Chapter 3: Hands-on Science and Student Achievement

In this chapter we examine the theoretical relationship of hands-on science and student achievement and the relationships that have been identified or tested by past research. Based on these, we propose the hypotheses linked to our three research questions. First, we examine the theoretical relationship of hands-on science and student achievement identifying the rationales that have been made for its proposed benefits. Second, we review the empirical literature on the testing of these rationales. As part of this review, we describe several issues that have affected the value of past research and discuss how this analysis will address them. Based on our reviews of the theoretical and empirical literature, we set out the hypotheses to be tested by this work. Last, we provide an overview of the two data sources to be used in testing these hypotheses leaving a more detailed description of each to Chapters 4 and 5.

Theoretical Rationales for Hands-On Science’s Role in Student Achievement

Hands-on science has been proposed as a means to increase student achievement in science education. A set of theories has been proposed to explain how hands-on science benefits student learning of science. Science educators identify two broad domains of scientific knowledge: content knowledge and process skills (Glynn and Duit 1995 chapter 1; Lawson 1995 chapter 3). Content knowledge (sometimes called declarative knowledge) includes the facts, principles, conceptual models, theories and laws which students are expected to understand and remember. Process skills (sometimes called procedural knowledge) are the techniques used in science, for

---

7 This discussion focuses on learning scientific knowledge and skills. Hands-on science has been justified on other grounds, such as improving lab techniques and motivation to learn (Shulman and Tamir 1973).
example, observation, measurement, and developing hypotheses, which students are to master. Both domains are considered necessary in order for students to fully understand science and be able to apply it (Glynn and Duit 1995; Champagne, Klopfer and Gunstone 1982; Eylon and Linn 1988). Hands-on science has been proposed as a means to increase students’ understanding of both types of knowledge.

Scientific content knowledge is often abstract and complex. Examining and manipulating objects may make this abstract knowledge more concrete and clearer. Through hands-on science students are able to see real-life illustrations of the knowledge and observe the effects of changes in different variables. These illustrations also provide references for discussion (Shulman and Tamir 1973; Friedlander and Tamir 1990).

The idea that hands-on science supports understanding of content knowledge is consistent with developmental theory’s positing successive stages (from three to five) of mental development through which humans pass. The highest stage includes the ability to work with abstractions. Before this stage can be reached, humans first pass through a stage in which thinking is confined to concrete matters. Interactions with the physical environment (along with other factors) support the mind’s passage through these stages (Piaget 1973, Gage and Berliner 1994, Lawson 1995). Under this view, hands-on science can help students move from the second highest stage to the highest stage as it offers concrete illustrations of abstract ideas at a time when the mind needs concrete representations for understanding. Once at the highest stage, however, hands-on science is of much less importance in helping the student gain understanding as the student is now capable of grasping and manipulating abstract ideas.
This argument for hands-on science is also consistent with cognitive theory’s information processing model of the mind which includes a long-term memory to store knowledge and a short-term memory to hold knowledge in immediate use. The ability to retrieve relevant knowledge from the long-term memory for use in the short-term memory is based on how the knowledge has been organized in the long-term memory and how strong the associations have been made between individual pieces of knowledge. Hands-on activities create additional associations between pieces of knowledge so that information can be referenced both by its abstract meaning and by a physical illustration of it. In this way, it improves information retrieval (Gage and Berliner 1984).

Hands-on science may also be used to address faults in information processing. According to cognitive theory, the separate bits of knowledge held in the long-term memory are organized using broader concepts known as schema. These schema are organizing principles which guide an individual’s understanding of the separate pieces of information and are used to organize and integrate new information. One can form schema that do not correspond to the real world. These misconceptions may prevent learning as new information may be synthesized in a way that justifies the misconception or may be ignored if it contradicts the misconception (Eylon and Linn 1988; Champagne, Klopfer and Gunstone 1982). One approach to instruction, known as conceptual change, attempts to identify these misconceptions, have the student realize they do not accurately explain phenomena and help the student adopt more realistic conceptions. Hands-on science, with its focus on real world phenomena, has been proposed as a method to help a teacher identify these misconceptions as well as provide a setting for students to explore
how their misconceptions falsely predict phenomena in preparation for reconsidering
them (Driver 1981; Driver and Bell 1986; Friedler and Tamir 1990).

Science education also entails the use of process skills which are the techniques of
science such as observation and measurement. What these process skills are and how to
teach them affects the use of hands-on science. Currently, a debate continues over many
facets of these skills: 1) their number and type, 2) at what age they should be taught, 3)
whether they need to be taught in a specific order or at the same time, and 4) whether
they can be taught separately from content knowledge. Gagne (1965) identifies eleven
skills and places them in two categories, basic and integrated. For him, basic skills must
be taught before integrated ones. Lowery (1992) identifies seven skills and attempts to
determine at what age students are developmentally ready to learn each. Resnick (1987)
argues that higher order skills are used along with basic skills when young students learn
and therefore the two types of skills must be taught together in all content areas. In the
1960s a new curriculum, known as SAPA, focused on teaching a hierarchical set of skills
without links to content. In contrast, studies comparing experts and novices have
concluded that content knowledge is critical to the correct use of process skills
(Champagne, Klopfer and Gunstone 1982; Eylon and Linn 1988).

Another concern regarding process skills is promoting the ability of students to
use all the individual process skills in combination for problem solving and carrying out a
scientific investigation on one’s own. Sometimes this goal has been set for only the best
students expected to go into the field of science and other times it has been geared to all
students through applications in daily life, be they personal or work related. The science
education instructional approaches seeking this goal have all included some level of hands-on science.

Some of these process skills are by their nature hands-on, for example measurement, and therefore are considered best learned through hands-on science. Others may be linked to hands-on activities but themselves are not, for example inferring is based on results derived from hands-on activity. Students can learn and practice these skills using results drawn from other sources than in-class activities leading to debate over the need for hands-on science when teaching them. Klopfer (1990), for example, argues for a larger role for hands-on science in learning the skills of gathering scientific information (e.g. through observation and measurement) than in the ability to make inferences and draw conclusions from experimentation.

