DOCUMENT RESUME

ED 387 519 TM 023 792

AUTHOR Saturnelli, Annette Miele; Repa, J. Theodore

TITLE Alternative Forms of Assessment in Elementary Science: The Interactive Effects of Reading, Race, Economic Level and the Elementary Science Specialist on Hands-On and Multiple-Choice Assessment of Science Process Skills.

PUB DATE Apr 95

NOTE 39p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).

PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS Academic Achievement; Cultural Differences; *Educational Assessment; Educationally Disadvantaged; *Elementary School Students; Ethnicity; Grade 4; Hands on Science; Intermediate Grades; *Learning; Manipulative Materials; Multiple Choice Tests; Poverty; Racial Differences; Reading Achievement; *Science Process Skills; Science Teachers; Science Tests; Sex Differences; Socioeconomic Status; *Test Construction; Test Reliability; Urban Schools

IDENTIFIERS *Alternative Assessment; Elementary Science Program Evaluation Test NY; New York City Board of Education; *Performance Based Evaluation

ABSTRACT

The specific focus of this research was to determine how the outcomes on two alternative forms of assessment (multiple-choice and hands-on/manipulative) for science process skills were related when students were grouped on the basis of sex, race/ethnicity, and poverty level. Subjects were 1,381 fourth graders in a culturally diverse city school district in New York state. Also explored was the relationship between the reading scores of students and their scores on the two alternative science assessments and the effect of the presence of an elementary science specialist on the student scores. The reliability of New York's Elementary Science Program Evaluation Test (ESPET) was also considered. All students performed better on the hands-on test, and the traditional gap associated with socioeconomic or racial/ethnic status was reduced by the hands-on test. Economic status appeared to have an effect on science achievement, and reading had a greater effect than poverty on science achievement. Results supported the reliability of the ESPET, but did not support the importance of an elementary science specialist to student achievement. Results show that all students can learn science given appropriate instruction. Thirteen tables and 17 figures present study findings. (Contains 48 references.) (SLD)

Alternative Forms of Assessment in Elementary Science: The Interactive Effects of Reading, Race, Economic Level and the Elementary Science Specialist on Hands-On and Multiple-Choice Assessment of Science Process Skills

by

Annette Miele Saturnelli
Newburgh (NY) Enlarged City School District

J. Theodore Repa
New York University

A Paper Presented at the Annual Conference
of the American Educational Research Association
April, 1995, San Francisco, California

Annette Miele Saturnelli, Ed.D.
Director of Science PreK-12
Newburgh Enlarged City School District
124 Grand Street
Newburgh, New York 12550
914-563-7462
914-563-7493 (f)

J. Theodore Repa, Ph.D.
Chair, Department of Administration, Leadership, and Technology
New York University, School of Education
239 Greene Street, Suite 300
New York, NY 10003-4721
212-998-5520
212-995-4041 (f)
repa@acfcluster.nyu.edu

Alternative Forms of Assessment in Elementary Science: The Interactive Effects of Sex, Reading, Race, Economic Level and the Elementary Science Specialist on Hands-on and Multiple-Choice Assessment of Science Process Skills

Introduction.

The publication of The Bell Curve (Hernstein & Murray, 1994), currently on the "best-seller" list, overtly contradicts a statement that many educators frequently hear and say: "All children can learn." The results of the research reported in this paper provide a counterargument for their stance. Regardless of economic level, gender, race/ethnicity, or reading level, this research provides evidence that all children can learn science given two prerequisites. First, to be certain that the science program is taught, designate specific teachers as 'science specialists' to deliver instruction, elementary teachers who want to teach science, who are enthusiastic about teaching science, who will give the time and energy to the teaching of science, and who will be accountable for doing so. This will provide the appropriate conditions for all children to have the opportunity to learn science. Second, in addition to traditional multiple-choice testing of science, provide an alternative, hands-on, performanced-based assessment. This will provide students with more than one way to demonstrate what they know and can do. Based upon the results of this research, it is clear that we need to broaden our concepts regarding how and by whom instruction is delivered and how we assess.

The research described in this paper was carried out in an attempt to resolve some of the descrepancies and confront some of the issues with which many science educators continue to struggie regarding assessment. However, when we assess and hold students responsible for what they have learned, we are also socially and intellectually obligated to address the assessment of their opportunity to learn the content and or skills which are being assessed. Educators must find and employ ways to measure the existence and the quality of the resources for teaching and learning science, as well as to identify and use alternative methods which allow students to better demonstrate what they know and can do. The literature reveals that some research has been done on the latter in the form of comparing performance on hands-on versus multiple-choice paper-and-pencil forms of assessment (Shavelson, Baxter, & Pine 1992; Doran, 1990; Doran & Tamir, 1992; Kuechle, 1990; Comber & Keeves, 1973). Other researchers (Doran & Tamir 1992; Kuechle, 1990; Comber & Keeves, 1973) have reported that students were observed to perform better on hands-on tests than on multiple-choice tests of science process skills. The research reported in this paper supports and goes beyond their work.

Using the New York State Elementary Science Program Evaluation Test (ESPET), the test score gap on the multiple-choice test that is observed between students of different economic, racial and ethnic backgrounds was found to be reduced on the hands-on test; and the hands-on test format section of the ESPET was able to discriminate between two groups of students, those who had had a science program with an elementary science specialist and those students who had not had a program with an elementary science specialist. Thus, it was observed that, when provided with the opportunity to learn, the test score gap between subgroups was reduced on the hands-on test. Students in science programs with a science specialist performed significantly better that those in programs without a science specialist, especially those students of racial/ethnic groups currently underrepresented in the sciences and those with low-reading and high poverty levels. When students were provided a science specialist, the opportunity to learn science increased for all subgroups of students, but because some subgroups exhibited a greater increase in scores with a science specialist than other subgroups, a decrease in the test score gap between subgroups was observed. The multiple-choice test of science process skills was not able to discriminate between the two groups of students (those provided with a science specialist and those without a science specialist) and the multiple-choice test of science process skills in no way indicated the impact of the science specialist on the various subgroups of students. Under the twin conditions of a science specialist (which provides the opportunity to learn science) and hands-on assessment techniques (which provide students the opportunity to demonstrate what they have learned), we observe that all students can learn science.

Alternative Forms of Assessment in Elementary Science: The Interactive Effects of Sex, Reading, Race, Economic Level and the Elementary Science Specialist on Hands-on and Multiple-Choice Assessment of Science Process Skills

Objectives.

The specific focus of this research was to determine, for fourth graders in a culturally diverse city school district in New York State, how the outcomes on two alternative forms of assessment (multiple-choice and hands-on/manipulative) for science process skills were related when students were grouped on the basis of sex, race/ethnicity, and poverty level. In addition, two subproblems were also explored: the relationship between the reading scores of students and their scores on the two alternative forms of science assessment; and the effect of the presence of an elementary science specialist on the scores attained by students on the two alternative forms of assessment. The research addresses twelve questions: Did the NYS ESPET prove to be a reliable test? What was the relationship between the multiple-choice and hands-on science process skill test scores? Did the measurement of science process skills vary with the method of assessment used? Was the test score gap between student subgroups reduced when students were tested in an alternative way? Were there differences in performance based upon race? Were there differences in performance based upon poverty level? Were there differences in performance based upon sex? What was the relationship between students' reading scores and their multiple-choice and hands-on science scores? Was test performance more affected by race or poverty? Was test performance more affected by poverty or reading? How did the presence of an elementary science specialist affect hands-on and multiple-choice scores? Does alternative assessment make a difference?

Conceptual Framework and Relationship to the Literature.

Educational theorists propose that assessment should match pedagogy. Hands-on tests of science process skills may be considered as coming closer to measuring what science educators want to measure (Doran, 1990; Kanis, 1988; Cizek, 1991; Petraitis, 1991; Kulm and Steussy, 1991, Shavelson, Baxter, and Pine, 1992, Meng & Doran, 1990; Mitchell, 1992a; Wiggins, 1989, 1990), i.e., the science process skills. The four-year study by Shavelson, Baxter, and Pine (1992) provides some of the first evidence that hands-on assessments measure aspects of science achievment that are different from those measured by multiple-choice tests. The First International Science Study (Comber and Keeves, 1973) and the Second International Science Study (Doran and Tamir, prepublication manuscript, 1992) reveal some of the first evidence that elementary students score higher on hands-on tests than on multiple-choice tests of science process skills. Keuchle (1990) and Marshall (1991) report similar results. Sattler (1988) reports that different subgroups of children (American Indian and Hispanic-American children) obtain higher scores on performance IQ tests than on verbal IQ tests. Hein (1987) conveys that most science testing at the elementary level conflicts with objectives and program emphasis. The teaching of the science process skills through hands-on science experiences should be assessed using hands-on, performance tasks and not by multiple-choice, paper and pencil tests. Pine (1990), Davis & Armstrong (1991) and Champagne (1990) support Hein and stress the close connection between assessment and pedagogy and the need to articulate how a particular form of assessment matches pedagogic beliefs. According to Champagne (1990), the closer the assessment task is to what one wants to assess, the closer the sores will be to a true measure of attainment of the skill or concept.

Wadsworth's (1984) explanation of Piaget's theory of cognitive development may be applied to models for assessing science process skills: the concrete operational child (ages 7-11) can use logical operations to solve only problems that involve concrete objects and events in the immediate present. Hands-on tests are at the concrete level; multiple-choice tests are at the abstract level. Students should be taught and tested at their cognitive level of development. The students in this study were fourth graders, most of whom were age nine and most likely in the middle of the concrete operational level.

School variables have also been shown to affect science achievment more than other subjects such as reading (Tamir, 1989). Zuzovsky and Tamir (1989) examined the relative status of school (alterable) variables versus home (fixed) variables and found that the contribution of school variables was both subject specific and system specific, e.g., school variables had more of an effect on science achievment especially in low socioeconomic schools. An elementary science specialist is a controllable school variable. Abell (1990, Williams (1990), and Hounshell and Swartz (1987) present their opinions for and against the use of elementary science specialists but there is little hard data that compare the science achievment of elementary students with and without a science specialist.

