DOCUMENT RESUME

ED 282 838 SP 028 842

AUTHOR Martinello, Marian L.; And Others
TITLE A School-Museum-University Program to Develop Hispanic Children's Scientific Observation and Inferential Thinking.
PUB DATE [86]
NOTE 41p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cognitive Development; *College School Cooperation; Higher Education; Hispanic Americans; Intermediate Grades; *Museums; Preservice Teacher Education; Questioning Techniques; *Science Instruction; Student Teachers

ABSTRACT

This document describes a school-museum-university collaborative model for the effective use of museum resources to develop scientific observation and thinking skills in minority children. Participants in the program were 268 Hispanic American fourth-graders, most of whom were from low socioeconomic backgrounds. Over a 12-week period the students, accompanied by student teachers, toured a museum science exhibit, had four classroom lessons related to the visit administered once a week, and in the 10th or 11th week revisited the museum exhibit. A pretest and tests given after each museum visit assessed children's thinking, observation, and inference abilities; the study also examined the project's influence on the student teachers' questioning skills. In order to test the influence of the time factor, a control group of 57 students of comparable background was given a 2-week intensive experience of the same program. Although great variability was found among test items, the control group's apparent growth in thinking skills argues that an intensive timeframe may be critical for science process learning. Overall, it was believed that the program and lessons had significantly improved students' skills and aided the preservice teachers in developing questioning skills to help guide the students' learning. (CB)

A SCHOOL-MUSEUM-UNIVERSITY PROGRAM TO DEVELOP HISPANIC CHILDREN'S SCIENTIFIC OBSERVATION AND INFERENTIAL THINKING

Marian L. Martinello
The University of Texas at San Antonio

in collaboration with:

Mary B. Kennedy
Megan Kromer
The University of Texas at San Antonio

Sarah Hepner
San Antonio Museum Association

Angelo Russo

Harlandale Independent School District
San Antonio, Texas

A SCHOOL-MUSEUM-UNIVERSITY PROGRAM TO DEVELOP HISPANIC CHILDREN'S SCIENTIFIC OBSERVATION AND INFERENTIAL THINKING

This research was funded by the Education for Economic Security Act, Title II Program for Mathematics, Science, and Critical Foreign Languages, Higher Education Grants Program, Coordinating Board, Texas College And University System, 1986

There is a critical difference between looking and seeing. The widely held assumption that all people with normal or corrected vision see equally well is in error. Most people's powers of observation are limited at best. Evidence for that fact is common and familiar: people regularly complain about forgetting details (books to improve memory are popular) and reliable eye witnesses are hard to find, whether in courtroom testimony or news reports). The related facts are: first, we can remember only as much as we have seen; if we do not recall details, it is probably because we have looked at them but have not seen them. And second, we can <u>see</u> with detail and precision only when we know how to differentiate observation from inference. These are life skills. They are also basic to scientific inquiry.

The process of observing is the keystone of scientific thinking. Indeed, the Texas Chapter 75 curriculum cites observing, defined as acquiring data through the senses, as the first essential element in science education at each of the elementary grade levels, K-6. The essential elements speak to the development of skills of drawing logical inferences in grades 2 through 6.

Learning to observe in detail and with precision and knowing how to differentiate between one's collection of data (observations) and one's interpretations of that data (inferences) requires specific and sustained instruction in these processes. Children cannot learn to observe, to infer, and to differentiate between observations and inferences unless they practice using those skills. To do so, they must have opportunities to observe real objects and phenomena and to interpret them. They need teachers who know how to guide their development of those skills and they need an environment that invites their practice to those skills. One of the best places in which children can learn to observe and infer--to read objects--is the museum.

Although a variety of programs are reported in the literature on museum education, few have evaluative components that attempt to relate museum experiences to children's growth in thinking skills or the teacher growth in abilities to develop children's scientific observation and inferential thinking. Also limited are tests of science process skills; indeed, none are available to assess the specific skills of observing and inferring. The reported study addresses the needs to better develop and assess museum education, curriculum and teaching for minority children's growth in scientific thinking.

PROBLEM

How can a museum exhibit be used to develop Hispanic children's scientific observation and inferential thinking? This inquiry applied a repeated measures research design to a school-museum-university collaborative program. It examined the separate and collective contributions of guided interactive inquiry tours of a science exhibit and follow-up classroom learning activities to children's growth in: (1) making detailed observations, (2) formulating valid inferences, (3) identifying supporting evidence for inferences, and (4) differentiating observations from inferences.

In addition to examining children's growth in selected thinking skills, the study inquired into the influence of project activities on the development of preservice teachers' questioning skills for guiding children's observing and inferring.

METHOD

The experimental treatment was sequenced as follows:

Week	Activity
1	pretest
2 or 3	museum tour 1
4	posttest 1
5	classroom lessons
6	classroom lessons
7	clasroom lessons
8	classroom lessons
9	posttest 2
10 or 11	museum tour 2
12	posttest 3

The control group was tested on the same instruments at the same times as the experimental group. To assess the value of a intensive two-week experimental treatment, as compared to the twelve-week program, the control group experienced a museum tour and four classroom lessons after posttest 3 had been administered to all groups. Following the treatment, the control group took a fourth posttest.

At the start and end of the project, participating teacher education students were tested on their skills of writing question sequences to guide children to scientific observation and inferential thinking.

Samples

The experimental group numbered 268 children from eleven fourth grade clases in three schools of a school district in San Antonio, Texas. Ninety-five percent of the enrollment in each of two of these schools is Hispanic and of lower socio-economic status (SES). Seven classes in the experimental group, totaling 162 children, were from those schools. The third school in the experimental group has an 85% Hispanic enrollment of lower middle SES. Its four fourth grade classes, 106 children, participated in the experimental group. The control group was comprised of 57 children in two classes from a school in the same district with enrollments of 95 percent lower SES Hispanic children.

The classes were selected for the study by the district science curriculum coordinator who asked for the participation of fourth grade teachers with interest in science education in schools with enrollments that are representative of the districts' minority population. All fourth grade teachers in each of four schools agreed to participate.

Forty-two preservice teacher education students, enrolled in a required undergraduate course in science education in the elementary school, comprised the teacher sample. The majority of the students were Anglo; 8 were of Hispanic background. All were involved in the study because they enrolled in the science education course during the spring 1986 semester. A control group of teacher education students was not established.