The theoretical rationales given for the impact of hands-on science on student achievement have not gone unquestioned. Critics argue that hands-on science may reduce student achievement as well as improve it. Whereas proponents argue that hands-on science helps students visualize abstract ideas, opponents argue that it has the ability to confuse as well as clarify. Hands-on science also offers students additional opportunities not to learn as they may be busy doing activities but not thinking about the topic. Additionally, some research has shown that students may not link hands-on activities to written activities concerning the topic being studied (Wellington 1998, Hodson 1996; Atkinson 1990; Resnick and Klopfer 1989).

A practical criticism concerns the time and monetary costs of hands-on science. From Smith & Hall in 1902 through today, critics have argued that hands-on science
takes up too much time, requires expensive recurrent purchases, and drastically reduces the amount of material that can be covered in a course.

This practical point raises an issue of equity between higher and lower ability students. Lower ability students are expected to benefit both from the concrete examples and the greater time per topic provided by hands-on science. Higher ability students may be able to understand a topic in a shorter period using less time consuming instructional methods. For them, the additional time spent on one topic when using hands-on science may lead to a reduction in topics covered during a course.

**Research on the Relationship of Hands-on Science and Student Achievement**

The importance of resolving the question of hands-on science’s relationship with student achievement has not gone unnoticed. There is a body of research on the topic, which we review in this section. We organize our review under three broad headings. First, past research has not led to a firm conclusion regarding the link between hands-on science and student achievement. We review the past research with an emphasis on its overall inconclusive nature. Second, we discuss some of the important issues in this research and how the inability to resolve these may contribute to the lack of agreement. Third, we discuss how we plan on improving upon the past work and note that our work will still be subject to some of the same difficulties.

**Inconclusiveness of Past Research**

Research on the relationship of hands-on science and student test scores has been conducted since the turn of the century. The research has been based on three techniques.
Under small-scale experiments, very small groups of teachers are assigned a particular teaching method (for example, lecture, text based or hands-on science) to use in their classes and the test scores of their students are compared. A second approach is to compare classrooms using different types of curricula. Test scores from students using different curricula are compared and conclusions are drawn based on different levels of hands-on science in each curriculum. Alternatively, surveys have been used to determine the quantity of hands-on science in a classroom and to collect student test scores. The relationships between test score and hands-on science are then examined. Because our work follows this approach, we devote the greater part of our review to this vein of past research.

Overall, reviews of the experimental studies have not found a positive correlation between hands-on science and test scores except for tests of lab skills. Cunningham’s (1946) review argued that the benefit of students receiving more lab instruction was in their ability to use lab apparatus but not in their achievement on tests. Shulman and Tamir (1973) reviewed the literature and also found that the majority of studies failed to show lab teaching as more effective than other instructional methods in regards to test scores. The Hofstein and Lunneta (1982) review again found that lab work showed no significant benefits over other methods of instruction when comparing scores on tests for achievement, critical thinking and understanding the processes of science. They did find a positive effect for lab skills. White and Tisher (1985) reviewed teacher and student surveys and found little agreement on the perceived benefits of hands-on science.

Reviews of the impacts of the curriculum developed from the late 1950s through the early 1970s, in which hands-on activities played an integral part, showed more
positive results. Bredderman (1983) and Shymansky, Kyle and Alport (1983) carried out meta-analyses to synthesize studies (57 and 105 respectively) on these curricula. The former article focused on elementary programs while the latter addressed K-12 curricula. They reported significant positive results regarding student test scores (both for content and analytical skills). Interestingly, Bredderman (1983) did not find wide variation in the effects of the three elementary science programs: SAPA (process approach), ESS (discovery approach) and SCIS (exploratory approach) but did find some sensitivity to the instructional approach used. For example, students using SAPA scored higher on process outcomes and students using SCIS scored higher on content outcomes. In addition, he found that student outcomes under these programs were higher than those of students using textbooks and only somewhat higher than students using other activity based programs. From these two findings, he concluded that the common features of the three programs, that they contained more hands-on science and gave more attention to process, were more important to student outcomes than any unique features of each program. Initial analyses of the new curricula developed in the 1990s have begun. As the 5th and 6th grade portions of the curricula were developed later and the middle grades curricula are being piloted now, it will be several years before similar meta-analyses can be done to determine the overall and specific effects of these curriculum on student achievement. If the same results are obtained, that will be a strong argument that hands-on science, a common component of the new curricula, is responsible for their benefits.

Much of the recent work on this issue has been carried out using data from international and national science studies. Two organizations run separate series of international studies: The International Association for the Evaluation of Educational
Achievement (IEA) has supported four studies and the International Assessment of Educational Progress (IAEP) has organized two studies. All six of these surveys are similar in that they are cross-sectional in nature, gather survey data from students, teachers and schools, and include student test scores, primarily on multiple choice tests. The results from these studies are mixed and the majority do not show a positive relationship between hands-on science and test scores. In several cases, the value of these studies is hampered by a lack of reporting of the full results.

The IEA’s First International Science Study surveyed 10 and 14 year olds and students completing secondary school in 1970-1971. However, only 10 year olds were asked whether or not they made observations and did experiments in science class. In 12 of the 15 countries or regions included in the survey, a yes answer was positively correlated with a higher test score. In the U.S. the correlation coefficient was .18 using a sample of about 5400 students. A further analysis using an OLS regression model with covariates such as student type, school type, home and attitude, found positive coefficients for the variable representing observations and experiments but the significance levels were not reported (Comber and Keeves 1973).

The IEA’s Second International Science Study surveyed 10 and 14 year olds in six countries from 1983-1986. The students were asked how often they did experiments in science class and could choose from three answers (never, sometimes and often). Teachers were asked how much time was spent on practical activities, defined as experiments or field work, and could choose from four answers (zero or little, one-fourth, one-half, or three-fourths or more). The teacher item was not included in the U.S. survey.

