DOCUMENT RESUME

ED 035 038 24 CG 004 800

AUTHOR Nicol, Elizabeth H.
TITLE The Development of Validated Museum Exhibits. Final Report.
INSTITUTION Childrens Museum, Boston, Mass.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO BR-5-0245
PUB DATE May 69
CONTRACT OEC1-6-050245-1015
NOTE 108p.
AVAILABLE FROM The Children's Museum, Jamaicaway, Boston, Massachusetts 02130 ($5.00 per Copy)

EDRS PRICE EDRS Price MF-$0.50 HC-$5.50
DESCRIPTORS *Autoinstructional Aids, Children, Educational Experience, *Exhibits, *Instructional Materials, *Learning, *Museums, Students, Testing

ABSTRACT

Exhibit development, as conceived in this report, is an evolutionary process, drawing the museum visitor into the collaborative venture of testing and improving the exhibits. The findings of contemporary learning research were put to work in the arrangement of activities and specimens that engaged children through self-instructional sequences. The series of learning stations, which dealt with concepts about animals' teeth, was accomplished in two phases. After each, the effectiveness of the series was measured in game-like tests given to several hundred children both before and after viewing the exhibits. Results showed that gains from pretest to posttest were highly significant. The learning effect appeared fairly uniform for all age groups and for boys and girls. The report concludes with a consideration of the findings and observations in relation to exhibit effectiveness research and possibilities for further development. (Author/KJ)

FINAL REPORT

Project No. 5-0245

Contract No. OECI-6-050245-1015

The Development of Validated Museum Exhibits

Elizabeth H. Nicol, Ph.D.
The Children's Museum
Jamaicaway
Boston, Massachusetts

May 1969

U. S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Office of Education, Bureau of Research

ED035038

FINAL REPORT

Project No. 5-0245

Contract No. OECI-6-050245-1015

THE DEVELOPMENT OF VALIDATED MUSEUM EXHIBITS

Elizabeth H. Nicol, Ph.D.
The Children's Museum
Jamaicaway
Boston, Massachusetts 02130

May 1969

The research reported herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U. S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Office of Education, Bureau of Research

Photos by Jerry Berndt and Steven Caney

Copies of this report may be obtained from The Children's Museum, Jamaicaway, Boston, Massachusetts 02130 $5.00 Postpaid

TABLE OF CONTENTS

TABLE OF CONTENTS continued

LIST OF FIGURES AND TABLES

ACKNOWLEDGMENTS

A project of this nature owes much to the cooperation of many persons:

Michael Spock. Director of the Children's Museum, who was an initiator of the project and a stimulating influence throughout its evolution.

Phyllis D. O'Connell, Assistant Director of the Museum, who skillfully husbanded our resources, watching over all monetary details.

Jenefer Merrill, Harold Hall, Allan Conrad and Binda Reich, who as members of the project staff bore the main burden of the work; and Signe Hanson, who took care of all our graphics needs.

Michael Sand of Michael Sand and Associates, Cambridge, Mass., who designed the developmental version of the exhibits and collaborated in other aspects of the work.

Allan Conrad of the Design and Production Department of the Museum who designed the permanent version of the exhibits.

Others to whom I owe thanks are:

Dr. Neil Todd, geneticist, New England Regional Primate Research Center, Southboro, Mass., who generously served as our subject-matter consultant.

Mr. Robert E. Horn, President, Information Resources, Inc., Cambridge, Mass., who was one of the project initiators.

The Ampex Foundation, Redwood City, Cal., and Mr. Frank Rush, Ampex Corporation, Waltham, Mass., who arranged a magnificent gift of recording and playback equipment for the project.

Mr. Edward B. Shaw, taxidermist, who made important donations of specimens.

Mr. Paul Rockwell, Director of Development and Public Relations for the Museum, whose voice gave life to our recorded scripts.

The many members of the Museum staff whose talents and experience often came to our aid.

And finally it remains for me to acknowledge our debt to the hundreds of children whose candid reactions guided our work.

E. H. N.

SUMMARY

Exhibit development, as conceived in this report, is an evolutionary process, drawing the museum visitor into the collaborative venture of testing and improving the exhibits.

In this view, evaluative testing becomes an instrument of design to be applied throughout the developmental period. It should not prescribe rules to hamper the creative process but should work only to clarify the exhibit message and to improve the techniques of presentation.

This development philosophy was put into practice in the preparation and evaluation of a series of instructional exhibits for children at the Children's Museum of Boston. At every stage from initial inception on, the children's reactions guided us in adjusting the series. Special construction materials afforded the flexibility that permitted quick response to the results of tryouts.

The findings of contemporary learning research were put to work in the arrangement of activities and specimens that engaged the children through the self-instructional sequences.

The series of learning stations, which dealt with concepts about animals' teeth, was accomplished in two phases. After each, the effectiveness of the series was measured in game-like tests given to several hundred children both before and after visiting the exhibits. Systematic observational studies were carried out as well.

Statistical analyses of the learning outcomes showed that the gains from pretest to posttest were very highly significant. They also revealed that in spite of varying backgrounds, improvement was shown by 76% to 78% of the children visiting the exhibit.