Sites

In addition to the classroom settings of each participating fourth grade class, the study included a museum exhibit. Texas Wild is a permanent exhibit on Texas ecology at the natural history and history museum of the San Antonio

Museum Association. The exhibit includes encased dioramas and visual panel presentations of plant and animal interrelationships within the diverse ecological regions of the state and a walk-through diorama of the Texas thorn-brush. It contains other features as well, but the dioramas figured prominently on the children's tours because they contain a wealth of visual resources for teacher-guided observing and inferring. In addition, they are illustrative of installations in most natural history museums.

Teaching Teams

The teacher education students were grouped in teams of four by self-selection. Each team was assigned to one of the experimental classes. Team members shared responsibilities for conducting the tours, testing the children and teaching the classroom lessons. Another student, who had completed the science education course and who was enrolled in an independent study, conducted all project activities with the control group.

The Tours

A tour plan was scripted for use with the Texas Wild exhibit; it uses question sequences to guide children's analytic observation. The intent was to develop viewers' visual literacy by helping them learn how to look at an exhibit and how to interpret what they see. Minority children are reputed to perceive in wholistic ways; their attention span is often characterized as immature, i.e., of short duration. Therefore, the tour plan was designed to promote the children's attention to detail and their sustained examination of each display included on the tour. Another problem cited by museum educators is a tendency among exhibit viewers to make incorrect inferences about the content of exhibits because they lack preceptiveness and relevant background information. The project's tour plan included interactive episodes that led the children to make inferences based on what they saw, knew, and where told.

To increase the comparability of experiences with the Texas Wild exhibit for the tour script included expected answers for each question. A typical question sequence is illustrated by one addressed to a display case on the concept of <u>niche</u> in which several birds appear, especially the woodpecker, bobwhite, and mockingbird:

<u>Information</u>
Habitat adaptations enable different plants and animals to do different jobs in their "neighborhoods." The job that a living thing does and the place where it does its job is called its <u>niche</u>. Look at the tree birds in this exhibit and discover adaptations: specialized body parts that help the animals do their jobs in the place where they live.

<u>Notice that the woodpecker has a strong, pointed beak</u>.

<u>How does the woodpecker use its beak</u>?

(Ans: Making a hole for nesting; pecking insects out of the wood of trees).

<u>Do his beak feathers blend well with the tree bark</u>?

(Ans: yes)

Look carefully at the woodpecker's feet. <u>Describe how the toes are positioned and how they are used</u>.

(Ans: Two toes in front; two in back — used for climbing and for holding onto the tree.)

Now, compare the woodpecker's feet to the feet of the mockingbird. <u>Describe the woodpecker's toes and how they are used</u>.

(Ans: Three in front; one in back for perching on a limb.)

Notice the feet of the bobwhite. <u>How are they different from the mockingbird's?</u>

(Ans: Stronger for scratching in the soil and walking on the ground.)

<u>Is the bobwhite well camouflaged?</u>

(Ans: yes)

Look at the beaks of these three birds. <u>How are they different in looks and how do the birds use them in different ways?</u>

Ans:

Bird	Beak	Use
woodpecker	long, strong	drilling in wood
bobwhite	shorter, blunt	seed eating
mockingbird	thinner, pointed	catching insects, plucking berries; to meet wide variety of dietary needs.

The characteristics of other animals and plants on display in the Texas Wild exhibit were examined in similar ways on the tour. Children were encouraged to observe and make inferences about exhibited animal shelters, preditor-prey relationships, reproduction and rearing of young, protection and a variety of adaptations to their environments. Plants were explored in similar detail.

The tour was modeled for the preservice teachers who were given the script. Although they were told to guide the children through all sections of Texas Wild exhibit included in the tour plan, they were also encouraged to digress from the script in response to the children's interests and questions. The only stipulation was that, as docents, the preservice teachers focus their tours on developing the children's skills of observing and inferring.

The children were taken on the Texas Wild tour in groups of ten to twelve, each with one preservice teacher as its docent. With six definable areas (thornbrush, desert, Edwards Plateau, Plains, Gulf Coast, and Piney Woods), the exhibit was easily toured simultaneously by several groups, following alternate routes. All visits were scheduled to avoid conflict with other school groups. The tour was an hour in duration.

During the second or third weeks of the project, each class in the experimental group toured the exhibit in subgroups of 10-12 children, each guided by one of two members of the teaching team assigned to the class. Every experimental class returned to the Texas Wild exhibit during the tenth or eleventh week of the project for their second tours, conducted in subgroups as before, by the remaining two team members.

The control group classes experienced the same tour after the last posttest had been administered.

The Tests

Children's Thinking

One dilemma in assessing children's growth in thinking skills is the absence of paper-and-pencil instruments for the purpose. Our special problem was to create several forms of a test that can assess children's abilities to make precise and detailed observations, to formulate inferences from observed data and given information, to identify supporting evidence for given inferences, and to differentiate between observations and inferences. We needed paper-and-pencil instruments to collect data from whole classes of children at one sitting by one test administrator — a member of the preservice teacher team assigned to each class.

Four tests were developed, each with four subsections:

Observation. Each item in this section presents a line drawing which is repeated five times below the sample. Four of the five reproductions have minor modifications. The child is told to mark the one that matches the sample, i.e., the one that is not altered. These items require attention to specific and minute details, as indicated in example of Figure 1.

Place Figure 1 Here

Inference. Items in this section visually present a situation with information necessary to make the requested inference. Several inferences are offered from which the child is asked to select the most appropriate as in the illustration in Figure 2. Here the child must infer what the owl ate from the skeletal remains extracted from the pellet the owl regurgitated.

Place Figure 2 Here

Supporting evidence. Figure 3 gives an illustration of items that ask the student to identify the best pieces of evidence, from those listed, that support a given inference. Figure 4 shows the open jaw of a snake's skull and asks which evidence suggests why a live mouse cannot wiggle free of the snake's grip.

Place Figure 3 Here

Differentiating observations from inferences. The first section of the test supplies a visual with enough explanation to help the child interpret the picture. Several statements are given; some are observations and some inferences. The child is asked to differentiate between the two types of statements. Figure 4 illustrates this type of item with drawings of a lizard's eye in varying degrees of light. The child must first mark the inferences in the mixed set of observations and inferences, then identify the observations.