---

8 The fourth study, the Third International Math and Science Study, has not yet released results regarding classroom instructional practices and their relationship with test scores.
The study reported only the average correlation between the hands-on item and test scores for all the countries as a group. For grade 5, the average correlation was .07 using the student responses and 0 using the teacher responses. For grade 9, the test was subdivided into content and process sections. For student responses, the average correlation was .09 for content score and .07 for process score. For teacher responses (with no U.S. participation) the average correlation was respectively .21 and .25 but these were skewed by upward by one country) (Tamir & Doran nd; Doran & Tamir 1992).

From 1983-1987, the IEA carried out its Classroom Environmental Study which included an examination in six countries (not including the U.S.) of how science classroom factors affected achievement scores. Classroom observation was used to collect data on the amount of lab work done. No country showed a positive significant correlation between the amount of lab work and gains from a pre-test to a post-test (five countries had non-significant results and one had a negative significant correlation) (Anderson, Ryan and Shapiro 1989).

The IAEP surveyed 24,000 13 year olds in 5 countries and 4 Canadian provinces in 1988. Students were asked how often they did experiments on their own and with other students. They choose from 5 responses: never, less than once a week, once a week, several times a week, and almost every day. Rather than report the actual correlations of these responses with test scores, the authors stated that the frequency of experiments was not consistently related to test performance (Lapointe, Mead and Phillips 1989, p. 41).

In 1991, the IAEP survey 25,000 9 year olds in 14 countries and almost 52,000 12 year olds in 20 countries. The students were asked the same two questions concerning
experiments as in 1988 and their answer was combined into a single measure. No country showed a positive significant correlation between the 9 year olds’ responses and their test scores (7 showed non-significant correlations including the U.S. and 7 showed significant negative correlations). Only one country showed a positive significant correlation between the 12 year olds’ responses and their test scores (8 had no significant correlation and 11 had significant negative correlations, including the U.S.). The actual correlations were not provided in the study (Lapointe, Askew, Mead 1992, pages 50 and 94).

Domestically, the U.S. has two national surveys that include science education. The National Assessment of Educational Progress (NAEP) is an ongoing cross-sectional survey and the National Longitudinal Survey (NELS:88) longitudinally followed students through secondary school from 1988 to 1992. NAEP national sciences surveys have been most recently carried out in 1986, 1990 and 1996. In 1986 they surveyed grades 3, 7 and 11 then switched in 1990 to grades 4, 8 and 12. NAEP analysis of hands-on science’s relationship to test scores has been confined to grouping students by the level of hands-on science they report, averaging the test scores for each level and comparing them. In the 1986 NAEP, students in grades 3 and 7 were asked how often they used different types of equipment. Their responses were grouped into three categories: low, medium and high. Students in grades 7 and 11 were asked how often they were involved in innovative classroom activities including hands-on activities. Their responses were grouped into the same three categories. In both cases, the high response group scored significantly more than the low response group (Mullis and Jenkins 1988).
In 1990, students in all three grades were asked which of six broad categories of equipment or materials they had used. Their answers were grouped into four categories: none, 1 or 2, 3 or 4, and 5 or 6. For grades 8 and 12, the mean student scores monotonically rose with the response category with a significant difference between the none category and the 5 or 6 response (Jones, Mulllis, Raizen, Weiss, and Weston 1992).

In 1996, students in all three grades were asked if they had done hands-on activities or projects with seven different types of materials or instruments (eight for grade 12). Except for several cases involving 4th grade, students who answered yes had a higher group score than those answering no. Students who had used none of the materials or instruments scored lower than those who had used some in all grades. In addition, for grades 4 and 8 teachers reported how often they used hands-on activities with four possible responses: never or hardly ever, once or twice a month, once or twice a week, and almost every day. There was no difference in the student test score means for each of these groups in the 4th grade but in 8th grade the two highest response categories scored more than the lowest response category. The study further grouped students into science proficiency categories based on their test scores. The 8th grade students who reported more hands-on activities were more likely to be ranked at or above the Proficient level. The same was found for 12th grade students who self-reported higher categories of hands-on activities (Sullivan and Weiss 1999).

The NELS:88 included 25,000 8th graders in 1988 and about 11,000 were surveyed on science. Follow-up surveys were done in 10th grade in 1990 and 12th grade in 1992 with additional students added to maintain a nationally representative cross-sectional sample for each year. Horn, Hafner and Owings (1992) grouped test scores by
answers to a teacher item concerning how often experiments were done in the classroom. Possible responses included never or less than once a month, once a month, once a week and almost everyday. The mean student scores for the categories of almost every day and once a week were significantly higher than those of the other two categories. Teacher reports on hands-on science done in the 8th grade were also found to have a positive relationship with a subset of the NELS 10th grade test dealing with quantitative operations, chemistry and scientific reasoning but not with subsets of the other test items (Hamilton, et. al. 1995). This positive relationship was not found in a similar analysis of 12th grade sub-scores (Nussbaum, Hamilton and Snow 1997) nor in analysis of the frequency of experiments and growth in test scores between 10th and 12th grades (Hoffer and Moore 1996). Another analysis of NELS subscores divided the test into life science and physical science scores. No relationship was found between either subscore and 8th grade student reports of taking a science course with a lab once a week but a positive relationship was found between the physical science subscore and an interaction effect for girls who reported yes (Lee and Burkham 1996). A further analysis of 10th grade students using both a student and a teacher reported hands-on composite variable found a positive relationship between the teacher reports and both subscores (Burkham, Lee and Smerdon 1997).

Up to this point, we have discussed research on the relationship of hands-on science and scores from multiple choice tests. Research on the relationship of hands-on work and performance assessments are few and inconsistent. In some cases, the analysis of the relationship was not done. For example, the First International Science Study included a hands-on test for 14 year olds in England and Japan and secondary school
completers in England but the analyses did not correlate scores from these tests with the amount of hands-on science (Comber and Keeves 1973; Tamir and Doran n.d.; Kojima 1974). The 1986 NAEP included a pilot study using hands-on tests but the purpose was to assess these tasks rather than determine their relationship to hands-on science or other factors (Blumberg, et al. 1986). The current NAEP use both short and long answer constructive responses, some based on hands-on tasks, along with multiple choice questions but the past analyses use only the total score (Sullivan and Weiss 1999).