The learning effects appeared fairly uniform for all age groups; no consistent differences were found between the scores of boys and girls.

The child visitors spent an average of 20 minutes in the completed exhibit (exclusive of waiting-line time); they made repeated visits; they liked headsets for listening to the recorded messages; they did not engage in indiscriminate button-pushing.

The report concludes with a consideration of the findings and observations in relation to exhibit effectiveness research and possibilities for further development.

Chapter 1

INTRODUCTION

The purpose of most exhibits is to tell the viewer something. Yet surveys suggest that if the average visitor stops at all, he is held only a few seconds -- hardly long enough to decipher the message so thoughtfully designed for him.

In recent years the museum world has grown increasingly concerned with problems of exhibit effectiveness and with ways of improving the communication value of museum displays. The project that we are now reporting proposed a strategy for exhibit development that relies on the evaluative techniques of modern science. The key idea is that effectiveness grows by subjecting an exhibit to repeated periods of tryout with visitors whose reactions are captured by evaluative methods.

In the course of a twenty-nine month program, we applied evaluation as a tool in the development of instructional exhibits for children visiting at the Children's Museum of Boston. Throughout the process of preparing the exhibits, child visitors worked with us trying out the materials. Their reactions, as recorded by tests and observations, told us where the exhibit was not working effectively to tell the story we intended.

The exhibit is called "Teeth"; using skulls, models, photographs, and a variety of tooth specimens, it shows how teeth differ in shape and in function. The particular topic, however, is only of secondary interest here for the main purposes of the project were concerned with methods of development and evaluation.

The approach was to evolve, as the first phase, a series of learning stations dealing with a simple concept or cluster of related concepts. When cycles of tryout and revision had brought these stations to an advanced stage, a formal period of evaluation followed. Tests of the children's knowledge before and after visiting the exhibit gave us a measure of its effectiveness.

The second phase of the project consisted of expanding the series to take in another closely related concept-cluster; then the combined series of stations was tested. The fact that the learning outcomes were measured gave rise to the name "validated exhibits" in the title of the project.

In the course of the tests and observational studies we learned a great deal about our visitors and how they typically behaved in areas where one must follow a given sequence from station to station and in free areas where any order may be chosen.

This report is the story of how the children collaborated with us in these endeavors, of how their reactions caused us to change our materials, of how much they learned about the topic and of how much we learned about building exhibits for children.

The next two chapters explain the philosophy of development and the general evaluative methods we used. Following pages show how the philosophy and methods were applied in evolving the series of instructional stations that eventually came to be called "Teeth."

It may help the reader to have a brief chronological framework for the project. A few landmark dates are provided to serve that function:

Project began	June 15, 1966
Topic selected	July 22, 1966
Tryouts began	August 1, 1966
Opening of completely automated version of shearing-molar series	August 8, 1967
Formal evaluation of that series	December, 1967 - January, 1968
Grinding-molar series installed	early February, 1968
Formal evaluation of combined shearer-grinder complex	February 17-25, 1968
Evaluative testing of revised series now including canine and incisor units	April 16-21, 1968
Opening of redesigned exhibit series, constructed in permanent materials and in duplicate	October 19, 1968
Project ended	October 31, 1968

Chapter 2

THE MUSEUM AS A LEARNING ENVIRONMENT

Museums design learning experiences.

No longer are they passive mausoleums for the preservation of the curious or the beautiful; skillful use of new techniques in the audio and visual arts is making them more attractive and exciting places to visit than ever before.

While seeking ways to improve their exhibits, the museum world has remained strangely impervious to the concepts and techniques that have helped other educational institutions become more effective. Exhibits often seem out of touch with modern learning principles and instructional methods. Most importantly, the question of whether exhibits do provide learning experiences is often left unanswered. A message is shot off into the air . . . Whether it lands on uncomprehending eyes or ears we seldom know.

If we accept the educational role of the museum and, as its task, the arranging of learning experiences, then several points follow. The first is that exhibit design should make use of the best that is known about the learning process. The second is that those who create exhibits must adopt evaluative procedures in order to determine whether the educational function is being served or not.

Education in the schools was slow to examine critically its own methods and goals. (Even yet teachers and textbooks are rarely evaluated.) But in the nineteen-fifties the demands of a rapidly changing world jolted the education field into a productive acquaintance with new ideas from science and technology. Two areas of discourse were particularly helpful:

1. The research findings on learning together with the hypotheses of learning theorists, and

2. The analytic techniques and evaluative concepts from the systems design field.

Education has profitted from this cross-fertilization process in several ways but primarily in increased insight into the nature of its own endeavors and into methods for appraising its own progress.

In view of the number of common elements in the learning environments of school and museum, there is reason to hope that a similar approach applied to design of exhibits might extend their quality and enhance their educational impact.

The objectives of the present project were to explore the potential value of similar concepts and methods in devising guidelines for exhibit planning and development.

Let us quickly consider some of the new ways science and technology can suggest for looking at the museum field.

The Systems Approach and the Communication Model

Systems design concepts, and especially communication system concepts, contain certain emphases that have influenced educational technology. Chief among these is the emphasis on:

- clear specification of objectives for the total system,
- analysis of system components, their characteristics and functions,
- the iterative nature of the system development process and therefore on the necessity for a feedback loop.