The ability of a child to read a science test may influence performance. Differences in reading ability exist between students of different economic, racial and ethnic backgrounds (Scott-Jones and Clark, 1986; Jones, 1984). Economically disadvantaged students were found to achieve more in activity-based science programs than in text-book based science programs (Beane, 1985; Bredderman, 1982; Shymansky, Kyle and Alport, 1982). Research of the literature revealed that limited information was available on how race, sex, and economic status might interact to affect the science achievement of blacks, females, and disadvantaged students (Oakes, 1990). Few researchers have been able to focus on both race and sex, and they have frequently confounded economic status with minority group membership. In the research reported here, test score data for students for whom all these characteristics were known and were available for analysis.

Methods and Data Sources.

This was a quasi-experimental study of 1381 fourth grade students in a city school district in New York State. (Table 11 describes the students in this study.) All fourth grade students in New York State take all three parts of the New York State Elementary Science Program Evaluation Test (ESPET) which contains two methods for assessment of science process skills: hands-on and multiple-choice. The ESPET, constructed by the New York State Department of Education and mandated to be administered each May to all fourth graders, is administered and scored according to standardized procedures as described in the document produced by the New York State Education Department entitled *Program Evaluation Test in Science Grade 4: Directions for Administering and Scoring*. The main and interactive effects of three learner attributes (sex, race/ethnicity, and poverty level) on each method of assessment were examined. Race/ethnicity group membership was assigned according to categories used by New York State for all reporting purposes: White, Black, Hispanic, Other. Poverty level was established as high poverty, low poverty, or no poverty according to the Free and Reduced Meal Policies Eligibility Guidelines as set forth by the NYS Education Department, Bureau of Food Management and Nutrition. To determine the effects of the three independent variables (sex, race/ethnicity, and poverty level) on each of the two dependent variables (multiple-choice score and hands-on score), the independent variables were organized into a 2 x 3 x 3 factorial design. In addition, the main and interactive effects of two other independent variables (reading level and the presence of an elementary science specialist) were also explored. Descriptive statistics such as group means and standard deviations were computed, various correlation coefficients were determined, and t-tests, analysis of variance, and Tukey post hoc procedures were carried out. Reliability coefficients (Cronbach alpha and split half) were determined for the test used (the 1989 New York State Elementary Science Program Evaluation Test) and for its subparts.

The 1989 ESPET consists of three parts. Part I has 29 items and assesses science content in multiple-choice format. Part II with 16 items uses the multiple-choice test format and assesses science process skills. Part III with 15 items also assesses science process skills but uses a hands-on format. Part I has a maximum raw score of 29. Part II has a maximum raw score of 16. Part III consists of 5 different stations with a variety of tasks for a maximum raw score of 22. Raw scores were converted and reported as percent correct on each part. Parts II and III both measured the process skills listed in the following paragraph except for two skills, creating models and replicating, which were not measured on either Part II or Part III.

Definitions of the process skills tested by the ESPET are given in the New York State Elementary Science Syllabus (1985): classifying, creating models, formulating hypotheses, generalizing, identifying variables, inferring, interpreting data, making decisions, manipulating materials, measuring (length, mass, volume, and temperature), observing, predicting, recording data, replicating, using cues, developing vocabulary, and using numbers. The ESPET was developed by committees of science educators (pre-college teachers and university researchers) from across New York State, not by commercial test publishers. The consultants wrote, edited, and selected the test items under the direction of members of the New York State Education Department Bureau of Science Education and Bureau of Testing. Manipulative items were developed and adapted from a pool of hands-on items from other assessment instruments {the First and Second International Association for the Evaluation of Educational Achievement Science Study (FISS and SISS) and the British Assessment of Performance Unit (APU)}. For the multiple-choice questions, writers submitted test items to a test pool. All test items were keyed directly to the 1985 New York State Elementary Science Syllabus. Other consultants edited the test questions and others selected the items to be field-tested in various schools across New York State. Other consultants chose the questions from those field-tested items for the final form of the test. This test construction process is used by the New York State Department of Education to construct all of its statewide tests and is assumed to produce an instrument of high content and construct validity.

Results.

The findings presented in the sections that follow are from a study conducted in a city school district in New York State where the ESPET scores of 1381 fourth graders were analyzed (Saturnelli, 1993).

Did the ESPET prove to be a reliable test?

A Cronbach Alpha of .87 was obtained for the entire ESPET (Parts I, II and III combined). For Part I, science content using multiple-choice format, a split-half reliability coefficient of .75 was obtained. For Part II, science process skills using multiple-choice format, a split-half reliability coefficient of .78 was obtained. For Part III, science process skills using hands-on format, a reliability coefficient of .72 (Cronbach's Coefficient Alpha) was obtained. The lower values for the subparts can be attributed to the smaller number of items on each part as compared to the number of items on the total test. In addition, the correlation between the total ESPET and the

Science Subtest of the Iowa Tests of Basic Skills (ITBS) was found to be .69, a moderately strong correlation according to Guilford (1956). This data gives support for the criterion related (concurrent) validity of the ESPET. Group difference data gives support for the construct validity of the hands-on test. Students who had a science program which included an elementary science specialist once every six days for instruction (N = 577) had significantly higher scores (t = 9.52, p = 0.001) on the hands-on science process skills section of the ESPET than those students who did not have a science specialist (N = 804). However, the presence of a science specialist did not significantly affect the scores of students on the multiple-choice science process skills section. The multiple-choice science process skills test was not able to discriminate between those students who did and those who did not have a science specialist but the hands-on science process skills test did detect differences between these two groups.

What Is the Relationship Between the Multiple-Choice and Hands-On Scores?

For the total sample of students in the study, a .51 correlation coefficient (PPMC) between the two tests of science process skills was obtained (Table 1). However, analysis of data disaggregated on the basis of race reveals interesting relationships otherwise not detected. For Hispanics, the relationship between their hand-on and the multiple-choice scores was higher than it was for the toal sample (r = .60) while for Blacks, the relationship was lower than it was for the total sample (r = .36). This higher correlation for Hispanic students and lower correlation for Black students between the hands-on and multiple-choice science skills scores must be further examined with the following information about reading scores (Table 2). For Hispanic students, the correlation between reading scores and their science process skills test scores is .59 (hands-on) and .70 (multiple-choice). For Black students, the correlation between reading scores and their science process skills test scores is .29 (hands-on) and .52 (multiple-choice). Hispanic students appeared to be more dependent upon reading for both the hands-on and multiple-choice tests and therefore performed about the same on both forms whereas the Black students, apparently less dependent upon reading on the hands-on test than on the multiple-choice test, performed better on the hands-on form than on the multiple-choice form. Because reading was found to be associated more with the multiple-choice test than with the hands-on test (Table 2), partial correlations were run controlling for reading. A partial correlation coefficient of .30 was obtained between the two tests of science process skills. These differences are discussed in further detail in the section on reading which follows. The correlations with the ITBS science subtest discussed in the preceeding section and the correlations with the ITBS reading scores discussed in this section indicate that, while measuring some of the same aspects of science achievment, the two tests of science process skills may very well be measuring some different aspects of science achievement.

Table 1. Relationship Between Multiple Choice (MC) and Hands-on (HO) Science Process Skills Test Scores for Various Student Subgroups

Group	HO mean	MC mean	PPMC r	p (2 tail)	N
Total sample	73	64	.51	.001	1381
Whites+Blacks+ Hispanics	73	63	.52	.01	1353
Whites	79	71	.41	.01	759
Blacks	66	54	.36	.01	368
Hispanics	65	56	.60	.01	227
Males	74	63	.52	.01	677
Females	73	64	.51	.01	704
White males	79	70	.44	.01	383
White females	78	71	.38	.01	376
Black males	67	54	.35	.01	174
Black females	66	55	.37	.01	194
Hispanic males	63	54	.59	.01	104
Hispanic females	66	57	.61	.01	123
High-poverty	65	55	.41	.01	388
Low-poverty	69	58	.47	.01	94
No-poverty	77	69	.50	.01	899

Table 2 Correlations: ITBS Reading Scores with Science Process Skills Test Scores by Sex and Race

Group	N	RDG mean	HO mean	r	r^2	MC mean	r	r^2
All students	1381	56	73	.47	.22	64	.64	.41
Males	678	55	73	.44	.19	63	.64	.41
Females	704	58	73	.51	.26	64	.63	.40
Whites+Blacks +Hispanics	1353	56	73	.47	.22	63	.63	.40
Whites	759	61	79	.39	.15	71	.58	.34
Blacks	368	50	66	.29	.08	54	.52	.27
Hispanics	227	50	65	.59	.35	56	.70	.49
High-poverty	388	49	65	.38	.14	55	.52	.27
Low -poverty	94	54	69	.41	.17	58	.61	.37
No-poverty	899	60	77	.45	.20	69	.63	.40
White males	383	59	79	.39	.15	70	.60	.36
White females	376	63	78	.39	.15	71	.54	.29
Black males	174	48	67	.28	.08	54	.49	.24
Black females	194	51	66	.33	.11	55	.57	.32
Hispanic males	104	48	64	.51	.26	54	.73	.53
Hispanic females	123	53	66	.68	.46	57	.68	.46

Does the Measurement of Science Process Skills Vary With the Method of Assessment?

Students of every subgroup except Others were found to score significantly higher on the hands-o test of science process skills than they did on the multiple-choice test of science process skills. In nearly every case, the difference was significant (Table 3).

Table 3 Variation in the Measurement of Science Process Skills for Student Subgroups

Group	HO mean	MC mean	Difference HO-MC	t-value	p	N	RDG mean
All students	73	64	9	17.2	.001	1381	56
Whites	79	71	8	11.8	.001	759	61
Blacks	66	54	12	11.3	.001	368	50
Hispanics	65	56	9	7.2	.001	227	50
Others	82	81	1	.17	.866	28	68
Males	74	63	11	13.1	.001	677	55
Females	73	64	9	11.9	.001	704	58
White males	79	70	9	8.9	.001	383	59
White females	78	71	7	7.8	.001	376	63
Black males	67	54	13	8.3	.001	174	48
Black females	66	55	11	7.7	.001	194	51
Hispanic males	63	54	9	5.2	.001	104	48
Hispanic females	66	57	9	4.9	.001	123	53
High-poverty	65	55	10	10.6	.001	388	49
Low-poverty	69	58	11	5.3	.001	94	54
No-poverty	77	69	8	13.3	.001	899	60

Was the Test Score Gap Between Student Subgroups Reduced When Students Were Tested in an Alternative Way?

