DOCUMENT RESUME

ED 344 740 SE 052 388

AUTHOR McDuffie, Thomas E., Jr.
TITLE A Framework for the Impact Evaluation of the Carroll County Hands-on Elementary Science Program.
PUB DATE 91
NOTE 23p.; Prepared for Hands-on Elementary Science Dissemination Center, Hood College, Frederick, MD.
PUB TYPE Information Analyses (070) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Education; *Elementary School Science; *Evaluation Criteria; Evaluation Methods; *Experiential Learning; Measurement; Outcomes of Education; *Process Education; Science Education; *Skill Development

ABSTRACT

The objective of this investigation was to review impact evaluations related to activity based elementary science programs. Of particular focus is the measurement of process skills attainment. The report presents a survey of current impact evaluation procedure for process-oriented programs as described by recent literature, identification of current process evaluation instruments, and recommendations for the design of impact evaluation procedures for Hands-On Elementary Science (HES). Sections include: (1) "Tyler's Paradigm and Process Science"; (2) "An Evolving Paradigm for Process Evaluation"; (3) "Major Current Assessment Efforts-National Assessment of Educational Progress, State Based, and National Science Foundation"; (4) "Conclusions Regarding Process Assessment"; and (5) "Guidelines for Developing an Impact Evaluation". (30 references) (KR)

ED344740

A Framework for the Impact Evaluation of the Carroll County Hands-on Elementary Science Program

Prepared for

Hands-on Elementary Science
Dissemination Center
Hood College
Frederick, MD 21701

Prepared by

Thomas E. McDuffie, Jr., Ed.D.
St. Joseph's University
Philadelphia, PA 19131

Table of Contents

Introduction

Reviews of the impact of hands-on curriculum projects upon children's learning are almost universally positive. In a brief review of outcomes Shymansky and his colleagues (1982) forcefully concluded that children actively involved in science, especially the three major NSF sponsored programs, "achieved more, like science more, and improved their skills more than children in traditional, textbook based classrooms". Reinforcing statements were published in *Science and Children* two years later (Kyle, et. al., 1985; Orlich, 1985) In a broad based study Bredderman (1983) reported that participation in hands-on programs was associated with much higher scores on science process measures and creativity, and higher scores on content in science and mathematics as well as developing language skills. A meta-analysis confirmed these advantages of hands-on approaches (Bredderman, 1985) but found that attitudes were only slightly more positive than the control groups. On the district level the by-products, especially reading skills enhancement seemed to be the driving force for teaching science! Active involvement with science materials led to improved language acquisition and reading skills; certainly it must be part of the curriculum seemed to be the argument (Wellman, 1978).

The synergy with reading was not a new observation. Ground breaking research by Renner (1973) had clearly established the association between reading skills development and involvement with hands-on science. Pre-schoolers taught about the properties and attributes of objects using SCIS's Material Objects fared significantly better on reading readiness than participants in more traditional approaches. Problem solving such as interpreting posters, maps, or graphs reflected similar differences. SCIS was teaching far more than science. Related topics in reading, math and social studies were integral to the program.

Such results accompanied the general satisfaction with the curriculum and materials produced. Of concern was a major missing element-- evaluation instruments. The critical, and unifying differences in their approaches to instruction was the active pursuit of science. Scientific processes emerged as central. Paradoxically, evaluation of their attainment was not a major concern of those most involved with the programs. Project developers, by default, considered it secondary to creating the actual curricula. Science educators often channeled their energies toward teacher training and project implementation. Those in basic education struggling with the new course concentrated upon content learned rather than the attainment of process or problem-solving skills.

Current levels of interest in hands-on science may be traced to the many calls for school reform, but the positive benefits, scientific and non-scientific, are well understood by basic education. Local and state agencies lead the demand for improved outcomes. Concurrently, new modes of assessment are being sought which more validly, accurately, and realistically measure programatic outcome (Shavelson, et.al., 1990).

Purpose

The objective of this investigation is to review impact evaluations related to activity based elementary science programs. Of particular focus is the measurement of process skills attainment. Three facets will be presented.