The analogy of the most familiar of systems, the communication system, has been applied to teaching: the teacher is the transmitter, the subject matter of the course is the message, the voice is the transmitting medium, and the student is the receiving mechanism.

The analogy is appropriate for exhibit development too. We can identify the designer[1] as the sender, his script as the message, the exhibit as the transmitting medium, and the visitor as the receiver. Calling these a system reminds us that the success of an exhibit is not a matter of the excellence of one part (the exhibit) but rather that it depends on how well we understand the characteristics of the audience and their interrelations

[1]For convenience and brevity, we will hereafter use the word 'designer' to refer to the person (perhaps the 'corporate' person) who directs exhibit development from inception to completion.

with the exhibit components, and how well we have defined what the elements are supposed to accomplish. Let us consider what this means in practice.

In our original project proposal (March 1, 1965), we mentioned certain deficiencies of the designer-exhibit-visitor triad as a communication system. The first shortcoming was the common failure to say precisely what the system is supposed to do. Any question of how effective a telephone system, a school lesson, or an exhibit is has little meaning unless the objective was stated carefully in terms specific enough to permit some measure of success to be made. In an educational endeavor, the purpose is best stated in terms of what the learner can do afterwards that he could not do before. Terms such as "to appreciate" and "to understand" are too vague to guide design. Yet it is in terms like these that more often than not, the exhibit designer formulates his goals.

In the systems approach, the nature and characteristics of the receiver are given close study. By contrast exhibit designers often neglect the intended recipient of their message. Museums often seem to care more about the effect of their exhibits on other designers and curators than upon the visitors they ostensibly serve. The better one understands the visitor's characteristics, interests, skills, and vocabulary, the more effective the communication will be. In the absence of information about the visitor, the message can fall far short of its mark.

Related to the need to understand the 'components' of the system is the need to select appropriate communicating media. Too often visitor and exhibit must interact in a narrow range of stimulus and response modes. The transmitting media of the exhibit tend to be limited in the kind of stimulation used to communicate their message, relying heavily on the static visual display of objects and of verbal labels. For his part the visitor is confined to looking, reading labels, and walking around. A more successful matching of exhibit, message and visitor can be achieved by calling upon a broader spectrum of response modes (touching, tasting, smelling, listening to a recording, manipulating objects or colors, taking things apart, drawing, watching a film, solving a puzzle).

The communications model includes another feature that the usual exhibit does not: control over the sequencing of message units. It would be traumatic if the telephone system transmitted our sentences in scrambled order. Yet the exhibit designer has no assurance that something like that will not happen to his message; the visitor can

dart about from one unit to another in any order he chooses, lingering over some and skipping others entirely. The communication model serves to remind us that exhibits whose messages are of an "argument-building" nature must have sequence-control measures to insure effectiveness.

The most glaring deficiency of the exhibit as a communication system is the absence of any feedback link to the designer from the visitor. In this respect the classroom teacher has a distinct advantage over the designer: he can easily find out whether his message has been received by the student -- by asking questions, by noting the stifled yawns or glazed expressions. In contrast, the exhibit designer seldom even inquires about what happened to his message. Talk of seeking guidelines for exhibit effectiveness is pointless until the practice is established of measuring the quality of the transmission.

The feedback link permits comparison of the actual output of the system with the intended output. A discrepancy shows that the original aims are not being satisfied and so of course adjustments are made. If the teacher finds his carefully thought-out plans had disappointing impact, he modifies his technique, his message, or even his goals. But few exhibits are tested and revised.

This brings us back to a concept mentioned earlier: that system development is an iterative process. One of the key ideas in the systems approach is that a system is not just created, plugged in, and forgotten -- it is achieved only through a process of successive approximations. The reason for this is that initially there are many unknowns; to get started at all it is necessary to make a number of assumptions or hypotheses about what will work and how parts will interact. Devices built in to monitor the system outputs provide the data for judging whether these assumptions are true or need to be modified or abandoned. By successive approximations, the gap between results and objectives is reduced.

That such a process of iteration would be fruitful for exhibit development is a major hypothesis of this project. Exhibit building is not a process of following well-established rules for combining objects, display techniques and conceptual information. It too proceeds on a series of assumptions. The assumptions need to be verified and retained, or disconfirmed and discarded. Through the conscious effort to learn by such reality testing, a viable set of guidelines to exhibit effectiveness may begin to take shape.

In summary, systems design concepts suggest some new ways of looking at exhibit-design problems and some new ideas about development procedures. We have tried to incorporate these in the developmental philosophy followed in this project.

In systems design, once the basic decisions about what the system is to do have been made, the question of how to set up conditions to accomplish it must be considered. Here the science and technology of the particular domain come into play. For our project, learning was the objective. Therefore, as designers of "educational events,"[2] we should take advantage of the world of learning research and instructional technology. We briefly review some of the concepts and methods relevant to our task.

Learning Research

To arrange effective learning environments, we should draw upon the best information about the different kinds of learning, the best methods for each, the effects of various conditions, and the possible differential effects for learners varying in abilities and personal characteristics.