The test score gap between various subgroups was reduced when students were tested using the hands-on format. The greatest reduction in test score gap was observed between the following groups: no-poverty White males and high-poverty White males; no-poverty Whites and high-poverty Whites; no-poverty White males and no-poverty Black males; no-poverty White females and no-poverty Hispanic females; White males and Black males; White females and Black females (Table 4).

Table 4 Test Score Gap Between Various Student Subgroups

Groups being compared	N	MC Test Score Gap	HO Test Score Gap	Gap Reduced by
White males no-poverty and	334			
high-poverty	41	13	6	7
Whites no-poverty and	652			
high-poverty	77	12	5	7
No-poverty females White and	322			
Hispanic	51	10	5	5
No-poverty males White and	334			
Black	74	14	10	4
Males White and	383			
Black	174	16	12	4
Females White and	375			
Black	194	16	12	4
Whites and	758			
Blacks	368	16	13	3
Males no-poverty and	457			
high-poverty	174	13	10	3
White females no-poverty and	322			
high-poverty	37	9	6	3
Black females no-poverty and	62			
high-poverty	112	7	4	3
No-poverty and	899			
high-poverty	388	14	12	2
Females no-poverty and	444			
high poverty	212	15	13	2
No-poverty females White and	322			
Black	62	12	10	2
High-poverty females White and	37			
Black	112	10	8	2
Females White and	375			
Hispanic	123	14	12	2
Whites and	758			
Hispanics	227	15	14	1
Blacks no-poverty and	137			
high-poverty	200	6	5	1
Males White and	383			
Hispanic	104	16	16	0
Black males no-poverty and	74			
high-poverty	104	4	4	0

Were There Differences in Performance Based upon Race?

When data was not disaggregated, Whites were observed to score significantly higher than Blacks and Hispanics on both the hands-on and multiple-choice tests of science process skills, and no significant difference was observed between Blacks and Hispanics. However, when data were disaggregated on the basis of sex, race, and poverty level, the following differences between Black and Hispanic racial subgroups were observed. For high-poverty groups, on the hands-on test, high-poverty Blacks scored significantly higher than high-poverty Hispanics ($t = 2.24$, $p = .026$) and high-poverty Black females scored significantly higher than high-poverty Hispanic females ($t = 2.14$, $p=.034$). Additional observations about the performance of racial groups were made on disaggregated data. Detailed discussion follows in the paragraphs below.

Were There Differences in Performance Based upon Poverty Level?

On both the hands-on and multiple-choice tests, no-poverty groups scored significantly higher than low- and high-poverty groups and no significant difference was observed between low- and high-poverty groups (Tukey, $p< .05$). For each racial group, as poverty level decreased the mean scores increased on both the hands-on and multiple-choice tests (Table 5, Figures 1 and 2). Poverty level appeared to have the greatest effect on the performance of Hispanic students on the hands-on test where the mean score of the no-poverty Hispanic students was observed to be 8.4 points higher than the performance of the high-poverty Hispanic students. Poverty level appeared to have affected the performance of White students more on the multiple-choice than on the hands-on test. On the multiple-choice test, the difference between the mean scores of high-poverty White students and no-poverty White students on the multiple-choice test was observed to be 11.3 points, whereas on the hands-on test the difference in mean scores between high-poverty and no-poverty Whites was only 6.1 points. This means that the hands-on test closed the gap between high-poverty and no-poverty White students by a factor of 1.9.

Table 5 Mean Scores by Race & Poverty Level on Hands-on and Multiple-Choice Tests of Science Process Skills

	Mean Scores (N)					
	Hands-on			Multiple-Choice		
	White	Black	Hispanic	White	Black	Hispanic
combined poverty levels	78.7 (759)	66.4 (368)	64.7 (226)	70.6 (759)	54.8 (368)	55.7 (226)
High-poverty	73.4 (77)	65.1 (201)	60.1 (107)	60.8 (77)	52.8 (201)	53.1 (107)
Low-poverty	74.5 (30)	61.8 (30)	70.2 (29)	62.6 (30)	52.5 (30)	56.2 (29)
No-poverty	79.5 (652)	69.4 (137)	68.5 (90)	72.1 (652)	58.2 (137)	58.5 (90)

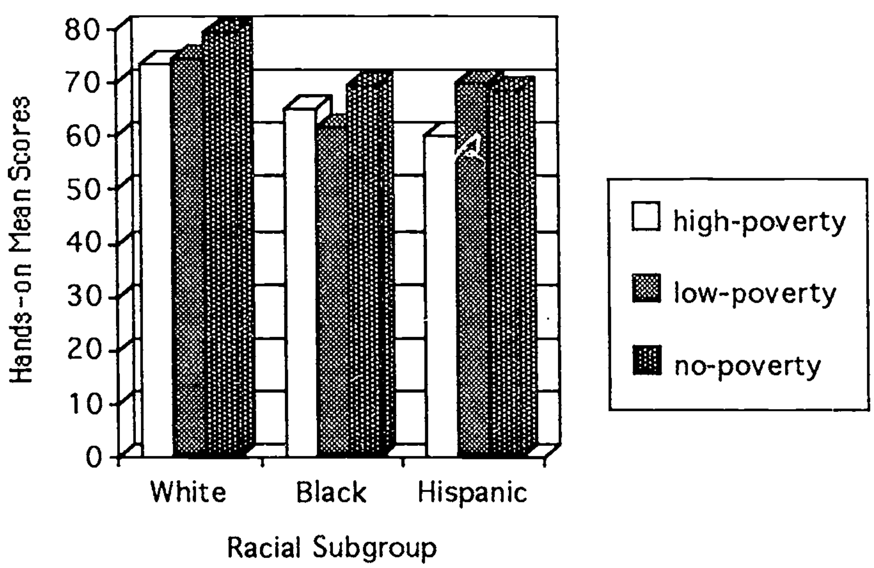

Figure 1. Effects of Poverty Level on Hands-on Test of Science Process Skills for Different Racial Subgroups

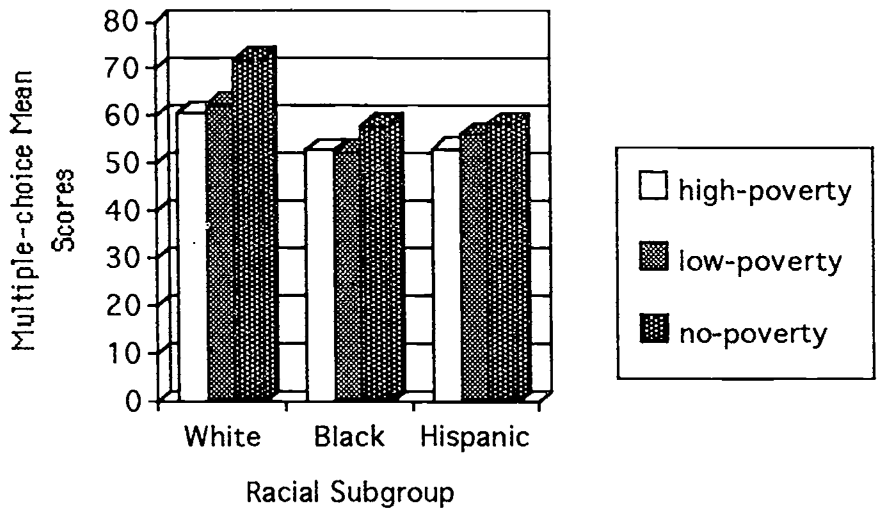

Figure 2. Effects of Poverty Level on Multiple Choice Test of Science Process Skills for Different Racial Subgroups

Were There Differences in Performance Based Upon Sex?

No significant differences were observed between the scores of males and females on either the hands-on test or the multiple-choice test until data disaggregated on the basis of sex, race and poverty level were analyzed. The following results were then observed. On the hands-on test, no-poverty Hispanic females scored significantly higher than no-poverty Hispanic males ($t = 2.59$, $p = .011$). On the multiple-choice test, no-poverty Hispanic females also scored higher that no-poverty Hispanic males but the difference was not significant at the .05 level ($t = 1.80, p = .075$). Other differences were observed between males and females on the multiple-choice test, where for no-poverty students without a science specialist, females scored significantly higher than males ($t = 1.97$, $p =.049$). With a science specialist, however, this gap between no-poverty males and females did not exist.

What Is the Relationship Between Students' Reading Scores and Their Scores on the Multiple-Choice and Hands-On Scores?

For all students, performance on the hands-on test of science process skills was found to be less associated with reading than was performance on the multiple-choice test of science process skills. The correlation coefficient for reading with the hands-on test was found to be .47 while the correlation coefficient for reading with the multiple-choice test was .64 (Table 2).

Scores of Hispanic students appeared to be more associated with reading than the scores of Black and White students. Both the hands-on and multiple-choice scores of high-poverty students were found to be less associated with reading than were the scores of no-poverty students (Table 2).

A complex pattern of differential performance in reading was observed. In middle- and low-reading groups, significant differences ($p < .05$) in reading scores were found to exist between males and females in favor of females. However in the high-reading group, a significant difference ($p < .01$) in reading scores was also found to exist but in favor of males (Table 6, Figure 3).