1. A survey of current impact evaluation procedure for process oriented programs as described in recent literature.
2. Identification of current process evaluation instruments.
3. Recommendations for the design of an impact evaluation procedure for Hands-On Elementary Science.

Background

Undoubtedly, the writing of appropriate evaluation instruments as part of the overall NSF elementary science program would have been a great benefit. Just as obviously this was recognized by project leaders. Two factors, among others, explain their non-existance: the focus of the efforts and their underlying philosophies. When the somewhat frenzied era of project development is coupled with the fact that the chief architects were scientists, then the lack is understandable. Creation of a new, and more valid approach to communicating the essential components of their disciplines with children was central. Decisions had to be made regarding the appropriateness of content and the manipulatives needed to convey scientific processes. Time, energy, and other resources for outcomes evaluation were limited. Any evaluation was targeted to curriculum formation. In fact, Guba and Lincoln (1989) consider these programs the driving force in the change toward formative evaluation.

A major exception, of course, was the SAPA instruments developed to evaluate process skills. The operational definitions they employed became the foundation for process evaluation and test development. Given the behavioristic, outcome orientation of SAPA or, in Pepper's (1941) world view, mechanistic approach of the project this makes sense. Similarly, the more organicist and developmental philosophies undergirding both ESS and SCIS lead to another set of psychological beliefs about learning. The child, in this framework, should be placed in an enriched environment which both challenges and provides opportunities for growth. Such a framework places a premium upon observations and anecdotal records as the basis for evaluating learning.

Tyler's Paradigm and Process Science

The process of evaluation is, according to Ralph Tyler (1949), "essentially the process of determining to what extent the educational objectives are actually being realized....However,

since educational objectives are essentially changes in human beings, that is, the objectives aimed at are to produce certain desirable changes in the behavior patterns of the students then evaluation is the process for determining the degree to which these changes in behavior are actually taking place" (p.69). Fundamental to Tyler's approach was the matching of objectives and content. He related these components using a matrix. Hands-on programs, at the very minimum, required the addition of a third dimension to accommodate process.

When modified, Tyler's paradigm provided guidance to test makers who incorporated process dimensions into evaluation. Several noteworthy examples show how "science processes" are operationally defined. The foundational role of SAPA in identifying, defining and establishing direction to process evaluation is evident. In time item writing team composed of classsroom teacher supported by science educators and reading specialists became standard .

The Test of Science Process developed in the late 1960's by Robert Tannenbaum (1971) is an example of a pioneering process instrument. Although intended for junior high school students, it exemplifies the wedding of objectives to evaluation. Development preceded along the following steps.

1. Defining behaviors related to the basic processes-- observing, comparing, classifying, quantifying, measuring, experimenting, infering and predicting.
2. Content validation by experts who met pre-established guidelines.
3. Preparing of draft of items.
4. Conducting a pilot study.
5. Reviewing and revising items for the final form.
6. Administering the test and determining statistical parameters.

While Tannembaum's linked a range of junior high school content to the processes,

others limited content to specific programs (McLeod, et. al. 1975; Tobin and Capie, 1982). The opposite tack was taken by Molitor and George (1976) who wrote and field tested an instrument which included familiar objects and events but was content free. Theoretically students in grades four through six who had experienced hands-on science instruction enjoyed no advantage on this measure compared to those who did not. Content driven issues and concerns created dilemmas for these researchers as it later would for evaluators associated with the Assessment Performance Unit. In essence items or activities must be based upon some content which always creates a situation where some students have more relevant experiences.

Smith and Welliever (1990) linked their instrument to the science competency continuum prepared by the Clarion University of Pennsylvania Curriculum Group (Mechling, et. al., 1984). Fourth graders ability to answer items on thirteen process categories was measured: observing, classifying, infering, predicting, measuring, communicating, using space time relationships, defining operationally, formulating hypotheses, experimenting, recognizing variables, interpreting data, and formulating models. Content included a range of common material from the physical, earth and space and biological sciences. A team composed of ten teachers, science educators, and science supervisors used a workshop format to write the 65 multiple-choice items which became the instrument. The workshop began with training in the Pennsylvania competency continuum and practice in writing test items. Next individuals prepared three items for each area which were later critiqued and revised. Like readability, validity was determined by experts, but the input of classroom teachers ensured that test items matched what was actually taught rather than what was supposedly included. After a pilot administration and further revisions, the final version was prepared and tested.