This sounds fairly straightforward. But it would be sheer fantasy to pretend that there exists a systematized body of unequivocal research findings ready to be consulted, or that there is any widely accepted theoretical point of view. The field is complex and the gaps are not inconspicuous.

It is important to remember that there are different kinds of learning (rote-learning, problem-solving learning, motor skill learning, to name only a few). The conditions that are advantageous for one type may not be best for another. The objectives of each are quite different, a point sometimes overlooked. Different hypotheses about the nature of the <u>learner</u> -- whether he is a passive receiver of stimuli or a selective self-organizing system -- will decidedly influence the way learning conditions are arranged.

This is not to say that the quest for help in these quarters should be abandoned but to caution against an expectation of finding ready-made

[2] Term of R. M. Gagné, Psychological Review, May, 1968

solutions. The exhibit designer should be acquainted with the range of options open to him, with some of the complexities of the area, with some of the mistakes that have been made. There does exist, for his use, a rich body of research findings with some well-established, oft-repeated results; there are some well-reasoned hypotheses and appealing theories, some good bets, and some educated guesses. Since so much of exhibit design must proceed on assumptions, as was mentioned before, it stands to reason that expert help, even if not in pat equation form, should nevertheless cut down on development and revision time.

The learning problem we ourselves chose for this project involved teaching several simple concepts based on multiple discriminations. This kind of learning has been called "reception learning" by Ausubel (1963), who defined it simply by saying: "the content of the learning task (what is to be learned) is presented to rather than independently discovered by the learner." This is of course the major way we acquire the concepts and principles we understand and use -- and it is also the quickest way.[3]

Use of reception learning does not imply a view of the learner as a passive receiver. In fact, as will be explained shortly, our learning conditions required the active participation of the child. The "givens" in the exhibit, the concepts and principles, were expected to acquire personal meaning for the child through his manipulation of the responsive materials of the exhibit.

The choice of reception learning as our starting point for this project influenced our methods, the kinds of learning research we drew upon, and so on. A different objective would have required a different kind of sieve for filtering through the storehouse of research findings.

---

[3]For a time, the widespread and often uncritical adoption of "discovery learning" led to the neglect in theory (not in practice) of this most common type of learning; the process was helped along by unnecessarily poor techniques on the part of teachers for reception learning. Current opinion is now coming about to view both discovery and reception learning more dispassionately and to recognize that both occupy valuable places in our learning armamentarium.

Given the kind of learning problem we chose for this project, the kind of environment in which we were to operate, and the age of the learners we expected, we then tried to arrange a learning situation that would include factors for which strong research support exists. The factors we used were:

1. self-pacing: allowing the child to proceed at his own rate provides for a range of individual differences in learning speed and in adjusting to novel situations;

2. active participation with concrete materials: for learning in general, overt responding seems to be superior to passive, covert responding. For children especially, learning seems to be enhanced by opportunities to interact with concrete materials; ideas and relations presented verbally can be confirmed in immediate experience by direct manipulation of objects, and in this way they acquire meaning;

3. information-response sequences: after each small unit of new information, it helps to have opportunities for trying out and confirming the ideas or relations before coping with more new material -- each gain is "solidified" before more is added;

4. multisensory involvement: for children, learning experiences involving several senses generally have certain advantages, some of which may be indirect in the sense that they make for an attractive, exciting situation that holds the child and so he learns more because he stays longer;

5. immediate knowledge of results (feedback): knowledge of whether or not one's responses are right usually has two effects: an informative one, confirming or correcting his conceptions and therefore guiding his later behavior; and second, a motivational effect, a feeling of satisfaction and an increase in confidence follow a success. In terms of the communication model, the receiver finds out whether he has received the message correctly. If not, he tries again until getting the 'correct' signal and thus always ends each unit with a success.

The way in which these learning conditions were worked into our own project will come up in a later chapter. We must first integrate the systems approach and learning research into a development procedure.

Implications for Exhibit Design

We have now reached the point where we can combine the ideas of this chapter into a practical strategy for exhibit development. The suggested steps to be followed are these:

1. Specify the purpose of the exhibits in measurable terms, saying what the visitor will be able to do or say after the experience that he could not do before.

2. Specify visitor characteristics, including the pre-exhibit knowledge they can be expected to have, their skills (e.g., reading), interests, ages, sizes, and so on.

3. Analyze the exhibit message into its component concepts, discriminations, principles, skills and so on.

4. Identify environmental effects. Where the exhibit has to operate may affect the success of the project. The need for accommodating to the surroundings must not be overlooked in the early planning stages.

5. Determine the conditions that will accomplish the objectives. If the purpose involves learning (as ours did), then the best estimate of the conditions for bringing about the learning experience is sought in this step; the ones we chose for our project were listed in the preceding section.

   The objective need not be learning of course; museums have other goals as well -- creating experiences of wonder and beauty, for instance, or changing values and attitudes. Whatever the objective, this step requires specifying as concisely as possible the techniques or arrangements that are expected to achieve the goal.

6. Design and construct the exhibit materials to conform to the requirements of the objectives and of the facilitating conditions. Here inexpensive, adaptable prototype construction

is required, we believe, for with this developmental philosophy, changes will be made.

7. Design information-gathering methods, ways of monitoring 'system performance' and assessing to what extent the objectives have been achieved.