Place Figure 4 Here

The method of administering each test was explained, with modeling, to the preservice teachers who comprised the teaching team. One member of the team was responsible for each of the tests (all control group tests were administered by one student). Arrangements were made by the teaching teams to administer tests to the children at times convenient to the class and classroom teacher during the weeks designated for testing on the project schedule. That schedule was strictly maintained throughout the project. Test administration was highly structured. The test administrator read each item to the class to offset differences in reading ability within the group. The children marked their responses before the class moved to the next item. The administrators set the pace, based on their assessments of the children's rhythm. When children asked for more time to ponder an item, they were permitted to return to it after all test pages were completed. Efforts were made to provide sufficient time for test completion. In most cases, all children completed the test within an hour.

Test items were scored as either right or wrong if only one answer was requested and required to correctly respond to the question, as indicated in Figures 1 and 2. Those items requesting more than one answer were scored for the total number of correctly marked and unmarked responses as indicated in Figures 3 and 4.

Teacher Questioning

The tests used to assess the preservice teachers' growth in questioning skills required each student to write a question sequence to guide children's observing and inferring about science content. The pretest asked the student to observe a live oak tree, identify an inference that could be made about the tree, and then write a series of questions, with anticipated responses, to guide children to make that inference, based on supporting observations and knowledge. Two posttests were administered; each required the preservice teachers to write question sequences with reference to different topics. Posttest 1 examined life along a 30 ft transept drawn on level ground. Posttest 2 focused on the iodine test for starch in a geranium leaf that has been stripped of its chlorophyll by an alcohol bath. As for the pretest, the student was asked to define an inference that could be drawn from the given situation and write a question sequence to elicit the inference from children. Posttest 1 was given under conditions similar to the pretest--a class exercise. Posttest 2 was part of the final exam for the science education course in which the students were enrolled.

The written questions were classified by their intent to elicit the following from children: Observation-general, observation-specific, observation-compare/contrast, inference, and knowledge. (A category of other was used for ambiguous questions.) Figures 5 and 6 present the classification system and rating scale. Three independent judges coded the questions by type. Table 1 shows the interrater agreements to be significant for all categories. One exception is the observation-specific category for which there was good agreement between raters 1 and 2 but not between 2 and 3. Rater 3 had only an hour's training in use of the system; indeed, the system was easily used without extensive preparation of raters, suggesting its ready applicability to teacher and docent training. The second exception is the category "other" for which there were too few examples to determine correlations among ratings.

Table 1

The preservice teachers' questioning sequences were coded to establish an objective basis for rating them. Each question sequence was rated on a five-point scale according to its apparent effectiveness in helping children make precise, detailed observations of the phenomenon under examination and to draw upon relevant information in order to support their formulation of valid inferences. A rating of 5 was assigned to the very best sequences, those that start with questions calling for observation, followed by requests for additional specific observations and information that support a logical line of reasoning toward the inference that is solicited by the last question in the sequence. The scale differentiates sequences rated 4 or 3 from those rated 5 by the number and position of questions calling for observation in the sequence and the logical consistency of the sequence. Those rated 2 have few questions calling for observations relevant to the desired inference and

Two raters applied the scale to a preservice teachers' questioning sequences; the coefficient of correlation was 0.895 and significant at the .001 level.

The Lessons

Four lessons were planned to follow the children's museum visit. Each was designed to develop their observing and inferring skills on the behavior of mealworms. The second explored animal tracks. The third was on fingerprints. Those three lessons contained structured observation tasks to lead the children toward specific inferences for which they had supporting evidence. The fourth lesson was adapted from activities with candles in Science--A Process Approach. This digression from nature study was made to focus sharply on the differentiation of observation from inference. All lessons were developed in detailed written form, with teacher questions, children's tasks and children's "lab" sheets. Each lesson was taught to the preservice teachers by having them engage in its learning activities as learners. To promote their ownership of the lessons, we asked them to discuss and to revise the written lesson plans during their college class sessions, after having read the plan and experienced the activities. Their suggestions were incorporated into the final version of the lessons. The teaching teams were supplied with all materials in class quantities needed for teaching the lessons to the experimental classes. One team member was responsible for teaching each lesson. In practice, several team members collaborated in the teaching activities, while each served as the principal teacher for one of the four lessons.

FINDINGS

The 12-Week Experience

A Between-Within Design was used to determine the significance of changes in scores from pretest to posttest, where the between factor was the group (experimental vs. control) and the within factor was time. Each of four scales was analyzed separately: (1) observing, (2) inferring, (3) finding supporting evidence for given inferences, and (4) differentiating between observations and inferences. F scores for the between factor were not significant for any of the four scales at any testing time during the twelve-week program, as shown in Table 2. The within factor was significant for scales 1 and 2, but not for scales 3 and 4. The interaction of the two factors was significant for only the inference scale, but not in the positive direction of growth in ability to infer.

A class-by-class analysis showed variable performance on each of the subscales across experimental and control classes. Most classes did less well on the observation subscale of the posttests as compared with the pretest. All scored well below their pretest performance on posttests 1 and 2 but on posttest 3, the experimental group classes and one control group class came within ten to twelve points of their pretest means, a decided improvement over their performances on the first two posttests. Only one experimental class showed consistent decline in scores on this scale over all four tests. For the subscale on inferring, performance was even more variable; most classes showed a decline in scores on the first two posttests. Seven experimental and both control classes had mean scores at least ten points higher on posttest 3 than on posttest 2. Two experimental classes actually exceeded their pretest means on the inferring scale by at least ten points and one by 5.6 points. The patterns of performance on the scale measuring ability to identify supporting evidence for inferences in scores were different for each class. No class showed improvement but differences in scores were not as sharp as they were for the other scales from test to test. And on the last scale, testing ability to differentiate between observation and inference, the majority of classes had remarkably stable means across test forms.