Table 6 T-tests: Differences in Science Test Performance and Reading Level Between Student Subgroups Categorized by Sex and Reading

Sub-group	N	Hands-on Skills mean	Hands-on Skills t-value	Hands-on Skills 2-tail prob	Multiple-Choice Skills mean	Multiple-Choice Skills t-value	Multiple-Choice Skills 2-tail prob	Reading ITBS mean	Reading ITBS t-value	Reading ITBS 2-tail prob
High-reading										
Male	208	83.9	1.77	.078	81.4	2.00	.046	76.6	2.64	.009
Female	257	81.7			78.7			74.6		
Middle-reading										
Male	212	73.8	.61	.540	64.2	1.73	.084	55.1	-2.11	.035
Female	251	72.9			61.6			56.0		
Low-reading										
Male	257	65.4	2.51	.013	49.8	.71	.478	36.6	-1.97	.049
Female	197	60.6			48.5			38.1		

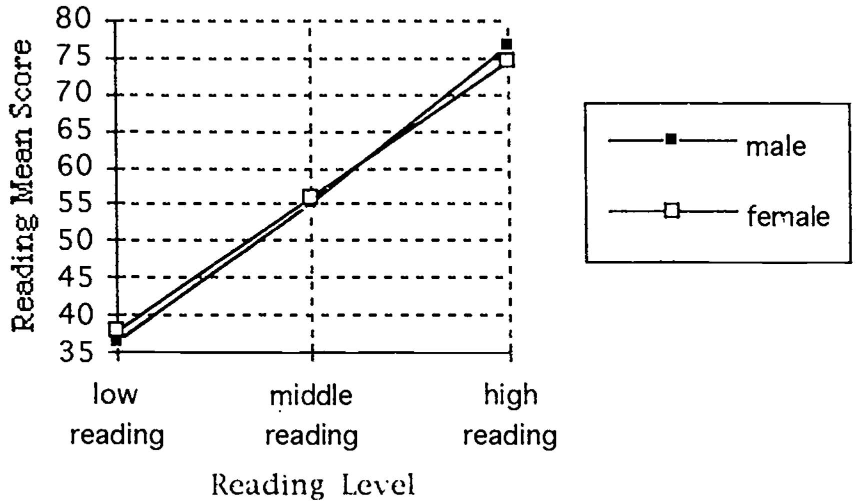

Figure 3. Differences in Performance of Males and Females of Different Reading Levels on the Iowa Tests of Basic Skills Reading Subtest

On the hands-on test of science process skills, for every reading level, males scored higher than females but the only significant difference was at the low-reading level ($t = 2.51$, $p = .013$) (Table 6, Figure 4).

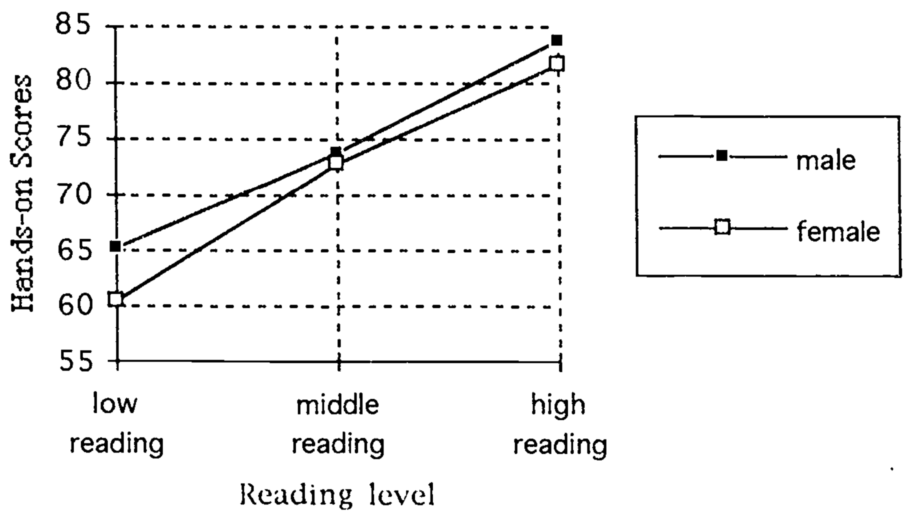

Figure 4. Differences in Performance of Males and Females of Different Reading Levels on Hands-on Test of Science Process Skills

On the multiple-choice test of science process skills, for every reading level, males also scored higher than females, but the only significant difference was in the high-reading group (t = 2.00, p = .046) (Table 6, Figure 5).

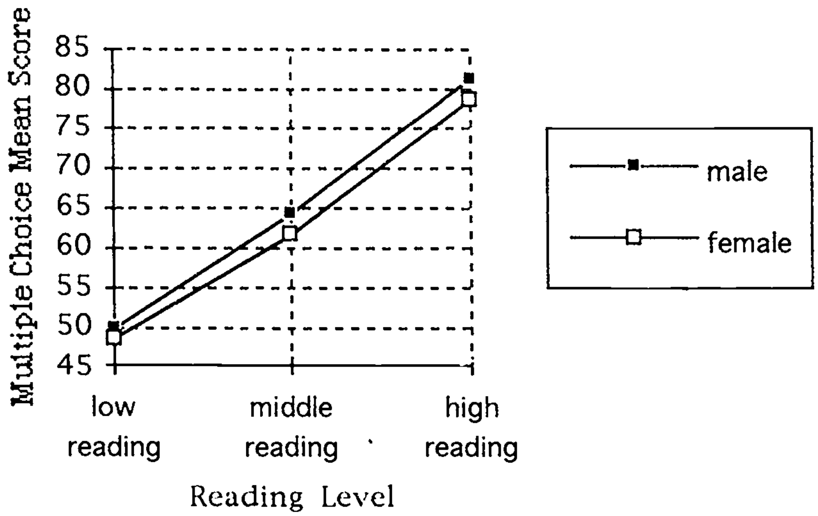

Figure 5. Differences in Performance of Males and Females of Different Reading Levels on Multiple-Choice Test of Science Process Skills

Significant two-way interactive effects of race and reading were observed for both the hands-on and multiple-choice tests. In the low-reading group, Hispanic students scored below White and Black students on both the hands-on and multiple-choice tests of science process skills. However, in the middle- and high-reading reading groups, Hispanic students scored above Black students but below White students on both the hands-on and multiple-choice tests (Tab' ; 7 and Figures 6, 7).

For each racial group as reading scores increased, scores increased on both the hands-on and multiple-choice tests of science process skills (Table 7, Figures 8, 9). Examination of the data in Table 2 as well as the data in Table 7 reveals that although reading affected the performance of all students, it appeared to have the greatest effect on Hispanic students. Mean score differences on the hands-on test between low-reading and high-reading Hispanic students was 29.5 points. Mean score differences on the multiple-choice test between low-reading and high-reading Hispanic students was 35.5 points.

On the hands-on test, the gap between mean scores of all White and all Hispanic students was 14.0 points, whereas the gap between scores of high-reading White and high-reading Hispanic students was only 3.2 points. Therefore, the gap between scores of White and Hispanic students on the hands-on test is reduced by 77% when reading is removed as a factor. On the multiple-choice test, the gap between mean scores of all White and all Hispanic students was 14.9 points, whereas the gap between scores of high-reading White and high-reading Hispanic students was only 5.3 points. This may be interpreted to mean that the gap between scores of White and Hispanic students on the hands-on test is reduced by 65% when reading is removed as a factor. Further analysis of the scores on the hands-on and multiple-choice tests for high-reading White and high-reading Hispanic students reveals that because there is a 3.2 point difference between the mean scores on the hands-on test and a 5.3 difference between scores on the multiple-choice test, by using the hands-on test, the gap between high-readingWhite and Hispanic students was reduced by 40%.

On the hands-on test, the gap between the scores of all White and all Black students is 12.3 points, whereas the gap between the scores of high-reading White and high-reading Black students is 9.3. This may be interpreted to mean that when reading is not an obstacle, the gap between the performance of White and Black students on the hands-on test is reduced by 24%. On the multiple-choice test, the gap between the scores of all White and all Black students is 14.9 points, whereas the gap between the scores of high-reading White and high-reading Black students is 9.5. This may be interrpreted to mean that when reading is not an obstacle, the gap between the performance of White and Black students on the multiple-choice test is reduced by 36%.

Table 7 Interactive Effects of Race and Reading on Science Process Skills Tests

	Mean Scores					
	Hands-On			Multiple-Choice		
	White (N)	Black (N)	Hispanic (N)	White (N)	Black (N)	Hispanic (N)
Low-Reading	71.5 (171)	61.8 (180)	51.5 (98)	56.4 (171)	46.6 (180)	40.9 (98)
Middle-Reading	76.1 (253)	68.7 (122)	71.6 (83)	65.5 (253)	57.4 (122)	61.8 (83)
High-Reading	84.2 (335)	74.9 (66)	81.0 (45)	81.7 (335)	72.2 (66)	76.4 (45)
All students	78.7 (759)	66.4 (368)	64.7 (226)	70.6 (759)	54.8 (368)	55.7 (226)

Note: For Hands-On Test, Race by Reading ANOVA $F = 10.399$ $p = .0001$; for Multiple-Choice Test, Race by Reading ANOVA $F = 6.321$, $p = .0001$.

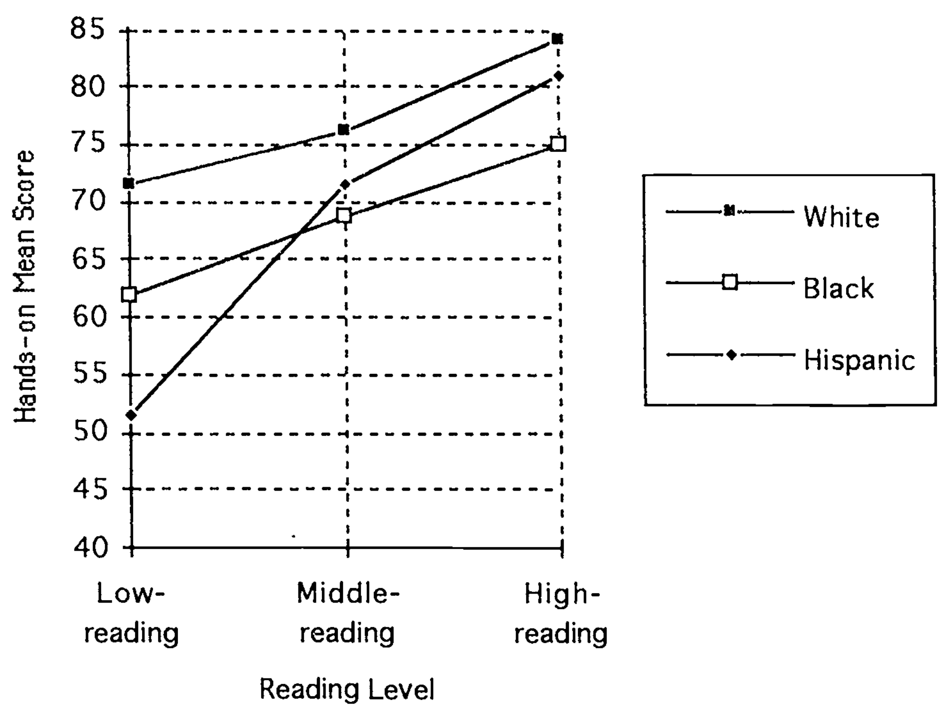

Figure 6. Interactive Effects of Race and Reading on Hands-on Test of Science Process Skills