8. Try out the exhibit with visitors of the type for whom the materials were intended. If the exhibit is composed of separable units, try out each segment as it becomes available. Even manual operation of units slated for eventual automation can pinpoint errors early and avoid costly adjustments later.

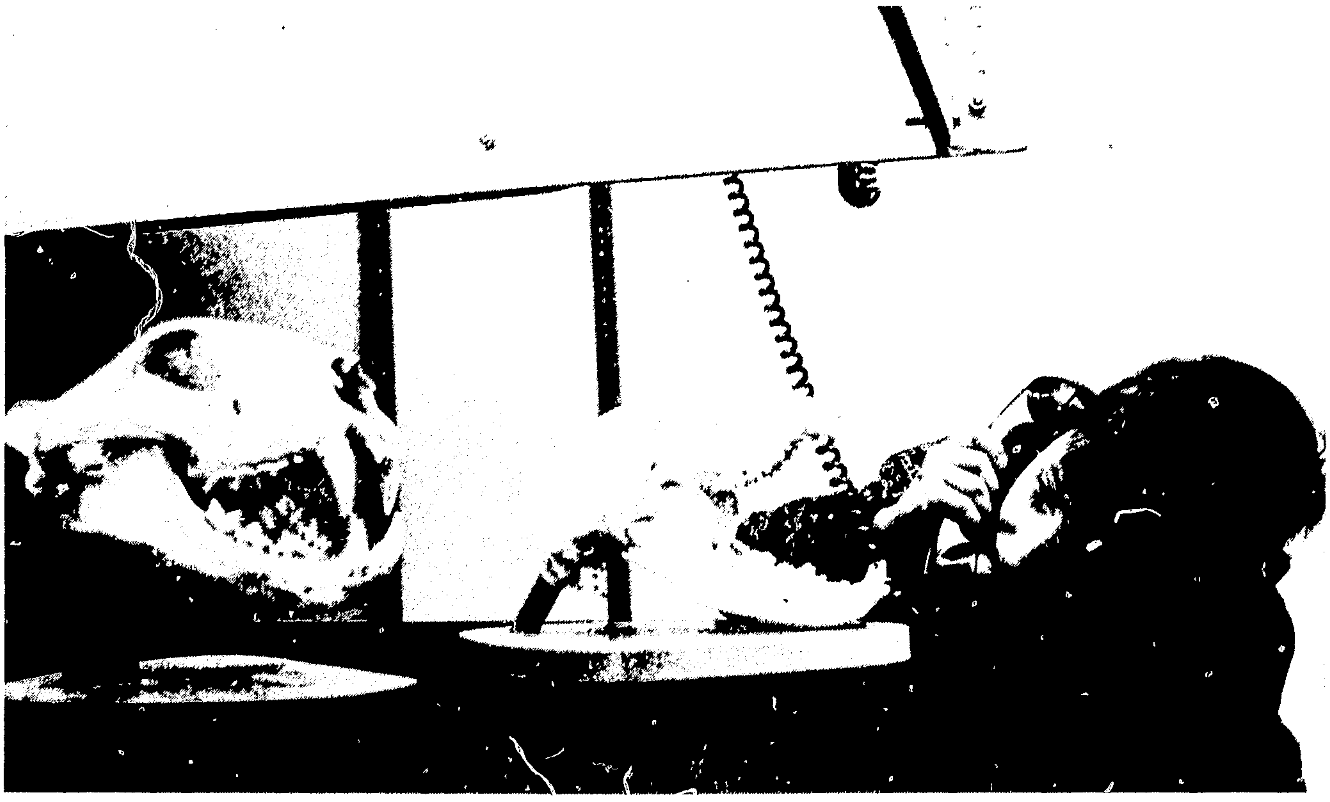
9. Apply performance data in the revision of objectives, redesign of construction, change of environmental situation and so on wherever modification is indicated.

10. Repeat the try-evaluate-revise cycles until the exhibit has reached a satisfactory level of effectiveness.

Summary

Borrowing ideas from modern technology, we have suggested how adaptive-corrective cycles of evaluation and revision might improve the process of exhibit development and foster the evolution of guidelines for effective exhibit-building.

In addition, we discussed how the accumulated research findings in psychology and education could be a strong force toward arranging more effective "educational events" in the museum setting. Some ideas about a practical strategy for exhibit development were outlined.




Start

Chapter 3

CHILD-PARTNERED EXHIBIT DEVELOPMENT

The exhibit-building practice we have adopted

- makes the exhibit audience (in our case, the child) a full partner in the development process,
- uses evaluative testing as an instrument of design throughout the development process.

In a very real sense the child helped write the script for our exhibit. At every stage from the initial formulation of the exhibit topic on, we were guided by the way the child interacted with the materials, concepts, and activities we were considering. Instead of the usual "duologue"[1] between museum and visitor, we tried to engage the children in a genuine dialogue by setting up situations where the children's behavior would indicate how effective our ideas were.

The way that our specific project took shape through this close involvement with the children will be the topic of the next chapter; here we outline the general strategy and discuss the evaluative methods that were applied.