An interesting extension of the definition of process highlights the relationships between problem solving, critical thinking and scientific processes (Ross and Maynes, 1983). By

implication, the roles of problem solving and critical thinking must also be considered when defining the scientific processes operationally. Experimental problem solving included: developing a focus (formulating a hypothesis), developing a framework (designing an experiment),judging the adequacy of collected data, recording information, observing relationships in data, drawing conclusions, making generalizations.

An Evolving Paradigm for Process Assessment

Valuable though these efforts were, several different manifestations of concerns with both the evaluation paradigm and process were voiced by such varied sources as science education researchers, classroom teachers, local and state administrators responsible for implementation and assessment as well as state and national policy makers (Shavelson, et. al., 1990). Researchers and policy makers seek nationally standardized, norm referenced instruments because of their need to make comparisons and generalizations confidently. Classroom teachers are primarily interested in individual achievement: Are the children learning 'the material'? Administrators charged with program assessment have broader concerns. What is science? What is the purpose of science instruction? Are beliefs about science woven through curriculum and instructional matters? Are they evident in evaluation? What is the purpose of assessment? How will/should results be used? What role(s) do stake holders-- policy makers, teachers, supervisors, science educators and consumers-- play in the assessment? What is the relative weighting of content and process? Are the processes conceived by test planners really being measured? What formative, summative or policy matters will be addressed as a consequence of assessment. What practical hurdles must be overcome? Nationally standardized tests fail to address many of these concerns worse they often created a gap between what was taught and measured.

Attempts to reconcile these concerns have led in different directions. Researchers have created a handful of process instruments, like those mentioned earlier, which ultimately address problems of limited scope. Teachers and local school districts have produced inventories for evaluating learning. Lastly large scale undertakings by the Educational Testing Service, the states of New York and Connecticut and the NSF have begun ground breaking approaches to assessment. Each of these will be discussed below.

Researcher's Concerns

At least one researcher's frustrations can be traced to invalid or suspect comparisons resulting from poor instrumentation. "The substance of test items often outweighed other considerations. Consequently, only 4 of 27 tests of science processes were rated as unbiased. All others were rated as favoring the laboratory program group. This result is a direct consequence of the fact that, for the most part, laboratory programs included the deliberate teaching of process while the control group programs did not. Further, the confounding of the influence of test format, standardization, and substance could not be resolved because no tests of process were nationally standardized and only 5 of 29 administrations of process tests were in a pencil and paper format" (Bredderman, 1985; p. 579). Granted the researcher's statistical needs exacerbated the problem, but the statement highlights the limitations, and lack of focus inhibiting process evaluation.

District Based Solutions

Two district based approaches to assessment resulted in very different solutions. Perhaps the most extensive set of inventories to date was prepared by the Fayette County Kentucky School District. A team of teachers, science educators and science supervisors created tests for each of the SCIS units (Atwood, et. al., 1984). Multiple choice items, more heavily content than process based, were written, reviewed, and revised to assure their

appropriateness, clarity of statements and readability. To reduce reading dependence, items were read out loud to children in levels one and two. Without a doubt these inventories were welcomed by classroom teachers. At the same time their multiple choice format, and relative emphasis upon content over process measurement has limitations for assessment.

Small (1988) used "the web of inquiry processes" to develop a grade specific test dubbed "an evaluation model". Eleven SAPA-like elements were included. Here also a summer teachers' workshop was used for item generation, but the product was far less extensive than the Fayette project.

Major Current Assessment Efforts

By comparison to today's concept of assessment these local initiatives are understandably primitive. Evaluation of student outcomes is necessary, but insufficient for activity based science curricula. If stakeholder needs are to be met and the relationship between evaluation, curriculum and instruction used beneficially, then assessment must be both formative and summative. Each of the major assessments is multi-faceted. By themselves multiple choice tests are clearly insufficient for impact evaluation because they provide too little information about thought processes additionally their validity is suspect.