The results from research on the IEA are not conclusive. The second IEA international science survey included three hands-on tests for both Grade 5 and Grade 9 students. Each test was composed of three hands-on tasks to be done by a student in one class period. Correlations were reported for the relationship of hands-on science and hands-on test scores. For grade 5 student reports, the mean correlation was 0 (-.03 for the U.S.) and for grade 5 teacher reports the mean correlation was again 0 (the U.S. did not collect teacher reports). For the grade 9 student reports, the mean correlation was .09 while for the teacher reports it was .30 but this figure was skewed upward by one country (Doran and Tamir 1992).

A small scale study found that 5th grade students using a strongly hands-on curriculum had a mean score on a hands-on test one-half standard deviation higher than those using a textbook based curriculum but this result may be confounded since the student with the hands-on curriculum also scored higher on a cognitive ability test (Baxter, Shavelson, Goldman and Pine 1992). Only in the case of lab skill tests has hands-on science been shown consistently associated with higher test scores (Yaeger, Engen and Snider 1969, for a review see Hofstein and Lunetta 1982). This research has
not determined how lab skills contribute to overall science achievement or achievement in specific domains.

Overall then, we do not find a consensus on the relationship between hands-on science and achievement. The experiment based research has not supported the relationship between hands-on science and student achievement. The curricula focused research has found a relationship with achievement but it is not clear whether hands-on science or some other aspect of the curricula are responsible for this. The survey based research provides mixed results.

Issues Surrounding this Research

The research on hands-on science has grown more sophisticated over time. However, there are still several issues that have yet to be resolved. These issues may in part be responsible for the lack of agreement in past research. Here we discuss four of these issues: 1) variables related to both achievement and hands-on science, 2) a lack of large data sets containing individual student performance test scores along with important explanatory variables, 3) a differential relationship of hands-on science to student achievement by student ability, and 4) the multiple facets of hands-on science.

There are a large number of variables that may affect achievement and need to be considered when analyzing test scores. Some of these are related to the level of hands-on science as well as achievement and if not controlled for, a spurious relationship between hands-on science and test score can result. This problem is especially true for correlational studies, such as those done with the NAEP, which did not control for any variables but may also be true for multivariate analyses as well.
Among the variables that past research has identified as linked to achievement, several may also be linked to the level of hands-on science. These include SES, prior science course taking, achievement level of the class, and student ability or past achievement. Research using both the NAEP and NELS:88 finds that higher SES students achieve higher scores (Jones, Mullis, Raizen, Weiss and Weston 1992; Horn, Hafner and Owings 1992). Research has also shown that higher SES students report a greater amount of hands-on science (Horn, Hafner and Owings 1992; Hoffer and Moore 1996) and we have obtained the same finding using the NELS data for 8th and 10th graders for both teacher and student reports (see Table 5-3). Similarly, the number of science courses a student takes affects their proficiency in science regardless of their SES, race or gender (Hoffer, Rasinski and Moore 1995; Madigan 1997) and the more science courses a student takes, the more likely they are to have hands-on science. Additionally, the academic level of the class and the academic ability of the individual student are related to both high test scores and higher levels of hands-on science (Hoffer and Moore 1996) and we have obtained the same finding in the NELS data for grades 8 and 10 (see Table 5-3) using both student and teacher reports as well as in the RAND data when using student reports (see Table 4-4). Research not controlling for these variables could produce results different from that which has.

Standardized testing used in large surveys and many of the smaller experiments and evaluations of curricula has relied primarily upon multiple choice items. These tests have many positive attributes including ease of administration, relatively inexpensive scoring, ease of standardization, good psychometric properties and the ability to provide coverage of content in a short testing period. During the 1990s the focus on multiple
choice testing came under a strong critique for a variety of reasons. Relevant to our research is the criticism that multiple choice tests cannot cover the wide range of skills that should be evaluated. Critics argue that multiple choice can only test narrow content areas and skills – especially short-term recall of facts and basic process skills - but cannot address the broader abilities of critical thinking, evaluation and problem solving (Miller and Legy 1993). Resnick and Resnick (1992) conclude that multiple choice standardized tests ask for quick bits of information, not linked to each other and not used for a whole task such as solving a problem or interpreting result: in short, they are not linked to thinking. Madeus et al. (1992) reviewed six standardized tests for math and science (with copyrights of 1985-1991) and found them to focus on short term recall questions concerning facts, definitions, and applications of formula, all of which could be learned by rote. The tests covered little procedural knowledge or problem solving and reasoning. A broader critique is that multiple choice tests are irrelevant to daily life and that testing should replicate the challenges people will face in the real world (Wiggens 1989).

These critiques are important to our research as they imply that focusing on multiple choice testing may overlook some of the benefits of hands-on science. If we use multiple choice tests to stand for achievement and if these tests cannot assess some of the benefits of hands-on science (such as teaching procedural knowledge and skills) then we may not capture the actual relationship between hands-on science and tests. Murnnane and Raizen (1988) make the specific charge that multiple choice testing cannot pick up what is taught through hands-on science.

A key conclusion drawn by the critics is that multiple choice testing needs to be supplemented (and in some cases supplanted) by alternative forms of assessments.
covering a wider range of knowledge and skills. These assessments are of many types and can be categorized in a number of different ways (OTA 1992; Herman, Aschbacher and Winters 1992). They may include such tasks as short or long written responses, oral presentations, the actual performance of a task, or the results of a performance (e.g. an exhibition or portfolio of past work). For the evaluation of hands-on science, alternative assessments may require the combination of the performance of hands-on activities with written or oral means. The performance section would allow the assessment of the hands-on skills while the written or oral section would address other procedural skills of science (e.g., developing hypotheses, recording data, and making conclusions based on results) as well as content knowledge associated with the activities.

The criticisms of multiple choice testing and the promotion of alternative testing represent one train of thought rather than the accepted viewpoint. Both in theory and practice, multiple choice remains widely defended and used for standardized testing. Mehrens (1992) provides a literature review to counter many of the criticisms of multiple choice testing including its narrowing of content and skills tested and inability to test higher thinking skills. He notes, though, that there is general acceptance that some types of procedural knowledge cannot be tested using multiple choice but require a performance assessment.