On first examination of these test scores, the lack of improvement expected for the experimental group over time and in comparison with the control classes and the variability in scores on subscales from class to class raised questions about the test items. Several test administrators reported that: (1) the children needed more time to complete the tests than had been planned, (2) the children appeared to become "test-tired" after the second test was administered, and (3) some children appeared to mark answers at random. Test administrators also reported that the second posttest in the series seemed more difficult for the children than had the pretest and first posttest. Some difficulties were reported in the children's interpretation of test illustrations. Many illustrations were used when constructing the test items to insure that the children were asked to make direct observations and to draw inferences from those observations rather than having to rely on their experimental backgrounds to determine answers. Unfortunately, however, the duplicating processes rendered some visuals unclear and several ambiguous. Another variable affecting the children's performance on the tests was the inexperience of preservice teachers who administered them. In fact, variability among classes could also be attributed to the differences in teaching skill among the preservice students who conducted the tours and who taught the lessons. But variability in the children's performance was evident between control classes and both those classes had the same test administrator. The test seemed to hold promise but also require revision toward greater standardization of items, administration, and scoring.

The 2-Week Intensive Experience

Several of the teachers of classes in the sample had commented that their students learn best under conditions of constant review and reinforcement. This raised the question of whether the experimental treatment was diluted by its "once-a-week" character. Perhaps the time gaps among the tours, tests, and lessons did not accommodate learning styles that require sustained attention to desired learnings and immersion in the instructional activities designed to develop them. If the children needed intensive learning experiences for content and basic skills learnings, the time factor might be especially critical in developing their thinking skills. The control group offered an opportunity to make an initial, if tentative, inquiry into the question. In the two weeks immediately following the administration of posttest 3 to the control group classes, the most experienced project staff led the children through an inquiry tour of the Texas Wild exhibit and conducted the four associated lessons in their classrooms. The form of the test used as pretest some fourteen weeks earlier was then administered to the control classes as their final test. Although the dependent t-test comparing control group means on posttest 3 and the final test must be interpreted cautiously, all were significant in the desired direction. Table 3 presents the t scores and significance levels for each scale. Means rose from 47.81 to 66.92 for observation (scale 1), from 66.34 to 76.30 for inference (scale 2), from 61.22 to 67.27 for identifying supporting evidence for inferences (scale 3), and from 56.53 to 66.38 for differentiating observations from inferences (scale 4). One might argue that these apparent growth scores are really a product of differences in the difficulty level of the two forms of the tests. However, considering that the children took the final test on the last day of school, that some classes in the sample showed strong performance on posttest 3, and that experimental group scores on scales 3 and 4 did not vary over time, we are justified in paying some attention to these findings. If the test forms are not exactly parallel on scales 1 and 2, they seem to be on scales 3 and 4. The bulk of the evidence suggests that a 2-week intensive learning experience that provides for review and reinforcement of scientific observation and inferential thinking may be superior to instruction over a more extended timeframe, especially for the development of thinking skills in children of lower SES and minority backgrounds.

To correct for the degree to which the control group's growth scores were an artifact of the test forms, the children's performance on each scale of the pretest was compared with their performance on the same form of the test after they had experienced the experimental treatment. Dependent t-tests found those scores not to be significantly different except on the subscale measuring ability to differentiate observations from inferences. On that scale the pretest mean of 57.58 was significantly exceeded by the final test mean of 66.47. The t score of 3.21 was significant at the .002 level (df=48). No significant change on that scale was discernable for the control group's performance on the pretest and posttests 1, 2, and 3 prior to their exposure to the experimental treatment. This may be interpreted as supporting evidence for the superiority of the intensive over the extended instructional experience for developing children's discrimination of data perceived and interpreted.

The Performance of Boys and Girls

Girls are considered a special minority in science education. Here the Hispanic girls in this sample outscored by the boys on tests of scientific observation and inferential thinking? Scores made by the girls and boys on each scale of each test were compared if scores for all between-group comparisons on each scale were not significant. It appears that the girls in our sample were able to observe, infer, find supporting evidence for inferences, and differentiate observations from inferences as well as their male cohorts. That finding supports the assumption that gender is not a critical factor in minority children's development of science process skills. The question deserves further study.

Preservice Teacher Growth in Questioning Skills

Good interrater reliability had been determined for the question categories and rating scales applied to the question sequences written by preservice teachers to develop children's scientific observation and inferential thinking. Two judges then applied the categories and scale to achieve consensus in the classifications and ratings for pretest and posttest question sequences. Numerial ratings were assigned to each sequence after all questions in the sequence had been coded; ratings were based on the types of questions and their function in the sequence.

The thirty-one students who completed both the pretest and posttest 1 had a mean rating of 2.9 on the pretest and 3.4 on posttest 1. Table 4 shows the 2-tailed t score of -1.87 to have probability level of .07. That table also presents the difference between the pretest mean and the mean of posttest 2 for the 38 students who had taken both tests to be less suggestive of improvement; the t score -1.57 had a probability level of 0.126.

The pretest and posttest 1 were comparable exercises in that the focus of each (Live Oak tree and life along on 30-foot transept) permitted many inferences that could be supported by direct observations. They were also completed in the context of a class learning activity. By contrast, the focus of posttest 2 on the iodine test for starch in leaves required more background information than direct observation to arrive at a particular inference. The topic had fewer instruction options. Therefore fewer variations were permissible in question sequences that were rated high on our scale. In addition, the completion of posttest 2 in a testing situation probably increased the anxiety level of the students. They had less time to reflect on their question-writing and, while questioning during teaching performance must be spontaneous, the subjects were still novices who need time to think. This experience suggests that the valuation of questioning skills by preservice students must compare performances that are comparable in degree of stress they generate as well as in the characteristics of the context focus for question writing. The pretest-posttest 1 comparison appears to be the better measure of student growth in questioning skills. While not dramatic, these findings argue for growth in the preservice teachers' development of questioning skills that can guide children's scientific observation and inferential thinking. Their participation in project activities does seem to have aided them in developing these important pedagogical skills.

DISCUSSION

School-Museum-University Model

This project's guiding idea was to develop a collaborative model of effective use of museum resources for the science education of minority children--children who are least likely to visit museums on a regular basis. The purpose was to promote children's interaction with exhibited science content so that they might become more observant viewers and more rational interpreters of what they see. We also sought to explore a fresh dimension in field work for prospective teachers by involving them in museum as well as classroom instruction. The curriculum that was developed to accomplish those purposes, with its interactive tour plan for museum teaching and its series of lessons for classroom follow-up, appears to be effective. The quantitative findings are encouraging. Qualitative assessments of classroom teachers, museum educators, preservice teachers, and the children themselves sparkle with evidence of excellent achievement. Each facet of the program is discussed here with recommendations for further study.