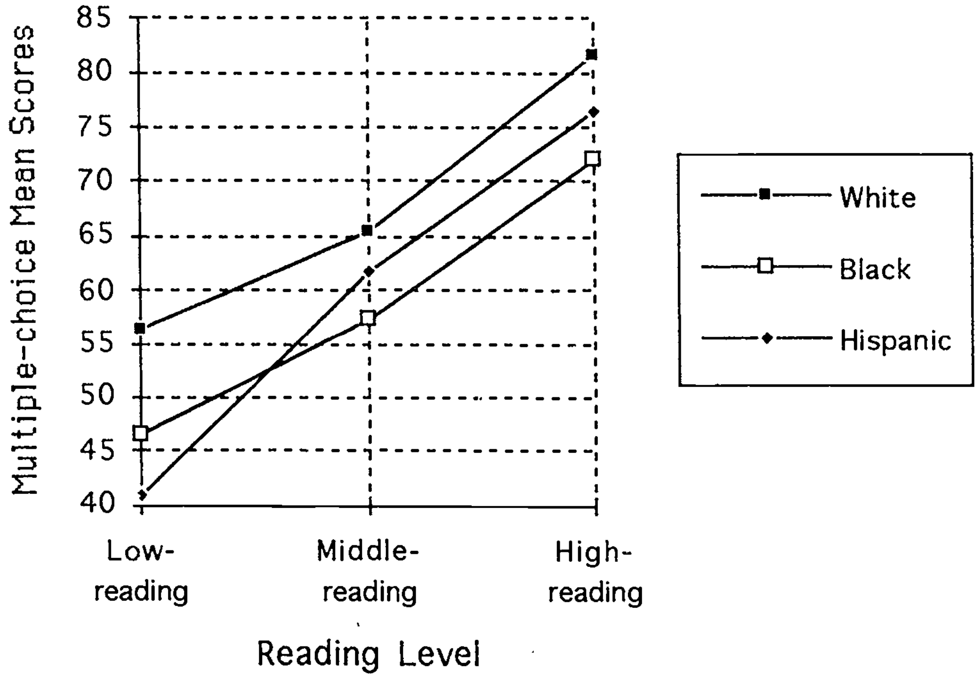

Figure 7. Interactive Effects of Race & Reading on Multiple-Choice Test of Science Process Skills

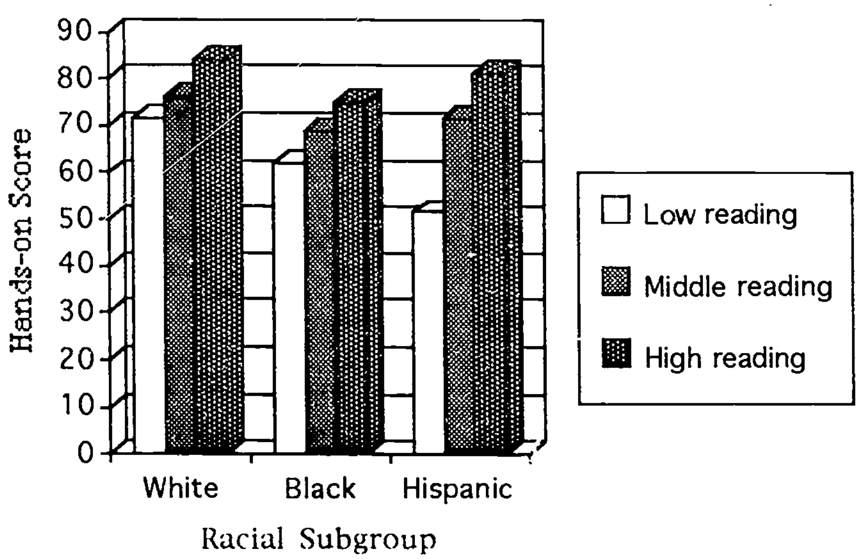

Figure 8. Effects of Reading Level on Hands-on Test of Science Process Skills for Different Racial Subgroups

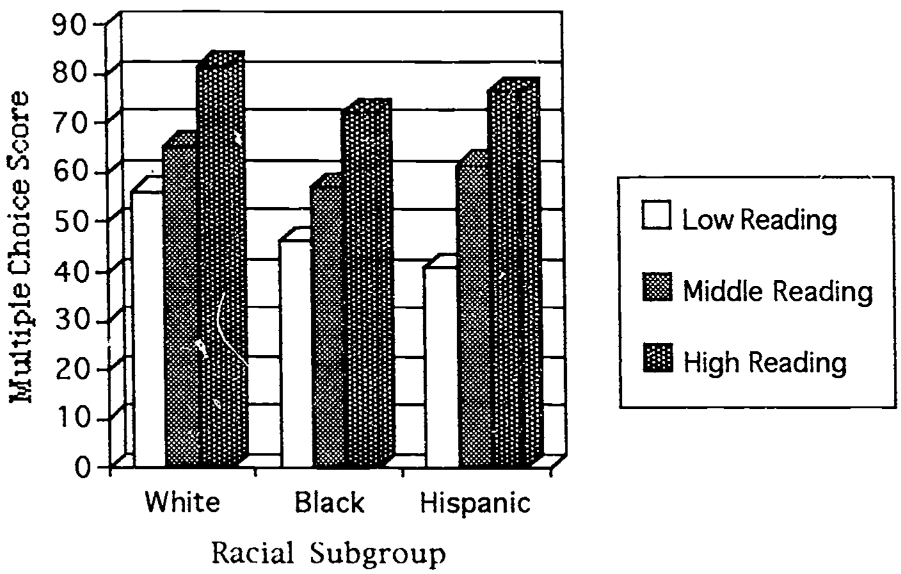

Figure 9. Effects of Reading Level on Multiple Choice Test of Science Process Skills for Different Racial Subgroups

Was Test Performance Affected More by Race or Poverty?

Poverty level was observed to have a greater effect than race on both the hands-on and multiple-choice science process skills test scores. Across each racial group (White, Black, Hispanic) scores were found to increase from high-poverty to no-poverty levels (Table 5). In all cases, the gap between racial subgroups was less where scores of students of the same poverty levels were compared than when scores of students of combined poverty levels were compared (Table 8).

Table 8 Effect of Poverty on Test Score Gap for Various Student Subgroups

		Mean Scores			
	N	Hands-on score	Gap	Multiple-choice score	Gap
All whites	759	78.7	12.3	70.6	15.8
All Blacks	368	66.4		54.8	
High-poverty					
Whites	77	73.4	8.3	60.8	8.0
Blacks	201			52.8	
No-poverty					
Whites	652	79.5	10.1	72.1	13.9
Blacks	137	69.4		58.2	
All Whites	759	78.7	14.0	70.6	14.9
All Hispanics	226	64.7		55.7	
High Poverty					
Whites	77	73.4	13.3	60.8	7.7
Hispanics	107	60.1		53.1	
No-poverty					
Whites	652	79.5	11.0	72.1	13.6
Hispanics	90	68.5		58.5	

Was Test Performance Affected More by Poverty or Reading?

Reading was found to have a greater effect than both poverty level and race on both the hands-on and multiple-choice science process skills tests. Across all racial groups, scores of no-poverty students were found to increase from low- to middle- to high-reading levels. Students at the no-poverty level were more affected by low-reading ability than students of high-reading level were affected by high-poverty (Table 9). On the hands-on test, for each race and sex, the mean score for the no-poverty low-reading group was lower than the mean score for the high-reading high-poverty group. Females appeared to be more affected by reading (about two times) than males.

Table 9. Hands-on Scores for Students With Different Reading and Poverty Levels

High Reading	White male	White female	Black male	Black female	Hispanic male	Hispanic female
High-poverty	78.6	82.7	70.5	73.0	72.8	79.8
No-poverty	86.7	83.0	79.3	77.9	81.7	84.0

No-Poverty	White male	White female	Black male	Black female	Hispanic male	Hispanic female
Low-reading	72.2	70.1	62.9	59.5	49.9	60.6
Middle-reading	77.2	77.3	71.5	70.0	70.6	76.9
High-reading	86.7	83.0	79.3	77.9	81.7	84.0

	White male	White female	Black male	Black female	Hispanic male	Hispanic female
High-poverty/ high-reading	78.6	82.7	70.5	73.0	72.8	79.8
No-poverty/ low-reading	72.2	70.1	62.9	59.5	49.8	60.6
Difference	6.4	12.6	7.6	13.5	22.9	19.2

How Does the Presence of an Elementary Science Specialist Affect Hands-On and Multiple Choice Scores?

Compared to students in a program without a science specialist, those students in a program with a science specialist were found to achieve significantly higher scores on the hands-on test (Part III of the ESPET), on the multiple-choice science content section of the ESPET (Part I), on the combined two multiple-choice sections of the ESPET (Parts I and II), on the total ESPET (Parts I, II, and III), and on the Iowa Tests of Basic Skills Science Subtest (Table 10, Figure 10).

Table 10 T-tests: Differences in Science Test Scores With and Without a Science Specialist

Science Test	Mean with Specialist	Mean without Specialist	t-value	2-tail prob
ESPET hands-on science process skills	78	69	9.52	.0001
ESPET multiple-choice science process skills	65	64	1.29	.196
ESPET multiple-choice science content	63	61	2.69	.008
ESPET multiple-choice skills plus content	64	62	2.84	.005
ESPET total test	69	64	5.65	.0001
ITBS science subtest	49	46	3.41	.001

Note: With a science specialist, N = 577; without a Science Specialist, N = 804.