In the past, only a small number of testing surveys included hands-on activities. Today, while there is more alternate assessment in use, much of it is not relevant for our work. In some cases, only writing items are used (e.g. commercial tests such as the Stanford Achievement Test 9). In other cases, hands-on activities are done but the reported data cannot be used (e.g., the 1999 NAEP mixed the constructive response
scores with the multiple choice scores, the Maryland’s MSPAP program does not provide individual student test scores, and data from the Third International Math & Science Study has not been released). There is then a paucity of data to assess the links between hands-on science and hands-on science assessment that has two implications for our research.

First, it is important to understand how hands-on science relates to multiple choice test results but it is also important to determine if the type of testing (be it multiple choice or some alternative form) changes this relationship. If so, we may be missing some connection between hands-on science and student achievement. Second, while performance tests may have greater face validity for testing hands-on instruction, they have not been proven to have more construct validity than multiple choice tests. Their justification on these grounds then remains open to question and analysis.

A third issue concerns how student ability affects the relationship of hands-on science with student achievement. Past research has examined certain interactions of hands-on science with specific student groupings, for example gender and race/ethnicity (Peng and Hill 1995; Lee and Burkham 1996; Burkham, Lee and Smerdon 1997). But there has been little work on whether the relationship differs for higher versus lower ability students. From a policy point of view, this interaction is important for the public acceptance of hands-on science. Methods of instruction are often pulled in two directions with regards to the issue of equity. On one hand, methods may be supported that have the goal of increasing equity by improving the performance of lower ability students. On the other hand, if these methods have adverse effects on higher ability students, their introduction may be opposed. Politically acceptable methods then must benefit both
groups or benefit one group without harming the other (i.e., one group may achieve more under the different method of instruction but the other group must achieve no less under it than when using other methods). In the case of adverse effects on one group, the method may be restricted to programs strictly for the other group. As our analysis is concerned with the broad application of hands-on science across the majority of classrooms, we are most interested in determining whether we find a benefit in the use of hands-on science without an adverse effect for any ability group.

From a theoretical point of view, it is expected that hands-on science should have a differential effect for higher versus lower ability students. Current theory, though, provides support for a differential effect that could favor either high or low ability students. Research on the interaction of student aptitudes (including ability) and instructional methods (known as Aptitude-Treatment Interaction or ATI) has found that teaching methods often have differential effects based upon student ability (Cronbach and Snow 1981). Specifically, instructional methods that require greater responsibility of the learner benefit higher ability learners more while methods that pre-process the material better help lower ability learners. Hands-on science requires students to do more on their own and draw conclusions for this work. This requirement for greater student responsibility argues that hands-on science may be of greater benefit to higher ability students or lessor benefit to lower ability students. There is some evidence for this from analysis of the hands-on curricula developed in the 1960s (Koran & Koran 1984). If true, hands-on science would be politically less acceptable as it would have an adverse impact on lower ability students.
At the same time, ATI also proposes that higher ability students benefit from instructional methods based on abstract concepts (Koran & Koran 1984). Hands-on science’s strength, conversely, is in making the abstract concrete through physical examples. From this point of view, hands-on science may better benefit the lower achieving student who is not yet ready for instructional methods based solely on abstractions and hinder the higher achieving student who is. Development theory would also support this view as hands-on science would help students move up through the concrete thinking stage of mental development to the higher abstract thinking stage. In this case, hands-on science may benefit lower achieving students but not higher ability ones.

Time constraints may also contribute to a differential impact of hands-on science based on student ability. Hands-on science requires more time to cover the same material than other methods thereby reducing the time available to cover other material. Higher ability students may be ready to cover more material more quickly using other methods besides hands-on science, especially if they can grasp abstract concepts with a minimum of concrete examples. If this is true, then when taught using hands-on science they will cover less material in class and learn less. Because this is an issue of material coverage, any differential effect of hands-on science for higher ability students will more obviously appear in tests covering more material. If true, there could strong opposition to the general introduction of hands-on science by higher achieving students and their parents.

A fourth issue for researching hands-on science is its multiple facets of quantity, quality, and instructional approach. Quantity is the measure of how much hands-on science is done in the classroom and/or how often it is done. Quality describes the skill
of the teacher providing instruction through hands-on science and the value of the activities used. Third, hands-on science can be taught through a number of instructional approaches, as detailed in Chapter 2. For both theoretical and practical reasons, research has focused on quantity measures.

The critical distinction between hands-on science and conventional science is students’ hands-on activities in classroom. Students’ involvement in hands-on activities exposes them to vivid connections from abstract concepts to concrete examples and from scientific principals to scientific applications, which contribute to students’ better understanding of the scientific knowledge and their ability to solve problems using their scientific knowledge. Quantity of hands-on activities in classroom captures this distinction between hand-on science and conventional science instructional methods as well as the degree to which hands-on activities are used. Thus, operationalizing hands-on science using quantity is valid.

In practice, data collection has focused on quantity. Quantity issues are considered more reliably surveyed as the amount of hands-on science done in a classroom is considered memorable. Teachers’ reports on hand-on amount are relatively more reliable since the amount of hands-on activities is included in their lesson plans. Students’ reports are relatively more reliable since hand-on activities are more interesting and more likely to be remembered. This assumed reliability combined with cost issues has led to a focus on the use of surveys versus classroom observation or use of teacher and student logs. Teachers and or students are most often surveyed as to the overall amount or frequency of hands-on work activities they have done in class. In addition, they may be asked about the use of specific materials or instruments or the completion of
activities on specific topics. Proxy questions assumed to be related to the level of hands-on science may also be asked, such as frequency of lab reports.