In the earliest stages the concepts, possible sequences and potential materials were tried informally with individual children or with small groups. For example, the staff person might take specimens and artifacts for a potential topic into the Museum where they would soon attract a circle of visiting children. Sitting around on the floor, they would explore the various things and wonder over the "why" of some aspects. The staff member would experiment with an approach to the topic concepts, noting the ideas and questions raised by the children, the sources of misunderstanding, and indications of interest or boredom. Variations that did not work were either modified or discarded. A successful technique was pursued with other children, modified, and so on until the script began to take form.

---

[1]Term suggested by Professor Abraham Kaplan, philosopher of the University of Michigan, to refer to the communication form in which both parties take turns speaking but neither listens (N. Y. Times, January 10, 1969).

Throughout these days, a record was kept of the various approaches, their outcomes, and the impressions or conclusions of the staff. These constitute the earliest stages of evaluation.

As the scope of the topic was narrowed and objectives clarified, construction of prototypes began; when segments of the exhibit became available, the cycles of tryout and revision continued with the visitors' help. When difficult problems of selecting alternative designs or plans arose, we soon learned not to argue the pros and cons extensively but rather to take a quick mockup out onto the Museum floor to find out what our child partners thought of it. We did not have long to wait for answers.

For the most part, in the early days the answers came from watching the children in action, but they came also from talking informally with them afterward, exploring how much they extracted from the experience.

As the project took shape, formal methods of seeking information about the effectiveness of the materials came into use. When tryouts resulted in a dwindling number of adjustments to the exhibit, a period of concentrated formal testing would follow. The nature of these methods and of their different objectives is the subject of the next section.

Evaluation -- A Finger on the Pulse

Evaluation is a most powerful tool for exhibit development. Unfortunately, the term has been the victim of more unpleasant connotations; for many it creates a feeling of confinement, restriction, a dampening of the creative impulse and a deadening of perceptive intuition, to mention only a few. But these qualities are not inherent in the evaluation process itself but merely reflect the limitations of evaluators.

The planning of any instructional program involves a large measure of art, intuition, and pure guesswork. Anything that interferes with the creative exercise of these must not be tolerated. But evaluation, properly conceived, can be a fine-tuned, sensitive instrument for judging the progress of the enterprise. It must not risk the probing and prying that might kill the very thing it is trying to foster.

Evaluation per se is not concerned with laying down prescriptions about how to do things. Its only role is to determine whether a project has accomplished the goals it set for itself (whether the project be an exhibit, a text-book, a mousetrap, or a hotdog). Its only requirement is a statement of the objectives; this, however, is not always easy.

The practice we have followed is to formulate objectives in terms of the things we expect a child to be able to do after visiting the exhibit that he could not do before. Defining objectives in terms of "observable behavioral changes" is fairly standard now in education. Among the extensive writings on the subject, a small book by Mager (1962) was especially influential in showing how ambiguous were ways of phrasing objectives in terms like "to know," "to appreciate," and "to understand."

The more specific a set of objectives can be, the more they help the designer zero in on the best strategies for accomplishing them. And it is from the statement of behavioral changes that the evaluative measures are easily derived.

There are different levels of evaluation and different methods. The two sources of evidence of progress toward our goals that we relied on primarily were: systematic observations of behavior and gain scores in performance tests taken before and after visiting the exhibit. Each of these is described briefly in the next sections.

Systematic Behavioral Observations

For the earliest tryouts, as noted before, records of approaches and observations were maintained along with evaluations of the results. As the tryouts grew more structured, we began to record data on the individual children's behavior in the situations. These gave us a much more solid basis for generalization and design decisions.

And finally when the exhibit units evolved to the point where they were mechanized and operating without staff involvement, then we turned to systematic recording of the children's responses. For each separate unit or activity in the project, data-gathering forms were designed to record what the child did with the materials -- how long he stayed, the kinds of behavior he showed in interacting with the materials there, and so on. Observers were instructed to follow a child unobtrusively, recording his reactions on the long check-list (one of which is reproduced in the Appendix).

The general purpose of these observations was to identify the points where the activities did not function well, where the instructions were not understood, or where the materials were not used in the ways we had anticipated -- in short, they helped pinpoint the short-comings of the materials, sequencing, mechanical devices and so on. They also provided valuable supplementary information about our visitors.

The detailed record form, we discovered, provided much more informative data than did observers' unguided accounts. The checklist focussed their attention on the specific classes of behavior that were important to us and insured that the same kinds of information were systematically recorded for each child. The cumulative data gave us a more reliable picture of what was happening; they had an important role in guiding many of the changes made during the development process.

Pretest-Posttest Comparisons

In the final stages of evaluating the exhibit, we measured its actual effect in changing the visitor's behavior. Before entering the exhibit, children were asked to try a "game" in which their actions gave a measure of their knowledge of the exhibit information. A second measure on a similar "game" was taken after they had been through the exhibit. Gains from pre-exhibit to post-exhibit may be attributed to the exhibit experience. It is primarily this measure that constitutes the "validation" from which this project took its name -- the "validated" exhibit is one for which there is concrete evidence of changes produced in its visitors' repertory.