Museum-classroom curriculum

The tour plan and follow-up classroom lessons for the Texas Wild exhibit were scripted and modeled for the preservice teachers as opposed to having them develop their own tour and lesson components. The students' response was stronger than expected: they greatly appreciated the completeness of the models provided--in print and in performance. All student team reports commented on the value of the structured materials in helping them teach with confidence in both museum and classroom settings. The students reported that they felt quite comfortable in developing their own styles of performing from the tour and lesson scripts which supplied the needed foundation that they "knew would work" in contrast to the more risky, less well designed, instructional plans they might prepare. The classroom teachers who observed the novices teach concurred. During the debriefing conferences held with the teachers after the last test was administered, all commented on the unusually high quality of the students' teaching. The classroom teachers were surprised and pleased that the children in their classes had responded so well to the preservice teachers' instruction and, in fact, waited with anticipation for "the science lesson." That curriculum package is complete enough for any teacher or museum educator to use in conjunction with the Texas Wild exhibit. An interesting question that deserves longitudinal study is whether and to what degree the preservice teachers who participated in this study will use the curriculum package for the permanent Texas Wild exhibit in their own teaching.

Children's process learnings

None of the quantitative analyses of the experimental group's test performances over time demonstrate growth in the children's abilities to observe, infer, find supporting evidence for inferences, or differentiate observations from inferences. None of the analyses suggest that the second visit to the museum served a significant instructional purpose. Nonetheless, the preservice teachers' team reports are full of testimony to the children's process learnings. Several teams reported improved perceptiveness on the part of their classes during the second tour of the Texas Wild exhibit. One team said it well for all who reported the children's independence in interpreting the exhibit on the second trip: "They didn't need us!" That team report continues: "It seems that they (the children) saw and observed more in the second tour. The students were able to make some good inferences during the first tour, but they seemed to make even better and more outstanding inferences during the second tour."

Some teams expressed the view that the children had developed a background of knowledge during the first tour and subsequent lessons that enabled them to literally see more during the second visit. When the children had difficulty making expected inferences, the teams' reports implicated experiential deficiencies. For instance, the display of buffalo with thick coats on dry grass was expected to elicit the inference that the season depicted is winter. The team report states: "In spite of the buffalo's thick coat, the students immediately inferred that it was summer. They supported this inference with the observation that the grass was dead and dry. When one considers that most of the students are probably accustomed to the dry grasses which characterize South Texas summers, this was good evidence to support their (italics ours) inference." This is also symptomatic of the children's tendencies to base inferences on limited data, i.e., jumping to conclusions instead of reserving judgment.

Some teams reported evidence of the children's learning how to observe with precision during each tour; they noted that experiences in observing details in the displays first viewed appeared to prepare the children to better observe details in subsequently viewed displays. Several teams also found that the children were especially attentive to details on animals they had not seen in close-up before. Cited as an example was one group's interest in the details of the body structures of birds. Those who took the children to other exhibits in the Witte Museum, notably Dinosaures: Vanished Texas found that the children were able to formulate valid inferences and give supporting observations with relative ease. One of the preservice teachers commented that she could see growth in the children's perceptiveness as the project progressed—they demonstrated observational and inferential skill that might not be documented by the paper-and-pencil tests.

Several classroom teachers reported evidence of their children's attention to detail and use of descriptive language at times and in contexts that were unrelated to project activities. One teacher attributed to project learnings a clear increase in her students' use of descriptive language in writing assignments. After experiencing project activities, she noticed that the same children who typically gave limited, two-sentence responses to a writing assignment were submitting paragraphs of several sentences in length, containing detailed observations. Another teacher noted that the children responded to art works with greater awareness of detail than had been evidenced before their project participation. These accounts, while based on impressions rather than precise analyses, were prompted by what appeared to be dramatic changes in the children's behavior. They suggest new questions for study of the impact of a museum-classroom science curriculum on children's use of descriptive written language and on their analyses of pictures, i.e., the development of their verbal and visual literacy.

Tests of science process skills

The literature contains relatively few references to the development of tests to assess children's use of science processes. Work done in that area seems to have been limited to the 1970s, when Science - A Process Approach was developed. That program's tests have been difficult to find with reliability measures. The tests developed to date for assessing skills of science processes in elementary school children contain relatively few items on observing and inferring (Beard, 1971; Molitor, 1971; Tannebaum, 1968). It was necessary to develop test items for this project and to use them without benefit of trial test. Indeed, this project was the pilot for item trials. The variable test performance of experimental and control groups suggests that the tests were not of comparable difficulty. Problems were encountered with the clarity of pictures in some cases. A few items were found to be ambiguous or unclear to the children. Sections of some tests were too long, tiring the children and causing some to mark answers in apparently random ways. Nonetheless, many items are clear and useful measures of children's scientific observation and inferential thinking. Others have good potential and, with editing, can contribute to test development for these processes. Sustained effort in this area is critical and inportant if the goals of science education, like those included in the essential elements legislated for elementary education by the State of Texas, emphasize science thinking skills. This study has developed an item pool from which tests can be developed with better visual quality, shorter administration time, and increased comparability across test forms. The question of test validity can be addressed by developing performance tasks that record children's orally reported observations and inferences, the latter with supporting evidence. When these data are compared with the same children's performances on the paper-and-pencil tests, item validity and comparability can be determined.

Intensive format curriculum

Even given the variable quality of test items and the difficulties encountered with test administration, the study's findings suggest that the Hispanic children of lower SES comprising the study's sample, profited from an intensive experience with the museum - classroom curriculum on Texas ecology and scientific observation and inferential thinking. The children's teachers supported the reasonableness of that interpretation of pre-posttest comparison, for the twelve week and the two week experiences with the program. They characterized their students as "losing track" of lesson sequences or "forgetting" material when daily reinforcement is not provided. Because this study asked the classroom teachers not to teach the project's program but only to reserve time for preservice teachers to do the teaching, the weekly separation of project learning activities may have adversely affected the development of science process skills in the experimental subjects. The control group's apparent growth in those skills, after experiencing the experimental treatment over two weeks, argues that an intenstive timeframe for science process learnings may be a critical factor for instructional planning. This deserves further study. Two-week experiences with museum tour and classroom lessons should be compared with the same program offered over six or more weeks. These timeframes should be studied for children of different socio-economic and cultural backgrounds, science learnings, and developmental and grade levels. The impact of intensive over extended programs may vary with the type of science learnings sought as well. The study of these issues will have important implications for science curriculum development, especially for minority children.