Surveys concerning behavior (such as the quantity of hands-on work) are subject to several sources of error, primarily memory, possible motivations to misreport, and failure to understand the question correctly (Sudman and Bradburn 1991). These sources of error can be reduced in the structure and administration of the survey. Memory problems can be reduced by making items very specific, covering longer time periods to avoid respondents compressing time (moving events forward into the time period reported on), and allowing respondents to review their records. Motivation problems can be reduced by using less threatening questions, allowing less specific answers (such as daily, weekly, or monthly), and using more anonymous methods in administration and recording. Problems of understanding can be reduced by using familiar words (Sudman and Bradburn 1991; Rossi, Wright and Anderson 1983). Surveys on the quantity of hands-on work use many of these techniques. Surveys have specific items on frequency of experiments, they cover fairly long specific time periods of one course, they allow teachers to review their records, they often use open answers, they are given with a promise of anonymity for both teachers and students, and they use words familiar to science classrooms (e.g. experiments, scientific equipment, and observations).

The impact of these sources of error can be identified using verification studies. The work that has been done specifically on verifying surveys on classroom instructional behaviors has supported the use of teacher surveys. Burnstein, et. al. (1995) found that surveys of mathematics teachers gave an accurate picture of their level of use of different instructional methods when compared against a five week teacher log but noted that this
finding might occur because there seemed to be little variation in teachers’ methods overall. Porter (1995) found a correlation of .65 between math and science teacher survey responses concerning the amount of lab work and their logs kept over a year. Verification of student reports of methods used in the classroom has not been done though students surveys have been found to have reasonable construct validity concerning grades, course taking and some family variables (Fetters, Stowe and Owings 1984; Valliga 1986).

Measuring the quality of the hands-on science has been constrained by two factors. First, there is little agreement on how to measure quality of instruction. Second, quality of instruction is more open to individual interpretation than a quantity value. A small scale study of algebra teaching that tried to validate survey responses using classroom observation found that a composite of survey items could separate teachers using certain practices from those who did not but could not differentiate the quality of the practices among teachers using them (Mayer 1999). To ensure reliability, classroom observation would be the desired method of measurement but the high costs of this approach have constrained its use. For these reasons, most studies of hands-on science have not been able to address the issue of instructional quality.

Measuring the instructional approaches to hands-on science also faces difficulties. First, the approaches outlined above are not always clearly defined, are expected to vary with student ability, and may be used in various combinations by teachers. These factors increase the complexity of data collection even for classroom observation. Some of the large-scale surveys, like the NAEP and NELS:88, asked questions regarding teacher’s overall instructional approaches but have not specifically asked about approaches used
with hands-on science. Small-scale experiments often compare hands-on science teaching versus other methods (e.g., lecture or demonstration - for example, Yager, Engen and Snider 1969). As a result, they compare the quantity of hands-on science (some set amount in one classroom versus none in the other classrooms) rather than the interaction of quantity and instructional approach.

From a policy standpoint quantity is the first factor to examine because it can be most easily increased. Increasing quality of instruction or changing the instructional approach would require a much more intensive professional development program for teachers due to the current mix in the quality of teaching and in the instructional approaches used. A strong relationship between quantity and achievement would support Bredderman’s (1992) inference that quantity of hands-on work and/or a focus on process may have a greater effect on achievement than instructional approach. If no link is found, though, concerns will remain whether quality or instructional approach play an unmeasured role. Ongoing attempts to improve the ability to measure both quality and instructional approach in the classroom, through both observation, teacher logs, and surveys will lead to future research better able to account for links between quality and approach of hands-on science and achievement. These will be important to refining suggested uses of hands-on science, e.g. only with specific instructional approaches or when teachers can provide a certain level of quality activities.

Improving Upon Past Work

Our goal is to improve upon past research on the relationship of hands-on science and student achievement. We do this by addressing the issues discussed above. First,
the RAND data set includes both multiple choice and performance test scores from the same students. The performance tests were based on actual hands-activities done by the students who in addition wrote up their results and answers to further questions requiring reflection on the results. In addition, the data set contains both teacher and student reports on the level of hands-on science in the classroom. Thus, the RAND data set contains the two types of assessments and the level of hands-on science needed for this research.

In addition, because the two types of tests were taken by the same students, we can analyze a further topic not much looked at in the literature. Assuming there is a positive relationship between hands-on science and test score, we can consider whether there is a stronger relationship between hands-on science and one of the two types of tests. Past research has noted the possibility of such a differential relationship (Doran & Tamir 1992) and proponents of alternative assessment and hands-on science assume it (Murnanne & Raizen 1988). We might predict such a result as process skills are often hands-on in nature and performance tests are thought to focus on process skills more than multiple choice tests but there has not been a direct test of it.

A further contribution of this comparison is to provide some evidence regarding the claims that performance tests are more valid indicators of what students learn through a hands-on approach. A differential relationship between hands-on science and the two types of tests would support this claim while the lack of one (or one in the wrong direction) would not.

Second, we can avoid many of the concerns for a possible spurious relationship between hands-on science and student achievement by using multivariate analysis and
including a broader array of variables that may be linked to both hands-on science and achievement. This is especially true for our analysis of the National Educational Longitudinal Survey (NELS:88) which contains a wide variety of this type of variable. NELS:88 provides us with the SES of the students’ families, measures of student ability or achievement, achievement level of the class, and prior science course taking. In addition, we create our own composite variables, discussed in Chapter 5, to better reflect our conception of hands-on science and carry out additional analyses to determine if the variables linked to hands-on science and achievement affect our findings. Much of this additional analysis focuses on course-taking because it has several dimensions such as type and order of course-taking.

Third, we will test for a differential relationship between hands-on science and student achievement by student ability. As noted above, we will be using measures of student ability or achievement in our analysis to avoid spurious results. We will look at the interactions of these ability measures with the levels of hands-on science to determine if ability affects hands-on science’s relationship with achievement. We will first determine if these interactions should be included in our analysis. If we find they should be included, then we will determine if they actually produce a differential relationship.

Fourth, we will continue the focus on quantity as our measure of hands-on science for both theoretical and practical reasons. Theoretically, quantity measures the primary difference between hands-on science and conventional instructional methods that we wish to measure. Practically, quantity is easier to measure, has higher reliability and is more widely available. Using the quantity of hands-on science also allows us to estimate the appropriate amount of hands-on work. As discussed previously, because of time
constraints too much hands-on activities will greatly reduce the coverage of scientific topics, leading to a negative relationship with achievement. Our analysis will include this possibility and we will attempt to determine if there is a level of hands-on science at which the relationship to test score diminishes.