Evaluation can take many forms-- informal observation, structured observations using check lists embedded in instruction, paper and pencil tests with multiple choice and open ended questions, and hands-on evaluation of science processes. Among other benefits, multiple approaches enables examination of childrens' question answering strategies. How were "correct" or "incorrect" responses generated? Miscue analysis provides a Rosetta stone of sorts for understanding item ambiguity and error patterns. While problem solving is hardly linear, it follows certain stages-- problem interpretation, problem reformulation, planning and carrying out a solution, recording and interpreting information and evaluating the solution.

Mistakes made early in the solution may be compounded even when proper strategies were employed. Results on the practicum indicate that students are more knowledgeable about the processes of science that previously thought (Murphy, 1990).

Educational policy makers need valid and reliable data which can be analyzed based upon a variety of subgroups-- classrooms, buildings, districts or demographic groups. Individual scores are a means not an end. Responsibility for learning has shifted discernably from students to the educators. Was instruction at the proper conceptual level? Were appropriate interventions employed? Was adequate time provided? Assessments must be designed to assist in instructional decision making rather than to judge individual status. "Today's assessment requires decisions that affect both the content and the pedagogy of tomorrow's instruction" (Harmon and Mokros, 1990; p.185).

Four major projects are in the vanguard of the assessment movement: the Nation Assessment of Educational Progress, New York and Connecticut State Departments of Education, and the new NSF programs. While distinctly different initiatives, a shared belief that assessment must be multifaceted joins them. Of particular relevance to Hands-on Elementary Science (HES) are the more summative aspects of their endeavors related to outcomes and attitudes.

Multiple choice tests remain the work horse for large group tests, but item selection has improved through better pre-testing often including discussions with respondents. Open ended questions, both short and long answer, permit participants to demonstrate problem solving strategies. Long held fears about reliability are being addressed using techniques established to measure writing samples (Stock & Robinson, 1989). Inclusion of hands on process assessments is the single most important innovative commonalty. Each project has incorporated a hands on portion which requires completion of various activities at a number of stations.

Specific procedures and necessary manipulative materials are provided for these activities.

National Assessment of Educational Progress

The single greatest influence on the style and content of the practical tests was imported from the United Kingdom (Murphy, 1990; NAEP, 1987). The Assessment of Performance Unit (APU) formerly housed at Kings College in London had accumulated about a decade of relevant experience prior to a pilot project inaugurated by NAEP (Blumberg, 1987). Given time pressures plus the quality of the APU's materials, ETS elected to utilize or adapt the British approach. As national leaders in the field, it is only natural that the NAEP approach is reflected in the state and NSF sponsored undertakings (Baron, et. al., 1989).

APU monitored performance on six science processes (Table 1). Three were pencil and paper, two processes were tied to student performance on a series of timed problems. Lastly individual experiments were monitored one-on-one. The NAEP format closely follows this outline, but is well worth reviewing for the excellence of multiple choice items plus the assessment approach used for scoring open ended statements.

State Based Efforts

Connecticut's Assessment of Educational Progress (CAPE) began in 1984-85. Its main stay was a multiple choice test which included a broad range of inquiry and content items, like the Pennsylvania continuum, from life sciences, physical sciences and earth and space science (Baron, 1990). (The practical component may have been added later.) To accommodate the broad range of content and to minimize costs, matrix sampling was employed, but all districts had the option to participate. The practicum was administered one-on-one and limited to 30 schools with 10 randomly selected participants from each grade level.

Table 1. APU Activity Categories

Category	Sub-category	Form
Use of graphs or symbolic representations	Reading information from graphs tables or charts	Written
	Creating graphs, tables or charts	
Using apparatus and measuring instruments	Using measuring devices Estimating quantities Following instructions	Group practical test
Observation	Making and interpreting observations	Written
Interpretation and application	Interpreting information Applying information to concepts in biology, physics, and chemistry	Written
Planning investigations	Planning both parts of and entire investigations	Written
Perform investigations	Perform investigations	Individual, practical

Adapted form Murphy, 1990; p. 152.