Teachers questioning

The growth found in preservice teachers' skills of writing question sequences to guide children's development of a valid inference argues for teacher training that includes field experience like that offered by the project. A key factor seems to have been modeling. The students commented about the usefulness of their first-hand experience, as learners, with the tour plan and classroom lessons. They also viewed the tour scripts and lesson plans as valuable aids for their teaching. One student commented that she felt secure with the scripts and plans provided because she knew that "they'd work" in contrast to the tours and learning activities that she might plan. The value of modeling and the provision of scripted teaching materials for teacher evaluation has some precedent. The concept deserves further test in practice for helping preservice teachers develop the skills of effective teaching practice.

Another avenue for research is the question of museum field work in teacher education. Exhibits can be extraordinary instructional resources if used effectively by the teacher. Will teachers who have been trained to use museum resources for instruction during their preservice education continue to use them during their classroom tenure? Also, can the museum exhibit help preservice teachers develop and hone their questioning skills? This study's straight forward method of coding and rating question sequences seems to hold promise as a teaching tool. Its clarity makes it especially suitable for giving novice teachers feedback on their questioning and helping them evaluate their question-asking skills to guide children's scientific observation and inferential thinking.

The connections between teacher training and docent training are readily apparent when museum exhibits are viewed as the instructional focus, when questioning is the instructional mode, and where children's thinking skills are the focus of learning objectives. This project's method of evaluating growth in preservice teachers' questioning may be applied to the training and assessment of the same teaching skills in docents. A study is now underway to test that assertion.

The Museum in Science and Teacher Education

The museum's most valuable contribution to education is its invitation to inquiry. The objects it displays are rich in potential for interpretation. Viewers who know how to make detailed and precise observations when examining the contents of exhibits and how to interpret their findings with care cannot fail to learn. Their learnings include the subject matter content of the exhibit and more: they are developing their cognitive skills as they engage in a form of detective work into the meanings held by exhibited objects. Those skills are the same ones that enable scientific investigation: observing and inferring. The science museum is a marvelous source of visual presentations that prompt scientific thinking.

Minority children are not the usual patrons of museums. For lower SES families, the museum may seem a world apart when it should be their unschool for life-long learning. Making it so requires school experiences that teach children how to use the museum: how to look, how to see, and how to inquire into the visual richness of exhibits. This demands special pedagogical expertise. Teachers must know how to engage children with visual presentations, how to question to direct children's observing and how to guide their inferring while looking at objects and visual displays. Teaching in museums is different from classroom practice; the print literacy required for most classroom learning is developed differently from the visual literacy demanded for museum learning.

Educational research must explore ways of teaching in museum settings to develop children's thinking. The literature in museum education does not yet explain the special characteristics of teaching with the visual presentations of exhibits. It does not clarify the pedagogical repertoire for teaching in museums. Therefore, it offers little direction for the education of teachers who are as effective in galleries as in classrooms. That teachers must be prepared for practice in both settings is critical if the museum is to become accessible to the less advantaged students who most need to use its resources to expand and enrich their knowledge, and to practice inquiry. This study's findings suggest that minority children can benefit from engagements with science exhibits in definable cognitive ways and that museum field work can contribute to teacher education for children's growth in logical thinking. The findings also make clear that there is a great deal more to learn about museum teaching and learning.

References

Beard, J. (1971). The Development of Group Achievement Tests for Two Basic Processes of AAA's Science Process Approach. Journal of Research in Science Teaching 8: 179-183.

Blosser, P. & Helgeson, S. (Eds.) (April 20-30, 1984). National Associateion for Research in Science Teaching Annual Conferences, Abstracts of Presented Papers. New Orleans, LA.

Dillashaw, F.G. & Okey, J.R., (1980). Test of Integrated Process Skills for Secondary School Students. Science Education. 64: 601-618.

Doran, R. L. (1979). Assessing the Acquiring of Basic Skills. School Science and Mathematics. 79: 618-26.

Fuller, M. W., Jr. (1970). The Development and Status of Science Center and Museums for Children in the United States. Unpublished doctoral dessertation, University of North Carolina at Chapel Hill.

History News Dispatch. (August, 1986). American Association for State and Local History.

Koran, J.J., Jr. & Koran, M.L. (1986). A Proposed Framework for Exploring Museum Education Research. The Journal of Museum Education, 11: 12-16.

Linn, M.C. (April 13-17, 1981). Evaluation in the Museum Setting: Focus on Expectations. Paper presented at the Annual Meeting of the American Education Research Association, Los Angeles, CA.

Mayer, V.J. Unpublished Evaluation Instruments in Science Education: A Handbook. ERIC. Information Analysis Center for Science, Mathematics, and Enviornmental Education, Columbus, Ohio, ED 095015.

Molitor, L. (1971). An Investigation to Develop and Instrument to Evaluate the Science Skills of Inference, Prediction, and Verification of Children in Grades Four, Five, and Six. Unpublished doctoral dessertation. University of Pennsylania.

Molitor, L.L. & George, K.D. (1976). Development of a Test of Science Process Skills. Journal of Research in Science Teaching. 13: 405-412.

Riley, J.W. (1972). The Development and Use of a Group Science Test for Selected Processes of the Science Curriculum Improvement Study. Unpublished doctoral dessertation. Michigan State University.

Tannenbaum, R.S. (1968). The Development of the Test of Science Processes. Unpublished doctoral dessertation. Columbia University.

Youngpeter, J.M. (1973). Science Museums and Elementary Teacher Education: Organizing and Implementing Programs in Natural Science. Unpublished doctoral dessertation. Columbia University.