It is important to note that while the amount of hands-on activities in the classroom captures the distinction between hands-on science and conventional science, it does not distinguish the variations within hands-on science. Different instructional approaches to hands-on science (e.g., discovery or inquiry) and the quality of instruction (e.g., quality of the kit, teachers’ preparation, and teachers’ ability to stimulate the classroom) affect students’ achievement just as they do conventional instructional methods. Our measure of quantity of hands-on science will not capture these variations within hands-on science. It would be ideal if we could assess various instruction approaches and quality of hands-on science by comparing various instruction approaches and quality of conventional methods of instruction. However, we lack data describing instruction approaches and quality for either hands-on science or conventional methods. Thus our estimate of the relationship of the quantity of hands-on science with achievement will capture the average effects of various instruction approaches and the quality variation in hands-on science in comparison with the average effects of various instruction approaches and the quality variation in conventional instruction methods.

**Hypotheses**

Our analysis will center on the testing of three hypotheses regarding the relationship of hands-on science and student achievement. More specifically we will
examine the relationship of the quantity of hands-on science with student scores on standardized tests.

Our first hypothesis concerns the expected positive relationship between hands-on science and achievement. We hypothesize that students engaged in more hands-on science will score better on standardized tests than students who carry out less hands-on science, all other things being equal. Ideally, we hypothesize that a student’s test scores would improve if they received greater hands-on science. As we cannot practically test this hypothesis, we have operationalized it by predicting that students who have received more hands-on science will score higher on tests than their counterparts who received less. We derive this hypothesis from the different theories of learning relevant to hands-on science, all of which see a benefit in the provision of concrete examples of abstract knowledge. These theories propose an especially important role for real-life illustrations of content knowledge to help students pass through the concrete stage to the abstract of thinking in the development of their minds, or to improve information retrieval from long-term memory by creating greater associations between pieces of knowledge, and to help prevent or correct misorganization of this knowledge. For process knowledge, these theories propose several benefits of hands-on science. First, many process skills are hands-on in nature and can only be learned through hands-on practice. Second, some of these skills link abstract concepts to empirical reality (e.g. the concept of variables and how to identify them) and learners would benefit from concrete examples of them. Third, hands-on activities offer opportunities for applying all the process skills in combination while focusing on specific content knowledge.
Past research has found mixed results regarding this hypothesis. We hope to obtain more consistent results using two data sets. Using multivariate analysis with increased control of variables we will address additional factors that may be confounding the relationship of hands-on work and test scores. Furthermore, we will test the hypothesis using two different types of tests, multiple choice and performance. As each type of test measures achievement with particular emphasis on a different domain, content versus process knowledge respectively, a finding of a positive relationship between hands-on work and both types of tests would provide more robust evidence for the relationship.

From a policy perspective, a rejection of the null hypothesis (that there is no relationship between hands-on science and test scores) would provide evidence to favor the increased use of hands-on science, the continuation of the policy of the 1960s. Current attempts to temper the use of hands-on science and redirect it under an inquiry approach might better be held off until further research examined the interactions of hands-on science and instructional approach. Conversely, a failure to reject the hypothesis would support the tempering of its use unless supportive results were obtained regarding the roles of instructional approach or quality on its relationship to achievement.

The second hypothesis grows out of measuring student achievement using the two types of tests. Specifically, we hypothesize that performance tests will reflect a greater difference in scores among students receiving different levels of hands-on science. Hands-on science benefits students taking performance tests because it teaches process skills required and assessed in performance assessment. Conversely, multiple choice
tests are theoretically criticized for an inability to assess process skills and even their proponents note they it cannot cover all procedural skills.

Policy-wise the results from this analysis will have implications for the adoption of hands-on science and for standardized testing. As the majority of standardized testing is done through multiple choice tests, a stronger relationship with performance tests will identify a deficit in assessing for process skills. As a result, revising multiple choice tests or supporting wider use of performance tests may be justified. These revisions could lead to further support for the adoption of hands-on science. For as tests are revised, student test scores would be expected to show a greater difference between results by level of hands-on science (though a lesser difference between the same level by test type).

Failure to find evidence in favor of this hypothesis, assuming that evidence had already been found for Hypothesis 1, would support the conclusion that both multiple choice and performance tests capture the benefits of hands-on science. This finding would provide less justification for revising multiple choice tests or increasing the use of performance tests (unless other research found that neither was well designed to capture process skills). With either finding, because of the debate over how to teach process skills, further research could be done on whether interactions of quantity of hands-on with instructional approach provide different results.

Third, we hypothesize that there is a differential effect of hands-on science for students of higher versus lower ability. While past research has identified similar effects for other types of instructional methods, the theoretical generalizations induced from that work is unclear on which way the differential effect will tilt. That hands-on science requires student to take more responsibility for their work should benefit higher ability
students and we should see a differential effect in favor of higher ability students. On the other hand, that hands-on science makes the abstract more concrete should benefit lower ability students and we should see a differential effect in favor of lower ability students. Furthermore, the time requirements of hands-on science may reduce the number of topics a student is exposed to in class. For higher ability students able to cover a greater number of topics, hands-on science may reduce the amount they learn. This should be reflected in lower test scores, especially tests that cover a broad range of topics. Considering these opposite effects together, we hypothesize that hands-on science has a smaller positive relationship with achievement for higher ability students than lower ability ones. In particular, we would expect that any differential effect that favors lower ability students would be more strongly seen in a comparison of multiple choice test scores as multiple choice tests reflect a wider range of topics than performance tests.

Past research has not directly examined this interaction. Ability-interaction research has not looked directly at hands-on science. Research on hands-on science has focused on racial/ethnic and gender groups rather than ability groups. Although in some cases this work has controlled for ability, it has not examined the interaction of ability and hands-on science.