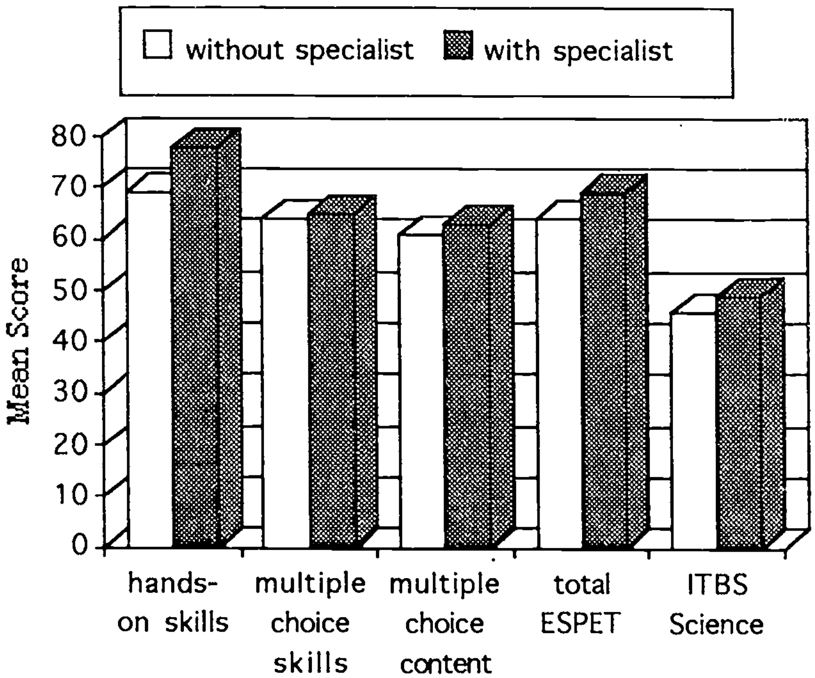

Figure 10. Effect of Science Specialist on Science Test Scores of Fourth Grade Students.

The mean score for students with a science specialist was also higher on the multiple-choice science process skills section of the ESPET (Part II), however the difference was not significant.

When disaggregated data is examined, it is observed that various student subgroups with a science specialist also score higher: males and females, Figure 11; high-, low-, and no-poverty groups, Figure 12; whites, blacks and Hispanics, Figure 13; no-poverty groups and high-reading groups, Table 11; low- , middle-, and high-reading groups, Table 12.

Table 11 Effect of Science Specialist on Science Test Scores for Racial Subgroups with No-Poverty and High-Reading Levels

	Mean Scores					
	Hands-on			Multiple-Choice		
	White (N)	Black (N)	Hispanic (N)	White (N)	Black (N)	Hispanic (N)
all students	78.7 (759)	66.4 (368)	64.7 (226)	70.6 (759)	54.8 (368)	55.7 (226)
without Specialist	75.5 (443)	62.8 (187)	60.0 (159)	69.7 (443)	55.8 (187)	54.2 (159)
with Specialist	83.0 (316)	70.2 (181)	76.0 (67)	71.8 (316)	53.7 (181)	59.2 (67)
No Poverty without specialist	76.5 (377)	65.2 (64)	65.3 (63)	71.2 (377)	57.8 (64)	57.6 (63)
No Poverty with specialist	83.5 (275)	73.0 (73)	76.0 (27)	73.2 (275)	58.5 (73)	60.5 (27)
High Reading without specialist	81.9 (183)	68.4 (32)	78.0 (28)	81.7 (183)	70.4 (32)	75.8 (28)
High Reading with specialist	87.0 (152)	81.0 (34)	85.9 (17)	81.9 (152)	73.8 (34)	77.4 (17)
No Poverty and High Reading without specialist	82.8 (165)	70.0 (15)	81.6 (14)	82.2 (165)	71.4 (15)	78.3 (14)
No Poverty and High Reading with specialist	86.8 (136)	84.5 (22)	86.2 (8)	82.4 (136)	76.0 (22)	82.1 (8)

Table 12 Interactive Effects of Race, Reading, Science Specialist Hands-on Tests of Science Process Skills

	Mean Scores								
	Low-Reading			Middle-Reading			High-Reading		
	White (N)	Black (N)	Hispanic (N)	White (N)	Black (N)	Hispanic (N)	White (N)	Black (N)	Hispanic (N)
With Specialist	76.8 (69)	65.4 (88)	68.8 (24)	81.2 (95)	71.3 (59)	76.2 (26)	87.0 (335)	81.0 (34)	85.9 (17)
Without Specialist	68.0 (102)	58.3 (92)	45.9 (74)	73.0 (158)	66.3 (63)	69.5 (57)	81.9 (183)	68.4 (32)	78.0 (28)

Note. ANOVA $F = 2.130$ $p = .075$.

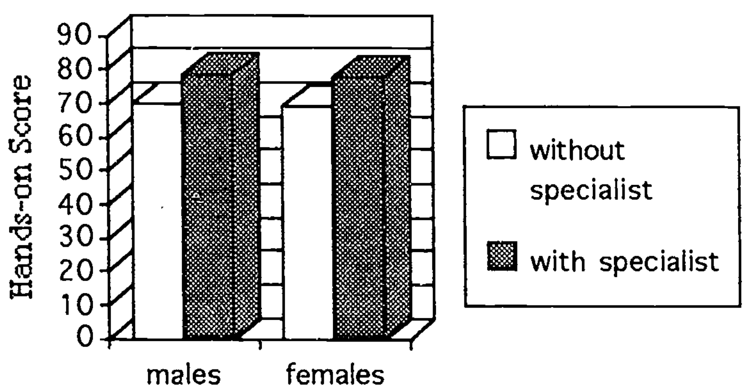

Figure 11. Differences in Performance on Hands-on Test of Science Process Skills for Males and Females With and Without a Science Specialist

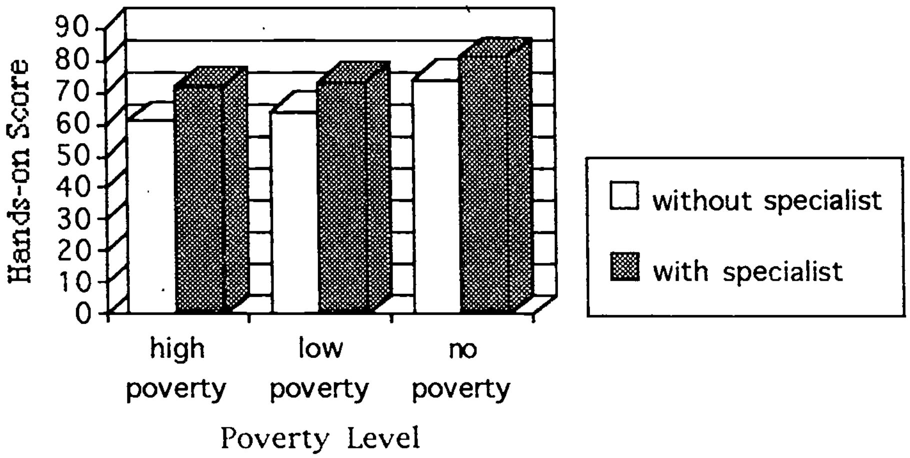

Figure 12. Hands-on Test Scores of Students of Different Poverty Levels With and Without a Science Specialist

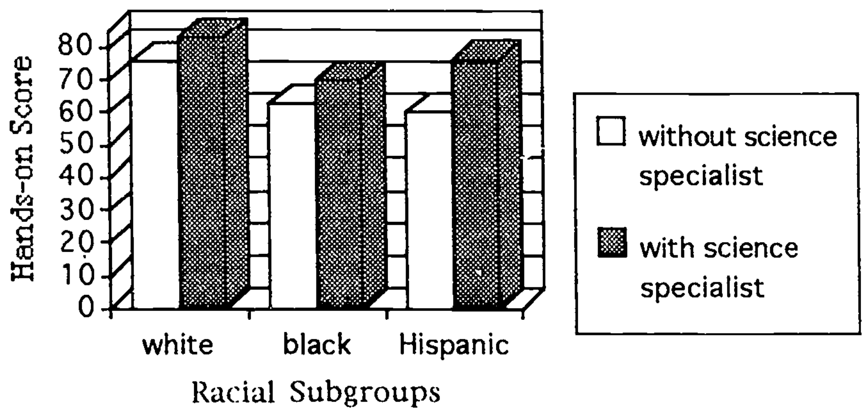

Figure 13. Hands-on Test Scores of Students of Different Racial Subgroups With and Without a Science Specialist

Numerous two-way and three-way interactive effects were observed between race and the science specialist and sex, race and the science specialist (Saturnelli, 1993). For example, see Figure 14.

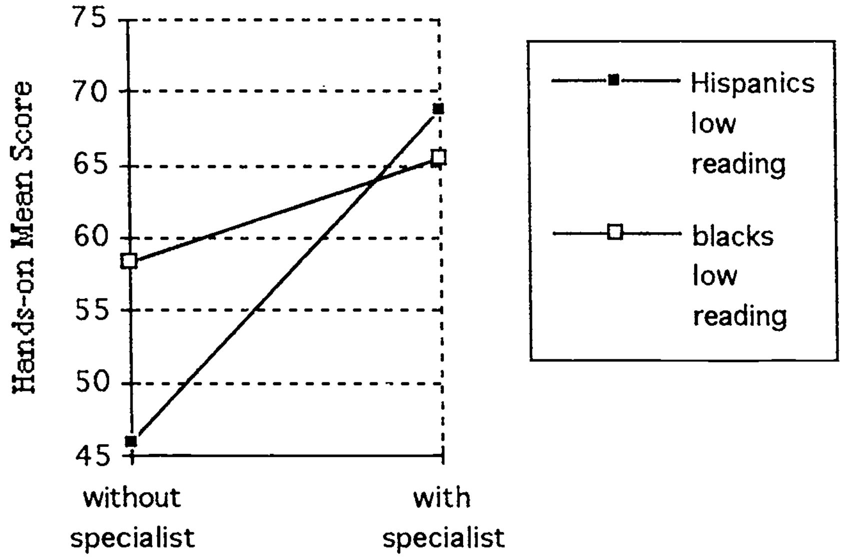

Figure 14. Interactive Effects of Race, Reading, and Science Specialist on hands-on Test of Science Process Skills: Blacks and Hispanics with and without a Science Specialist.

