student accomplishments. Therefore, they should serve as but one portion of the evaluation package, and must be supported by performance assessments and attitudinal surveys.

Given these recommendations, further research on activity-based programs is certainly warranted. Questions raised as a result of this research need to be studied empirically. Current definitions of hands-on science need to be examined and, if necessary, revised or expanded. These definitions must then be studied in the context of the classroom. Ultimately, it will be the interpretation of such empirical data that will allow educators to make curricular decisions regarding hands-on science that will positively impact the students in these programs.

References

Atash, M.N., & Dawson, G.O. (1986). Some effects of the ISCS program: A meta analysis. *Journal of Research in Science Teaching, 23,* 377–385.

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

Davis, T., Raymond, A., MacRawls, C., & Jordan, J. (1976). A comparison of achievement and creativity of elementary school students using project vs. textbook programs. *Journal of Research in Science Teaching, 13,* 205–212.

Dunfee, M. (1967). *Elementary school science: A guide to current research.* Washington, DC: National Education Association.

Gagne, R.M. (1967). The AAAS project: Science—A process approach. In E. Victor & M.S. Lerner (Eds.), *Readings in science education for the elementary school.* New York: Collier-Macmillan Limited.

Glasson, G.E. (1989). The effects of hands-on and teacher demonstration laboratory methods on science achievement in relation to reasoning ability and prior knowledge. *Journal of Research in Science Teaching, 26,* 121–131.

Jacobson, W.J., Doran, R.L., Humrich, E., & Kanis, I.B. (1988). *International science report card.* Washington, DC: National Science Teachers Association.

Kessen, W. (1964). Statement of purposes and objectives of science education in school. *Journal of Research in Science Teaching, 2,* 4–6.

Klopfer, L.E., & Champagne, A.B. (1990). Ghosts of crisis past. *Science Education, 74,* 133–154.

Koballa, T.R. (1984). Goals of science education. In D. Holdzkom & P.B. Lutz (Eds.), *Research within reach: Science education* (p. 31). Washington, DC: National Science Teachers Association.

Kyle, W.C., Jr., Shymansky, J.A., & Alport, J.M. (1982). Alphabet soup science. *The Science Teacher, 49,* 49–53.

Lawson, A.E. (1990). Science education in Japan and the United States: Are the Japanese beating us at our own game? *Science Education, 74,* 495–501.

Lee, E.C. (1967). *New developments in science teaching*. Belmont: Wadsworth Publishing Company.

National Center for Education Statistics. (1989). *National Education Longitudinal Study of 1988: A User's manual* (Contract No. 90-464). Washington, DC: Office of Educational Research and Improvement.

National Center for Education Statistics. (1991). *Psychometric Report for the NELS:88 Base Year Test Battery* (Contract No. 300-86-0010). Washington, DC: Office of Educational Research and Improvement.

Shaw, T.J. (1983). The defect of a process-oriented science curriculum upon problem-solving ability. *Science Education, 67*, 615–623.

Shymansky, J.A. (1989). What research says about ESS, SCIS, and SAPA. *Science and Children, 26*, 33–35.

Shymansky, J.A., Kyle, W.C., Jr., & Alport, J.M. (1982). How effective were the hands-on science programs of yesterday? *Science and Children, 20*, 14–15.

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

Siegel, B.D., & Raven, R. (1971). The effects of manipulation on the acquisition of the compensatory concepts of speed, force, and work. *Journal of Research in Science Teaching, 8*, 373–378.

Vanek, E.P., & Montean, J.J. (1977). The effect of two science programs (ESS and Laidlaw) on student classification skills, science achievement, and attitudes. *Journal of Research in Science Teaching, 14*, 57–62.

Washton, N.S. (1974). *Teaching science in elementary and middle schools*. New York: David McKay Company.

Wideen, M.F. (1975). Comparison of student outcomes for Science—A process approach and traditional teaching for third, fourth, fifth, and sixth grade classes: A product evaluation. *Journal of Research in Science Teaching, 12*, 31–39.