In interpreting gain scores, one must always consider the possibility that the pre- and posttests themselves might be instructive. Various control checks are possible; for example control subjects can take both the pretest and posttest before going through the exhibit. Gain in scores from such controls can be attributed to the informative value of the tests themselves or to the possibility that the child gradually adapted to a novel situation and only during later trials began to show evidence of a previously learned behavior. As later chapters mention, we did run certain control tests to check on these possibilities.

The question answered by the kinds of evaluation we used in this project concerns what a child can do immediately after leaving the exhibit. Other kinds of issues, such as questions of long-term effects, might have been raised; however the one of immediate effects not only suited our resources best, but it also provided us with a sensitive measure of the effects of exhibit changes and corrections.

Supplementary Data

In part of the exhibit, electric counters recorded right and wrong answers to a series of discrimination tasks. The counters were of course concealed from the children's view and were massed in a remote alcove for easy reading. They provided data that were later collated with those gathered in the behavioral observation records.

A second source of information, while not an evaluative method, should be mentioned. Two logbooks were kept for recording the history of project development and revision. One log was a diary of the project events; the other documented specific details of the exhibit units, the date when each went into operation and when changes were made.

As a case history of the evolution of the exhibit, we find these interesting (and humbling) records of our successes and failures, the latter including good ideas that could not be made to work, promising ideas that never got a trial, or bad ideas that we clung to too long. These are things that soon blur in memory or were not even recognized at the time; for us personally then, these records hold valuable lessons. They are also rich sources of research ideas.

Evaluation Schedules

For the most part we were able to schedule the main testing and systematic observation periods to coincide with school holiday weeks within the school year -- namely Christmas vacation, mid-winter vacation week (February) and spring vacation week (April). The advantages to us were twofold:

1. a larger population of children was present during these periods, while on ordinary school days visitors were predominantly neighborhood "regulars." It was expensive to keep an evaluation crew standing by when only a trickle of new visitors arrived;

2. university students were also available then to augment the project staff for these concentrated evaluation periods. After initial training, they remained at a higher level of efficiency in daily testing than if they had participated in the procedure only once every week or two with the week-end visitor populations.

Problems of Evaluation in a Museum Setting

For someone who has in mind conducting formal tests according to a carefully controlled experimental design, an operating children's museum presents problems. They are not insurmountable, but they make life difficult. Some of those that interfere with the precision of measurement are:

1. arranging a testing environment free from distractions;

2. the testing process must be confined to a very short time span -- visitors out for an afternoon's amusement cannot be detained long in a test; this means that we must be content with less reliable estimates of exhibit effects than if more test time were available;

3. selection of subjects and controls is not easy in a situation where groups of subjects from different economic and educational levels appear in a non-random order;

4. detection of "repeaters," children who pay return visits, is not a trivial problem. Their presence in an evaluation sample tends to dilute any difference between pre- and posttests. In our case, although there was no restriction on the number of visits to the exhibit, we wanted to test only those who were there for the first time. But because the tests themselves were attractive, asking the child if he had been there before often produced unreliable answers. A rapid-access card file of previous visitors was impractical to maintain on a daily basis with hundreds of children going through;

5. to try to include time controls is especially sticky in this setting; this is the case, for instance, when one might wish to make the time between pretest and posttest comparable for exhibit visitors and for control subjects -- but casual visitors who have perhaps waited in line for fifteen minutes cannot decently be shunted off to a substitute activity to serve as a control; we did not attempt to deal with this issue in our setup;