The New York Board of Regents as part of its massive effort to improve science instruction, implemented a state wide science curriculum taught by a hands-on approach. Accompanying implementation was the assessment of all fourth graders: The Elementary Science Program Evaluation Test (ESPET). By design it provides local and state agencies with an index of their science program's effectiveness. By extension concern is with group results as opposed to individual scores.

Five components make up the battery; two are required. A pencil and paper test containing 29 multiple choice, content items plus 16 based on process. The practicum contains 15 exercises at 5 stations: measuring basic physical properties, predicting, developing a classification scheme, making generalizations and making inferences. Optional components, fundamentally attitudinal, include student, teacher, and parent/guardian measures. Each building is responsible for setting up and administering the assessment, but the State Science Office has provided training opportunities, and sample materials.

National Science Foundation Programs

Evaluators are an integral part of the teams developing the new NSF programs (Harmon & Makros, 1990). Efforts are in their early stages for the most part, but evaluators have been a part of the efforts since inception. They have participated in forming strategies; ask the throny questions about purpose, definition and objectives; help match conceptual levels of children and content; and suggest instructional alternatives.

Conclusions Regarding Process Assessment

Given the above a number of specific conclusions can be made regarding the summative aspects of the most current practices in assessment.

1. There is no nationally standardized test of science process appropriate to Hands-on Elementary Science

2. The development of such an instrument would provide both a timely and valuable contribution
3. To be useful, the measure must be anchored to the objectives and content of HES.
4. A paper and pencil format has many advantages, but these should be complemented by the inclusion of a hands on component.
5. Assessment which is limited to a single grade level is incomplete.
6. To be manageable complete sets of evaluative materials--paper and manipulatives--should be provided.
7. Administration of practical tests require expertise uncommon among classroom teachers.

Guidelines for Developing an Impact Evaluation for HES

Outlined below is a framework for the development of a valid, reliable, and implementable impact evaluation of HES. The intent is to build a foundation for a state of the art outcomes measure given the developmental stage of the project. Summative is a more apt descriptor for the intent, but the evolving nature of HES suggests potential use for formative, instructional purposes. The outline which follows includes a statement of purpose, suggested format, and guidelines for implementation.

Purpose

Of prime concern is the assessment of the classroom impact of HES. Childrens' results when aggregated by class, district, or other unit can provide data to determine:

1. Attainment of the program's goals and objectives in terms of children's:
    a. Understanding of the processes of science;
    b. Knowledge and understanding of the content related to process instruction;

c. Ability to apply the processes of science;

2. Appropriateness of HES curriculum regarding:

a. An inquiry oriented philosophy of science;

b. Beliefs about developmental nature of psychological growth;

c. The varied populations implementing the program;

3. Deficiencies in terms of common misunderstandings

Format

A best approach to assessment would include measurement of cognitive outcomes and attitudes to science. The former would depend upon a paper and pencil test which combines a multiple choice component with open ended questions. Additionally, a practical component would simulate the instructional environment and requires demonstrated ability to apply processes. An optional addition would be pre-tests to help teachers diagnose weaknesses and develop intervention strategies. Related to this are check lists embedded in instruction which would guide teacher observations of class performance. Attitudinal aspects could measure children's feelings about science and instruction while inventories of teachers and administrators opinions and concerns could help guide ;implementation efforts. In both these cases, no new instruments need be developed as other efforts could be readily adopted or adapted to meet these needs.

Two separate impact assessments should be made. Although New York and Connectic and the NAEP are limited to the fourth grade, their purposes are more global. A more indep probing is a natural byproduct of a targeted evaluation. Further, the developmental distinctions between the primary and intermediate grades provide a natural division.