TABLE 1

INTERRATER AGREEMENTS FOR QUESTION CATEGORIES

Question Categories	R1 & R2	R1 & R3	R2 & R3
OBS-GEN	.895 ***	.852 ***	.738 ***
OBS-SPEC	.681 ***	.580 **	.145
OBS-C/C	.388 **	.836 ***	.434 *
INF	.784 ***	.798 ***	.603 **
SUP EVD	.913 ***	.917 ***	.833 ***
KNOW	.936 ***	.797 ***	.771 ***

* $p < .01$

** $p < .05$

*** $p < .001$

TABLE 2

Between-Within Design ANOVA Comparing Experimental and Control Group Performance Over Time on Tests of Scientific Observation and Inferential Thinking

		F Scores for each Scale			
Source	df	Observation	Inference	Supporting Evidence	Differentiating Difference From Observation
Between					
Group	1	1.02	1.47	.20	2.47
Within					
Time	3	85.54**	17.93**	1.66	.12
Time (x) Group	3	0.15	2.89*	.62	.31

** $p < .001$
* $p < .03$

TABLE 3

Pretest-Posttest Comparisons for 2-Week Intensive Program

Test Scale	Means Pretest	Posttest	F (df = 40)
Observation	47.81	66.93	3.77***
Inference	66.34	76.30	2.08*
Supporting Evidence	61.22	67.27	3.16*
Differentiating Obs./Inf.	56.53	66.38	3.13**

*** $p < .001$
** $p < .01$
* $p < .05$

TABLE 4

Comparison of Pretest and Posttest Mean Ratings of Question Sequence Written by Preservice Teachers to Guide Children's Scientific Observation and Inferential Thinking

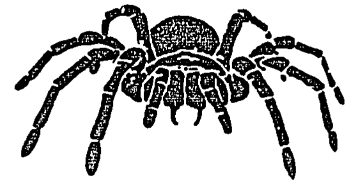
Test	N	Mean	SD	t
Pretest	31	2.9032	1.399	1.87*
Posttest 1		3.4355	1.039	
Pretest	38	2.6842	1.353	1.57
Posttest 2		3.0921	1.493	

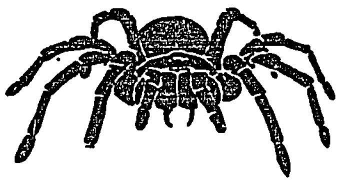
* $p < .07$

FIND THE SPIDER LOOK-ALIKE

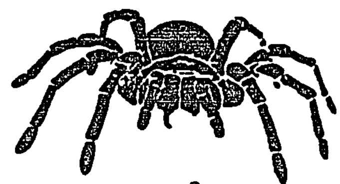
WHICH SPIDER LOOKS EXACTLY LIKE THIS ONE?

MARK IT WITH AN "X".



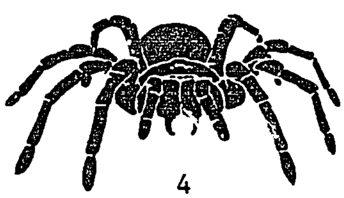


1

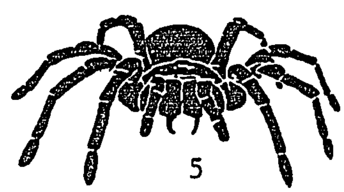


2

3



4



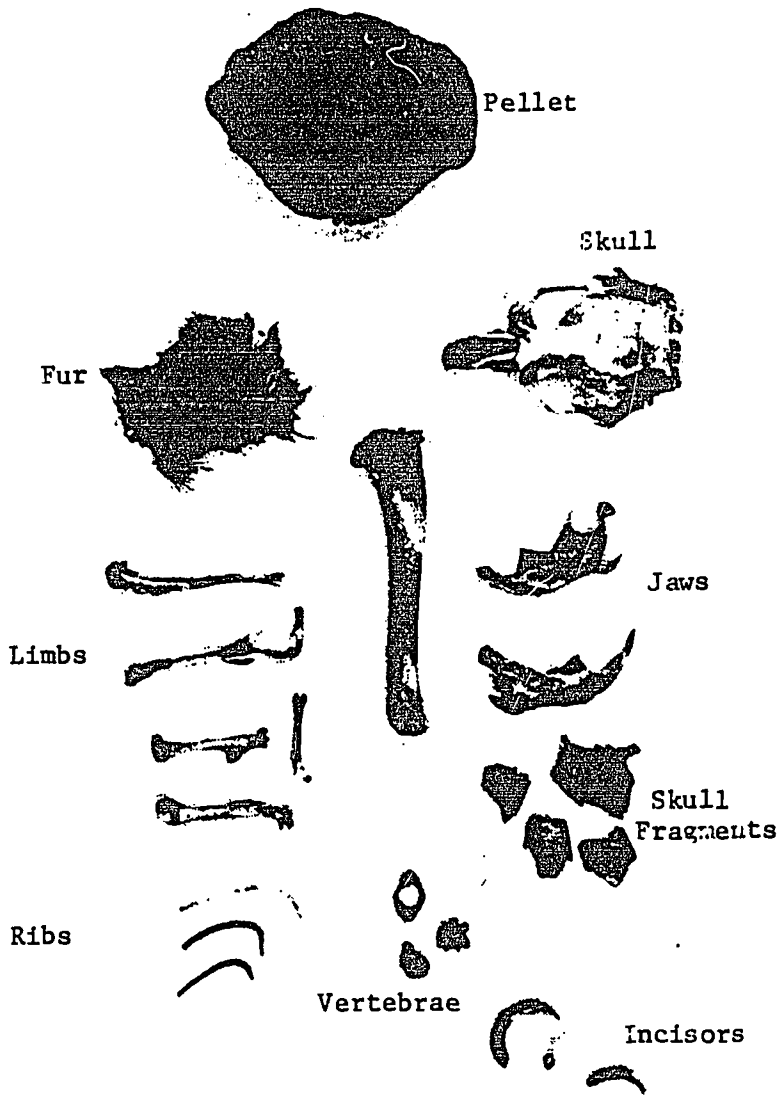
5

FIGURE 1

Sample page from the first subscale to test for ability to observe details. The correct answer is number 4.

WHAT DID THE OWL EAT?

This pellet was coughed up by an owl because he could not digest some of the things he ate. What do you think he ate?



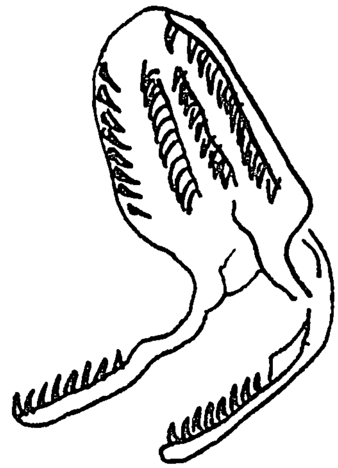
Cirlce the letter of the food you think the owl ate:

a. a bird
b. a plant
c. a fish
d. a rat
e. a snake

FIGURE 2

Sample question from the section on inferring. The correct answer is "d".