From a policy perspective, evidence in favor of the hypothesis of the existence of a differential impact by student ability (especially if one group found zero or a negative relationship) would make a general acceptance of hands-on science more complicated. Such a result would lend support for the differential use of hands-on science for only that group that benefited. The rejection of the hypothesis or if a weak differential was found
(e.g., positive association with achievement for all ability groups though greater for one) would support the widespread use of hands-on science.

**Data Sources**

To tests these hypothesis we will analyze two separate data sets based on surveys of students and teachers. Both data sets include information on the level of hands-on work, student test scores, and other student and school covariates. The level of hands-on science is provided by both teacher and student reports allowing us to compare results from both. In both surveys, students took standardized science tests and the scores are used to represent their level of achievement in science. As noted in Chapters 4 and 5, these tests exhibit the same characteristics of other tests used for this purpose. Students of higher ability and social economic status produce higher test scores while minority students (Black and Hispanic) have lower scores. The surveys also contain additional information on the students, their families and their schools that may be related to both hands-on science and test score. With them, we can use the hands-on science and test score data to perform a multivariate regression analysis and thereby obtain estimates of the relationship of hands-on science to test scores holding the covariates constant.

Our primary data set is the RAND 1994 study on 1400 8th graders in Southern California. The study surveyed students and teachers about the amount of hands-on science they did. Students took both a multiple choice and a hands-on standardized test. In addition, some student demographics and abilities were also surveyed.

The RAND data will be used to test all three hypotheses (see Table 3-1). We can determine whether there is a positive relationship of hands-on science to achievement...
(Hypothesis 1) using both multiple choice and performance tests score. If so, we next can test whether the relationship is stronger for performance tests (Hypothesis 2). Third, we can test for a differential relationship depending upon ability level of the student (Hypothesis 3) using both multiple choice and performance test scores.

Table 3-1: Data Sets Used in Testing the Three Hypotheses

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Hypothesis 1</th>
<th>Hypothesis 2</th>
<th>Hypothesis 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAND</td>
<td>Using MC &amp; PA Tests</td>
<td>Test</td>
<td>Using MC &amp; PA Tests</td>
</tr>
<tr>
<td>NELS:88</td>
<td>Using MC Tests</td>
<td>Not Test</td>
<td>Using MC Tests</td>
</tr>
</tbody>
</table>

The advantages of the RAND data are the large student sample and test scores from both multiple choice and performance tests for the same students. However, the RAND survey has a small teacher sample and a limited number of covariates for the variables we are concerned may be linked to both hands-on science and achievement. Specifically, the RAND data contains an ability measure for students. The issue of past course-taking is less important as 8th grade students normally have taken a similar number and type of courses.

The National Educational Longitudinal Survey (NELS:88) will allow us to check and extend some of the results from the RAND analysis. NELS:88 followed a nationally representative sample of about 25,000 8th grade students in 1988 to 10th grade in 1990 to 12th grade in 1992. Students and teachers were surveyed on the level of hands-on science in the classroom. Students took a standardized multiple choice science test. Data were also collected on student demographics, past course work, ability, classroom level and school characteristics.
The NELS data will be used to test the first and third hypotheses. Multiple choice test scores will be used to determine whether there is a positive relationship between hands-on science and student achievement (Hypothesis 1) and whether there is a differential relationship for higher and lower ability students. These hypotheses will be tested for each of the three grades.

Like the RAND data, NELS:88 contains a large sample of students. Another advantage is its large sample of teachers. In addition, NELS:88 also contains more of the covariates that may be linked both to hands-on science and test scores. In this way, we can extend our analysis of the RAND data to see if the results are robust when these covariates are controlled. Further, we can extend our analysis to the upper grades and here. With this extension, two points must be considered. First, one of the theoretical rationales for hands-on science is the provision of concrete experiences in support of understanding abstract concepts. As students mature, they are better able to directly understand abstract ideas and require fewer concrete demonstrations (Piaget 1973; Gage and Berliner 1994; Lawson 1995). Under this theory, we would expect the relationship of hands-on science to test score to be weaker for older students. Comparisons of the 8th grade students from the RAND survey with the older NELS students, especially the 12th graders, may illustrate this theorized difference in the usefulness of hands-on science. Second, in middle school most students take the same science courses while in high school there is greater student variation in science course taking. For analyses using students in higher grades, more attention will need to be given to the covariate of past course-taking.
NELS:88 has some disadvantages as well. First, its 8th grade student survey lacks the items necessary to develop a student measure of hands-on work and its 10th grade survey lacks a key item. This constrains the value of comparing the two data sets when considering student reports of hands-on science. Second, it does not contain performance test scores among the three waves of data. This lack prevents us from testing Hypothesis 2 and focuses our testing of Hypotheses 1 & 3 on multiple choice tests. While we will not be able to draw any conclusions on performance tests and hands-on science and their differential relationship versus multiple choice tests, we will have further evidence on the links between hands-on science and the most widely used, standardized, measure of student achievement.

As noted earlier in this chapter, NELS:88 has been used to examine the relationship of hands-on science to multiple choice test scores. This study complements and extends that work. Through the use of multivariate analysis and a broad array of variables that may be linked to both hands-on science and achievement we extend those studies that lacked these characteristics while complementing, and possibly confirming, those studies which have them. Second, the bulk of the NELS research operationalized the concept of hands-on science using variables with strong face validity. This work develops a composite hands-on science variable using both face validity and factor analysis. By better addressing the measurement error in hands-on science variable we have greater confidence that we are indeed measuring the underlying construct of hands-on science and reduce the bias in the coefficient estimating the relationship of hands-on science and test score. Third, this study will include the teacher and student measures of hands-on science available from each of the three years of data whereas the past studies
did not. Fourth, this study will examine how ability may affect the relationship of hands-on science and test score, an issue to which the NELS data has not yet been applied.

The following two chapters describe the analysis using each data set, RAND and NELS:88, respectively. Each chapter discusses the data set in greater detail including the measures of hands-on science and test scores, along with the covariates. Each chapter provides a descriptive analysis of the data, a description of the models used in the analysis, the results from the models, and tests of robustness of these results.