Development and Implementation

With the philosophy, developmental psychology and activity orientation of HES as given, validity seems the paramount consideration for the proposed evaluation. Does the evaluation

measure what is being taught? Good instruction and sound learning can produce poor results when mis-measured. "...assessment must be matched to the specific curriculum planned for a given setting or, if it can be determined, the curriculum actually delivered to the students" (Raizen, et. al., 1989). Two notions may be extracted from this. First a practical component alleviates concerns of ecological validity. Secondly, involvement of teachers bolsters confidence in content validly.

Composition of the evaluation instrument and possible pre-tests should involve a writing team composed of teachers, policy makers or implementers, science educators, plus reading and testing specialists. Full participation by teachers not only promotes validity but helps ensure credibility at the building level; moreover, it is in harmony with the programs development. Science educators provide both expertise and leadership. They keep the process on track and moving forward. Reading specialists offer guidance regarding wording and children's interpretation of items, and determine readability.

Writing can occur under a variety of arrangements, but summer workshops, prepared and structured by a leadership team, have been fruitful. Preparatory efforts by the leadership group are critical. If the proposed framework is employed, a fundamental need is to determine the distribution of questions and activities using the matrix in Table 2. First content and process emphasis should be established for each grade. Second content should be chosen for evaluating process skill attainment. Third the balance between multiple choice, open ended, and practical components of the test must be set. Lastly the group could select a range of model items to be used as standards. Existing questions on the NAEP, APU, Watson-Glazer Test of Critical Think or the Cornell Test, for example, could be used for model. If progress check lists are to be included, this group should prepare prototypes to be adapted for each grade or unit.

With this advanced work completed, a writing team which represented the diverse districts which have implemented HES could be gathered to complete the task. After training in teams, they would be charged with writing questions which would be reviewed, revised and rewritten. Reading experts would assist in determining appropriateness and areas needing revision. Simultaneously a sub-group could be adapting existing procedures and approaches for the hands-on portion of the evaluation including model materials kits. Finally, the instrument would be placed into final draft form based upon this effort.

Table 2. Relating question levels to science processes

NAEP	Modified Bloom's	Science Process
Knowing Science	Knowledge	Observing, measuring
	Understanding	Infering, communicating Predicting, Operational definitions
Solving Problems	Applications	Classifying, Using Space/time Relations Recognizing variables
Conducting Inquiries Formulating hypotheses	Higher Level	Interpreting data, Experimenting Formulating models

Piloting would occur during the middle of the following fall at the third and sixth grade levels at the school who had sent representatives to the writing conference plus others chosen to provide a proper cross section. Finally these results would be analyzed and the final form set for spring distribution.

End Comments

Based upon the literature and conversations with school officials, there is a demand for better evaluation of outcomes. Program accountability may finally be a reality. Hands-on Elementary Science should participate by developing a multi-faceted program impact evaluation. If this direction is chosen it would be advantageous to:

1. Become familiar with the Final Report of the pilot study by NAEP as well as the book edited by Hein. Both are cited among the references.

2. Communicate directly with those involved in existing state based assessments plus the evolving NSF efforts.

3. Develop an estimate of resource needs and possible support sources for completing the assessment such as the recently announced NSF initiative (NSF, 1991).

References

National Science Foundation. (1991). *Assessing Student Learning: Science, Mathematics and Related Technology Instruction at the Precollege Level in Formal and Informal Settings.* Washington, DC: National Science Foundation.

Atwood, R.K., Neal, A.A., & Oldhan, B. (1984). Developing classroom evaluation materials for SCIS. *Science Education 68*(2): 163-168.

Baron, J.B., Forgione, P.D., Jr., Rindore, D.A., Kruglanski, H., &
Davy, B. (1989). *Toward a new generation of student outcome measures: Connecticut's common core of learning assessment.* San Francisco, CA: American Education Research Association,.

Baron, J.B. (1990). What we learn from state assessments of elementary school science, ed. George Hein *The Assessment of Hands-on Elementary Science Programs.* Grand Folks: University of North Dakota, Center for Teaching and Learning.

Blumberg, F. et.al. (1986). *A pilot study of higher-order thinking skills assessment techniques in science and mathematics: Final report.* (ERIC Document Ed 278718) Princeton, NJ: Nation Assessment of Educational Progress.