THIS IS A PICTURE OF A SNAKE'S JAWS.



WHEN A SNAKE CAPTURES A LIVE MOUSE, THE MOUSE CANNOT ESCAPE FROM THE SNAKE'S JAWS. WHAT CAN YOU SEE IN THE SNAKE'S JAWBONE THAT TELLS YOU THAT THE LIVE PREY CANNOT EASILY WIGGLE OUT? MARK WITH AN "X" THE BEST EVIDENCE.

_________a. The teeth point backwards.

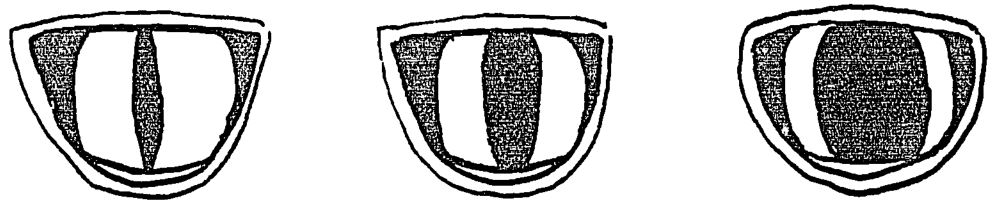
_________b. The jaws are large.

_________c. The jaws can open wide.

_________d. There are two rows of upper teeth.

FIGURE 3

Sample item to test for ability to identify supporting evidence for an inference. The correct responses are (a) and (d). The item is scored for the total number of responses correctly marked and unmarked. Total points on this item are <u>four</u>.



A

IN BRIGHT LIGHT

B

IN SHADED LIGHT

C

IN THE DARK

THESE ARE PICTURES OF A LIZARD'S EYES IN DIFFERENT AMOUNTS OF LIGHT. MARK ALL THE STATEMENTS BELOW WITH AN "X" THAT ARE INFERENCES--THINGS YOU THINK BUT DO NOT SEE WHEN YOU EXAMINE THE EYES.

_____ a. The eyes open wider in the dark.

_____ b. The eyes grow smaller in the light.

_____ c. The lizard hunts at night.

_____ d. The lizard's eyes are protected from bright light.

NOW MARK ALL THE THINGS YOU ACTUALLY SEE--YOUR OBSERVATIONS, NOT THE THINGS YOU INFER ABOUT THE LIZARD OR ITS EYES.

_____ 1. The eyes are important to the lizard.

_____ 2. The eyes vary in size in response to light.

_____ 3. The eyes can narrow to slits.

_____ 4. The eyes can open very wide.

_____ 5. The lizard's eyes are sensitive to light.

FIGURE 4

Sample item to test for ability to differentiate observations from inferences. Correct responses are c, d, 2, 3, 4. A total score on this item is 9, including all correctly marked and unmarked responses.

FIGURE 5

QUESTION CATEGOR[illegible] [illegible] STUDENT QUESTIONING SEQUENCES TO D[illegible]LOP CHILDREN'S INFERENCES

OBS OBSERVATION: Asks child to collect data through the senses

OBS-GEN (General) --in an open-ended way

E.g., "What do you see in this plot of earth?"

OBS-SPEC (Specific) -directed toward a specific focus

E.g., "What color is the butterfly's wing?"
"How does it feel?"

OBS-C/C (Compare/Contrast) --with reference to comparisons and/or contrasts

E.g., "How do these insects look alike?"
"How are they different in size?"

INF INFERENCE: Asks child to:

--make judgements from knowledge and observation

E.g., "Would this be a good home for squirrels?"

--interpret findings

E.g., "By examining all the insects on this plot of earth, which one seems best adapted to this area?"

--extrapolate

E.g., "What do the tracks you see tell you about who has been here?

--hypothesize

E.g., "How long do you think this plot of earth will continue to look the way it does today?"

--apply

E.g., "If water beads on waxed paper because of its cohesive forces, why does it bead on the hood of a waxed car?"

SUP EVD SUPPORTING EVIDENCE: Asks child to cite observations and/or information to justify inferences made by children or teacher

E.g., "What clues can you see that tell us that raccoons were eating here?"

KNOW KNOWLEDGE: Asks child to recall factual content or past experience

E.g., "How did/can we tell the age of a tree?"

"How long does a butterly live?"
"What's the name of this wildflower?"

OTHER This category is for question types that cannot be categorized as any of the foregoing types.

FIGURE 6

Rating Scale for Questioning Sequence to Guide Children's Scientific Observation and Inferential Thinking

RATING	CRITERIA
A=5	(5) Starts with an observation question, followed by several additional observation questions and, maybe, a few knowledge questions. The observation questions clearly guide students to attend to details and information that are relevant to the desired inference. The questioning sequence ends with one or two inference questions and, perhaps, one calling for supporting evidence. The sequence is distinguished by its ability to cause the learner to pay attention to specific details that inform the reasoning process toward formulation of the desired inference.
B=4	(4) Starts with an observation or knowledge question, followed by several additional observation questions and, perhaps, some knowledge questions. The sequence ends with an inference question or two but no request for supporting evidence. The sequence is marked by questions that lead the students to piece together information and observations that can support the desired inference, but the sequence is not so tightly developed as one that might be rated "A". The learner may have to rely on unsolicited past experience or knowledge or guess work to arrive at the desired inference. However, the questions guide sufficiently well to help the learner make a correct inference.
C=3	(3) Starts with an inference, knowledge, or observation question followed by additional knowledge or observation questions. The sequence ends with an inference question but the logical relationship of observation or knowledge questions to the intended inference is not always straightforward. There are obvious gaps. The learner could arrive at an incorrect inference.
D=2	(2) Starts with an inference or observation question but few observation questions of any type are evident. The majority of questions call for knowledge or inference without a clear indication of the direction thinking is expected to take. The relationship between the line of questioning and the ultimate goal is unclear.
F=1	(1) This sequence contains only inference or knowledge questions. No questions calling for observations are included. There is no logical structure to the sequence that clarifies the thinking desired. An inference is not developed.