Does Alternative Assessment Make a Difference?

Students of all reading abilities and all poverty levels in all racial subgroups appear to be able to demonstrate what they know and can do better on the hands-on test than on the multiple-choice test of science process skills (Tables 5, 6 and 7; Figures 1, 2, 8 and 9). The variation in performance on the two tests of science process skills resulted in a reduction in the test score gap between various subgroups of students including the following: no- and high-poverty Whites; White and Black males; no-poverty White and Hispanic females; White and Black females (Table 4).

Conclusions.

All students performed better on the hands-on test of science process skills test, a test that was found to not rely heavily on reading, and that was at their cognitive level of development. The test score gap that has been observed to exist between students of different economic and racial or ethnic backgrounds was found to be reduced. On the hands-on test, economically disadvantaged (high poverty) students were provided the opportunity to demonstrate what they knew and could do and the gap between low-reading level and high-reading level students of all racial groups was less than on the multiple choice test.

The hands-on test of science process skills may be considered as coming closer to measuring what science educators want to measure (Doran, 1990; Kanis, 1988; Cizek, 1991; Petraitis, 1991; Kulm and Stuessy, 1991; Maeroff, 1991; Shavelson, Baxter, and Pine, 1992), because it appears to be less dependent upon reading ability than the multiple-choice test of science process skills; because it appears to be a more developmentally approriate test (more concrete, less abstract) for fourth graders; and because it matches instruction, especially in hands-on science programs where there is a science specialist.

The results of this study provide data needed to answer part of the question proposed by Jeannie Oakes (1990) who asks if it is race or economic status that has a greater effect on science and math achievement. For science, it appears to be that the answer is economic status. Within each racial group, test scores were found to increase significantly from high- to no-poverty levels and the gap between racial subgroups was less when scores of students of the same pover'· level were compared than when scores of racial subgroups of combined poverty levels were compared. The results of this study also allow the concerns about reading, expressed by Scott-Jones and Clark (1986) and Tolman, Sudweeks, Baird and Tolman (1991), to be addressed. Data obtained leads to the conclusion that reading has an even greater effect than poverty level on science achievement; and, as Scott-Jones and Clark suggest, a complex pattern of differential performance in reading exists for males and females. For low-reading ability students, females scored higher than males whereas for high-reading ability students, males were found to significantly score higher than females. In addition, the results of this study corroborate the work of Doran and Tamir (1992) and Shavelson, Baxter, & Pine (1992) regarding the mode of assessment of science process skills. Hands-on test scores of all students were found to be higher than their multiple-choice test scores of science process skills. The difference was greater for some subgroups of student than for others.

The results presented here (Table 10) support the studies of Beane (1985), Bredderman (1983), and Kyle, Shymansky and Alport (1982) who found that economically and/or educationally disadvantaged students in hands-on activity-based science programs performed better than those in text-book-based programs. From the results of this study it can be concluded that students in science programs with a science specialist, where the focus is teaching science through hands-on experiences, performed significantly better than those in programs without a science specialist, especially those students of low-reading and high-poverty levels. The results of this study also add strength to the conclusions drawn by Tamir (1989) and Zuzovsky and Tamir (1989) and Staver and Walberg (1986) regarding the effect of alterable school variables on certain subjects such as science, especially in low socio-economic schools. The findings of this study also provide information for those educators involved in the debate about the elementary science specialist (Abell, 1990; Hounshell & Swartz, 1987). It was found that with elementary science specialists, science was assured of being taught, and it was apparently learned. When students were provided a science specialist, the opportunity to learn science increased for all subgroups of students. All students demonstrated that they had learned more science with a science specialist than without but because some subgroups exhibited a greater increase in scores with a science specialist than other subgroups the gaps between subgroups which differ on the basis of race, poverty, and reading is reduced. See Figures 14, 15 and 16 where scores of various subgroups of students are compared with and without a science specialist. Figure 17 clearly shows that when the three factors (poverty, reading, and science specialist) are accounted for, the gap between racial subgroups is nearly eliminated when the hands-on scores of no-poverty, high-reading students are in a program with a science specialist.

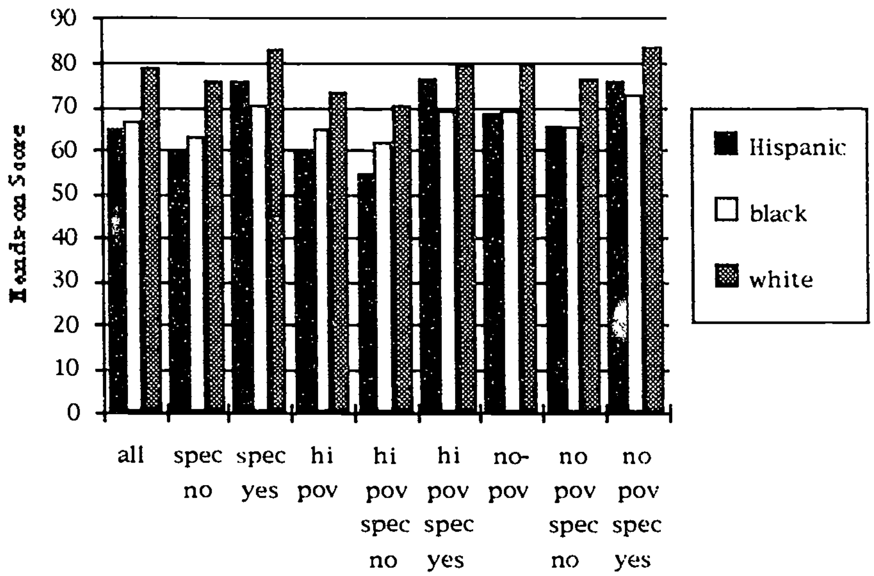

Figure 15. Effect of Poverty and Science Specialist: High- and No-Poverty Hispanic, Black and White Students With and Without a Science Specialist

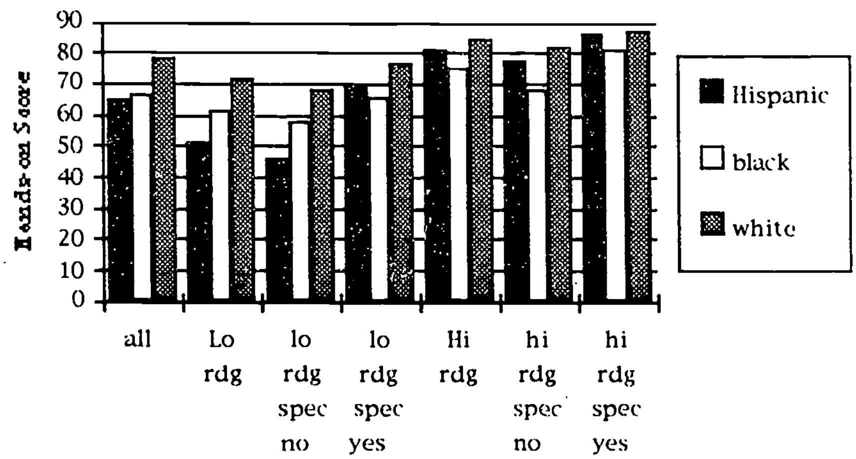

Figure 16. Effect of Reading and Science Specialist: Low- and High- Reading Hispanic, Black and White Students With and Without a Science Specialist

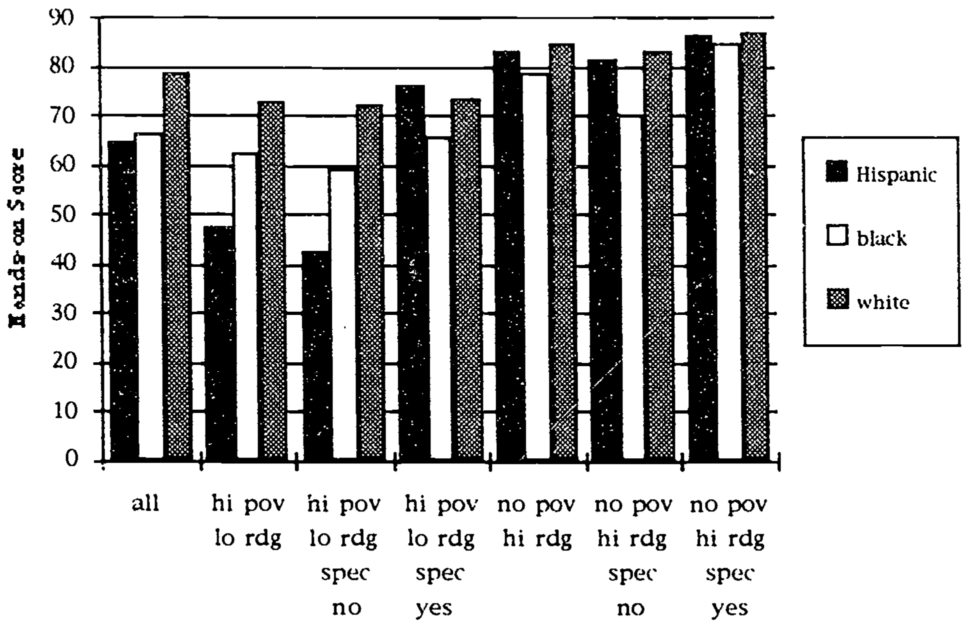

Figure 17. Effect of Poverty, Reading and Science Specialist on Hands- on Test Scores: High-Poverty Low Reading and No-Poverty High-Reading Hispanic, Black and White Students With and Without a Science Specialist

Table 13 Student Sample Available for Study

Student Subgroup	N	Student Subgroup	N
Total sample	1381	White females	
Males	677	High-poverty	36
Females	704	Low-poverty	18
Whites	759	No-poverty	322
Blacks	368	Black males	
Hispanics	226	High-poverty	87
Others	28	Low-poverty	13
high-poverty	388	No-poverty	74
Low-poverty	94	Black females	
No-poverty	899	High-poverty	114
White males	383	Low-poverty	17
White females	376	No-poverty	63
Black males	174	Hispanic males	
Black females	194	High-poverty	46
Hispanic males	103	Low-poverty	18
Hispanic females	123	No-poverty	39
Other males	16	Hispanic females	
Other females	12	High-poverty	62
White males		Low-poverty	11
High-poverty	41	No-poverty	50
Low-poverty	12		
No-poverty	330		
Other males		Other females	
High-poverty	1	High-poverty	0
Low-poverty	3	Low-poverty	2
No-poverty	12	No-poverty	10
Science specialist			
With	577		
Without	804		

Educational Importance.