Bredderman, T. (1982). Activity science--The evidence shows it matters. *Science and Children 20*(1): 39-41.

Bredderman, T. (1985). Laboratory programs for elementary school science: A meta-analysis of effects on learning. *Science Education 69*(4): 577-591.

Guba, E.C. & Lincoln, Y. S. (1989). *Fourth Generation Evaluation.* Newbury Park: Sage Publications.

Educational Testing Service. (1989). *Science objectives: 1990 assessment.* (ERIC Document Ed 309031). Princeton, NJ: National Assessment of Educational Progress.

Harmon, M. & Mokros, J. (1990). Assessment in the new NSF elementary science curricula: An emerging role, ed. George Hein *The Assessment of Hands-on Elementary Science Programs.* Grand Folks: University of North Dakota, Center for Teaching and Learning.

Kyle, W.C., Jr., et. al. (1985). What research says: Science through discovery: Children love it. *Science and Children 23*(2): 39-41

McLeod, R.J., Berkheimer, G.D., Fyffe, D.W., & Robinson, R.W. (1975). The development of criterion validated test items for integrated science process. *Journal of Research in Science Teaching 12*: 415-421.

Mechling, K., et al. (Ca.1984). *A recommended science competency Continuum for grades K to 6 for Pennsylvania schools.* Harrisburg, PA: Pennsylvania Department of Education.

Moliter, L.C. & George, K. D. (197[illegible]). Development of a Test of Science Process Skills. *Journal*

of Research in Science Teaching 18(5): 405-412.

Murphy, P. (1990). What has been learnt about assessment from the work of the APU science project? ed. George Hein. Grand Folks: University of North Dakota, Center for Teaching and Learning.

National Assessment of Educational Progress. (1987). *Learning by doing: A manual for teaching and assessing higher order thinking in science and mathematics.* (Report no: 17-HOS-80) Princeton, NJ: Educational Testing Service.

Orlich, D.C. (1985). Picking and Choosing. *Science and Children 23*(1): 10-12.

Pepper, S. (1942). *World Hypotheses: A Study in Evidence.* Berkeley: University of California Press.

Raizen, S. Baron, J., Champagne, A., Haertel, E., Mullis, E. & Oakes, J. (1989). *Assessment in Elementary School Science.* Washington, DC: National Center for Improving Science Education.

Renner, J.W., et. al. (1973). An evaluation of the Science Curriculum Improvement Study. *School Science and Mathematics 73*(4): 291-318.

Ross, J.A. and Maynes, F.J. (1983). Development of a test of experimental problem-solving skills. *Journal of Research in Science Teaching 20*(1): 63-75.

Shavelson, R.J., Carey, N.B., & Webb, N.M. (1990). Indicators of science achievement: Options for a powerful policy instrument. *Phi Delta Kappa 71*(9): 692-697.

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.

Small, L. (1988). Science process evaluation model. Presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA April 5-9). American Education Research Association.

Smith, K.A. (1987). The development of a science process assessment for fourth-grade students. Unpublished Doctoral Dissertation. State College: The Pennsylvania State University.

Smith, K.A. & Welliver, P.W. (1990). The development of a science process assessment for fourth-grade students. *Journal of Research in Science Teaching 27*(8): 727-738.

Stock, P.L. & Robinson, J.L. (1987). Taking on testing: Teachers as tester-researchers. *English Education 19*(2): 93-121.

Tannenbaum, R.S. (1971). The development of the Test of Science Process. *Journal of Research in Science Teaching 8*(2): 123-136.

Tobin, K.G. & Capie, W. (1982). Development and validation of a group test of integrated

science processes. *Journal of Research in Science Teaching 19*: 133-141.

Tyler, R.W. (1949). *Basic Principles of Curriculum and Instruction.* Chicago: University of Chicago Press.

Wellman, R.T. (1978). Science: A basic for language and reading development, ed Mary Budd Rowe. *What research says to the Science Teacher, Volume 1.* Washington, DC: National Science Teachers Association.