Jeannie Oakes (1990) asked if it is race or economic status that has a greater effect on science and math achievement. Based upon the results of this study, it appears that for science the answer is economic status. Within each racial group, test scores were found to increase significantly from high-poverty to no-poverty levels. This study shows that when economically disadvantaged (high-poverty) students are tested in alternative ways, they are better able to demonstrate what they know and can do.

The results of this study also allow concerns about reading as expressed by Scott-Jones and Clark (1986) and Tolman, Sudweeks, Baird, & Tolman (1991) to be addressed. On the hands-on test, students with low reading levels were apparently less handicapped by their inability to read and therefore performed better on the hands-on performance test than they did on the multiple-choice test which relies more upon reading. Data obtained lead to the conclusion that reading has an even greater effect than poverty level on science achievement; and, as Scott-Jones and Clark suggest, a complex pattern of differential performance in reading was found to exist for males and females. For low-reading ability students, females scored higher than males, whereas for high-reading ability students, males were found to score higher than females.

In addition, the results of this study corroborate the work of Doran and Tamir (1992) regarding the mode of assessment of science process skills. Hands-on test scores of all students were found to be higher than their multiple-choice test scores of science process skills. The difference was greater for some subgroups of students than for others and because of this, the gap between certain subgroups of students was greatly reduced.

The results of this study provide additional evidence to support the studies of Beane (1985), Bredderman (1983), and Kyle, Shymansky, & Alport (1982) who found that economically and or educationally disadvantaged students in hands-on activity-based science programs performed better than those in textbook-based programs. From the results of this study, it can be concluded that students in science programs where the focus is teaching science through hands-on activity-based experiences (with a science specialist), performed significantly better than those in programs without a hands-on program (without a science specialist), especially those students of low-reading and high-poverty levels.

The results of this study also add strength to the conclusions drawn by Zuzovsky & Tamir (1989) and Staver & Walberg (1986) regarding the effects of alterable school variables on certain subjects such as science, especially in low-socio-economic schools. A science specialist is an alterable school variable. The science specialists in the schools in this study were not teachers with special science degrees or certification. They were elementary teachers who chose and were selected to teach science in their schools. They wanted to teach science. Therefore, it can be inferred that not only are we assured that science was taught regularly and frequently, but that they were probably enthusiastic about teaching it and, in turn, these teachers most probably conveyed this positive attitude about science to their students. The final result is that someone who wanted to teach science was accountable for teaching it and did so regularly. Hence, science was taught using a hands-on approach and when it was taught more science was apparently learned by all students.

Based upon the results of this study, it is clear that all students can learn science provided that two conditions are met: (1) students must be provided with appropriate instruction so that they have the opportunity to learn science (this can be assured by providing a science specialist); and (2) students must be able to demonstrate what they know and can do (this condition can be met by providing hands-on, performance-based assessment).

REFERENCES

Abell, S. K. (1990). A case for the elementary science specialist. School Science and Mathematics, 90(4), 291-301.

Beane, D. B. (1985). Mathematics and Science: Critical Filters for the Future of Minority Students. Washington, DC: The American University Mid Atlantic Center for Race Equity.

Bredderman, T. (1982). What research says: Activity science - the evidence says it matters. Science and Children, 20(1), 39-41.

__________ (1983). Effects of activity-based elementary science on student outcomes: A quantitative synthesis. Review of Educational Research, 53(4), 499-518.

Champagne, A. B. (1990). Assessment and Teaching of Thinking Skills. In G. Hein (Ed.), The Assessment of Hands-on Elementary Science Programs. Grand Forks, North Dakota: University of North Dakota Center for Teaching and Learning.

Cizek, G.J. (1991). Innovation or enervation? Performance assessment in perspective. Phi Delta Kappan, 72(9), 695-699.

Comber, L. C., & Keeves, J. P. (1973). Science Education in Nineteen Countries. New York: Wiley and Sons.

Davis, A. & Armstrong, J. (1991). State Initiatives in Assessing Science Education. In G. Kulm & S.M. Malcum (Eds.). Science Assessment in the Service of Reform. Washington, DC: American Association for the Advancement of Science.

Doran, R. L. (1990). What research says about assessment. Science and Children, 27(8), 26-27.

__________ & Tamir, P. (Eds.). (1992) An international assessment of science practical skills. A collection of articles which will appear in Studies in Educational Evaluation, Vol. 18 Issue I.

Guilford, J. P. (1956). Fundamental Statistics in Psychology and Education (p. 145). New York: McGraw Hill.

Hein, G. (1987). The right test for hands-on learning. Science and Children, 25(2), 8-12.

__________ (Ed.) (1990). The Assessment of Hands-on Elementary Science Programs. Grand Forks, North Dakota: University of North Dakota Center for Teaching and Learning.

Hernstein, R. J. & Murray, C. (1994). The Bell Curve: Intelligence and Class Structure in American Life. Free Press.

Hounshell, P. B., & Swartz, C. E. (1987). Elementary science specialists? Definitely!/We know better......Science and Children, 24(4), 20-21, 157.

Jones, L. V. (1984). White-black achievment differences: The narrowing gap. American Psychlogist, 39, 1207-13.

Kanis, I. B. (1988). An Analysis of the Science Process Practical Examination Administered to Grade Five and Grade Nine Students in the United States of Am rica. Doctoral Dissertation, Teachers College, Columbia University, New York.

Kean, M. (1991). In Test Publishers Defend Standard Exams, Warn Against Alternatives. Report on Educational Research, 23(21), 5-6.

Kean, T. H. & Neill, M. (1991). Do we need a national achievement exam? Education Week, X(31), 36.

Kellogg, T. M. (1987). Science and State Assessment Programs. Science and Children. 24(7), 23-29.

Kuechle, J. A. (1990). The Effects of Written Planning on 6th Grade Boys' and Girls' Ability to Control Variables During Practical Assessment (Sixth Grade). Doctoral Dissertation. University of Minnesota.

Kulm,G. & Malcom, S. M.,(Eds.). (1991). Science Assessment in the Service of Reform. Washington, DC: American Association for the Advancement of Science.

_________, & Steussy, C. (1990). Assessment in science and mathematics education reform. In G.Kulm & S. M. Malcom, (Eds.), Science Assessment in the Service of Reform. Washington, DC: American Association for the Advancement of Science.

Kyle, W. C., Jr., Shymansky, J. A., & Alport, J. M. (1982). Alphabet soup science: A second look at the NSF-funded science curricula. The Science Teacher, 49(8), 49-53.

Maeroff, G. I. (1991). Assessing alternative assessment. Phi Delta Kappan, 73(4), 272-281.

Marshall, J. E. (1991, April 7-10). Construct validity of multiple-choice and performance-based assessments of basic science process skills: A multitrait-multimethod anal Paper presented at the Annual Meeting of the National Association for Research in Science Teaching. Lake Geneva, WI. Center for Open Education.

Meng, E. & Doran, R. L. (1990). What research says about appropriate methods of assessment. Science and Children, 28(1), 42-45.

Mitchell, R. (1992a). Testing for Learning: How New Approaches to Evaluation Can Improve American Schools. New York: The Free Press, A Division of MacMillan, Inc.

New York State Education Department (1985). Elementary Science Syllabus. Albany: The University of the State of New York.

Oakes, J. (1990). Multiplying Inequalities: The Effects of Race, Social Class, and Tracking on Opportunities to Learn Mathematics and Science. Santa Monica, CA: The RAND Corporation.

Petraitis, B. J. (1991, October 4-5). The future of testing and assessment: The College Board's perspective. Presentation at the New York State Council of Educational Associations 17th Annual Leadership Conference. Albany, New York.

Pine, J. (1990). Validity of science assessments. In G. Hein (Ed.), The Assessment of Hands-on Elementary Science Programs. Grand Forks, North Dakota: University of North Dakota Center For Teaching and Learning.

Sattler, J. M. (1988). Assessment of Children. San Diego, CA: Jerome Sattler, Publishers.

Saturnelli, A. M. (1993). Alternative Assessment in Elementary Science: What Difference Does It Make? Doctoral Dissertation, New York University, New York.

Scott-Jones, D., & Clark, M. L. (1986). The school experiences of black girls: The interaction of gender, race, and socioeconomic status. Phi Delta Kappan, 67(7), 520-526.

Shavelson, R. J., & Baxter, G. P. (1992). What we've learned about assessing hands-on science. Educational Leadership, 49(8), 20-25.

__________ & Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. Educational Researcher, 21(4), 22-27.

__________ & Pine, J. (1991). Performance assessment in science. Applied Measurement in Education, 4(4), 347-362.

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

Shymansky, J., Kyle, W., & Alport, J. (1982). How effective were the hands-on science programs of yesterday? Science and Children, 20(3), 14-15.

Staver, J. R., & Walberg, H. J. (1986). An analysis of factors that affect public and private school science achievement. Journal of Research in Science Teaching, 23, 91-112.

Tamir, P. (1989). Home and school effects on science achievement of high school students in Israel. Journal of Educational Research, 83(1), 30-39.

Tolman, Marvin N., Sudweeks, R. Baird, H., & Tolman, R. (1991). What research says: Does reading ability affect science test scores? Science and Children, 29(1), 44-47.

Wadsworth, B. J. (1984). Piaget's Theory of Cognitive and Affective Development. NY: Longman.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. Phi Delta Kappan, 70(9), 703-713.

__________ (1990). Alternative Forms of Assessment. A workshop presented by Mr. Wiggins at Holiday Inn, Newburgh, New York, November 7, 1990.

Williams, D. H. (1990). Making a case for the science specialist. Science and Children, 27(4), 30-32.

Zuzovsky, R., & Tamir, P. (1989). Home and school contributions to science achievment in elementary schools in Israel. Journal of Research in Science Teaching, 26(8), 703-714.