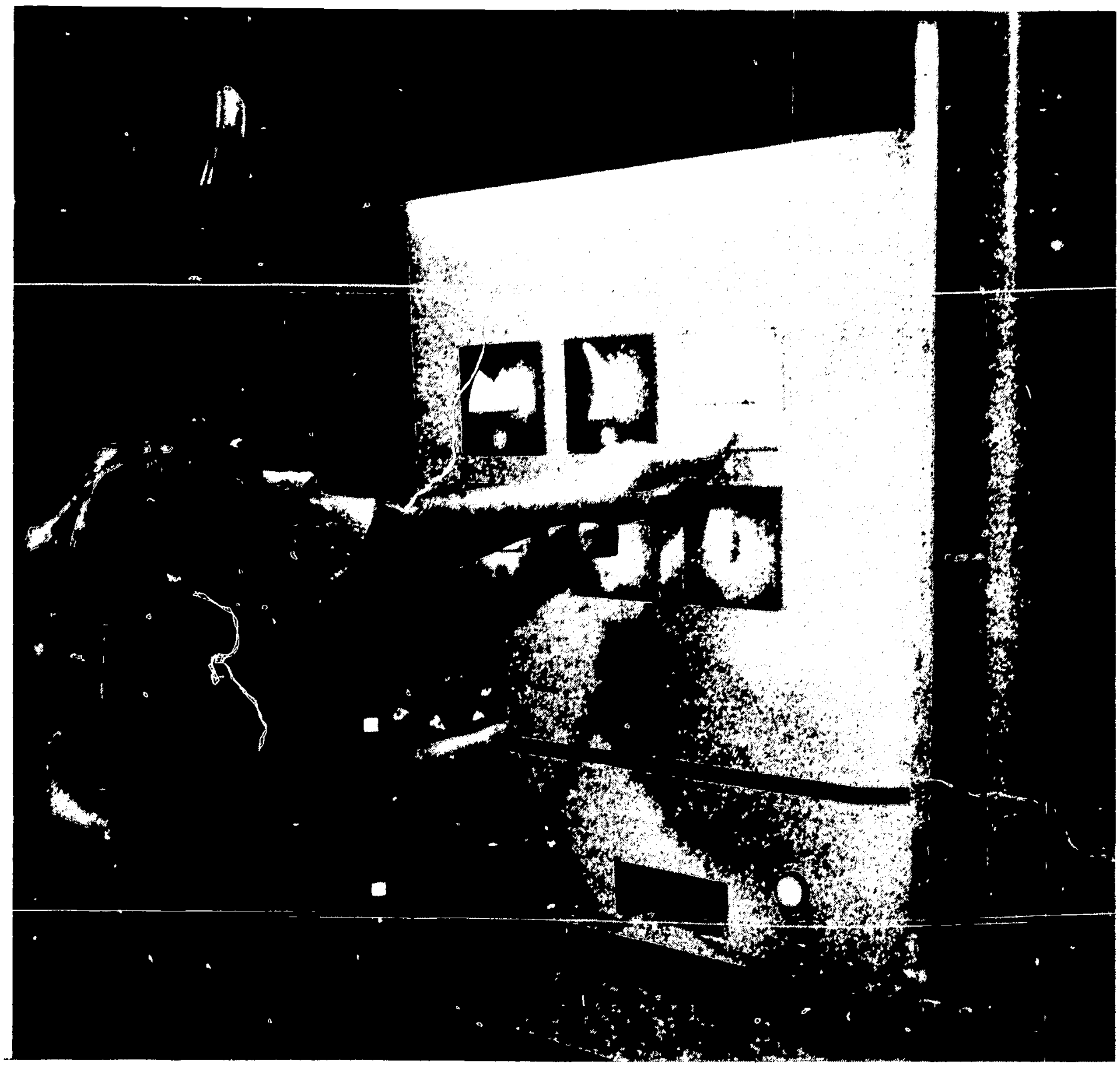
6. because it is hard to get reliable estimates of the average size of the effects and of the range of variations, it is difficult to specify the number of visitors that should be tested; in such cases one seeks a fairly large sample (whatever that may mean) in the hope that the effects will emerge in spite of the variations;

7. the drop-out problem is present, although for us it was not serious; but sometimes parents do reach the limit of their endurance and insist on extracting a small visitor in the midst of the evaluation process.

These problems are common to evaluation in other real-life settings where one must accept the best estimate of the effect he can achieve under the circumstances. Our own efforts were simply addressed to the question of the instructional value of these exhibits in an operational setting.

Summary

Evaluative procedures are powerful techniques for evolving instructional exhibits. To say this is to say that the exhibit audience is in a very real sense a partner in the development process, for it is their reactions as captured by the evaluation methods that guide adjustments to the exhibit. In our project many children joined us in evolving a series of instructional exhibits. The methods we used for measuring their reactions and some of the problems we encountered are outlined here.



Chapter 4

EARLY EVOLUTION OF THE EXHIBITS

Criteria for Subject-Matter Selection

We had set ourselves the task of developing two related exhibits in a development process that was committed to repeated testing and evaluation. The exhibits were to be derived from clearly defined, measurable objectives and known learner-visitor characteristics; in addition they were to involve the visitor actively in a varied and responsive environment, permitting feedback both to him and to the designer.

These commitments went a long way toward providing a framework for selecting a subject-matter area. In addition, several other features that seemed desirable to the project staff were added to the list of selection criteria against which prospective topics were weighed and judged for suitability. These criteria specified a topic that:

- was in the natural or social sciences,
- was not regularly covered in the elementary curriculum or by some other source in the child's experience,
- would appeal to the natural interest of children,
- could be presented through a variety of media,
- could be responded to in a variety of ways,
- could provide opportunities for the child to interact with real materials (as opposed to models or graphic representations only),
- could be dealt with within the limited attention span of the child visitor,
- was rich enough to support the later expansion into a second exhibit,
- was compatible with the interests and talents of the staff,
- was about equally interesting to boys and girls in the 8-12 year age range,
- might be transportable,
- would be recognized by other museums as a museum-type topic.

A score or so of potential topics was surveyed; possibilities for developing responsive devices and attractive displays for them were discussed. After initial screening sessions, six subject-matter areas remained as potential candidates. Briefly, they were:

a) elementary development of concepts of genetic inheritance through use of animal, insect and plant configurations showing frequency of key characteristics in succeeding generations;

b) some concepts about feet - either dealing with interpretation of footprints or with the relation of foot structure to terrain and life habits;

c) concepts in the communication area - such as how writing developed in different cultures, or how animals communicate (cries, calls, hormone traces, bees' dances and so on);

d) camouflage principles in nature explored by manipulation of backgrounds, coats, coverings of various kinds;

e) a walk-in model of a tree showing structures, fluid-carrying tubes, interrelations with insects, birds, etc.;

f) teeth structures and functions in mammals -- basic shapes and actions, along with interesting variations in certain species (beavers' incisors, elephants' and walruses' tusks, tiger canines, and so on).

The Choice of Topic for the Two Exhibits

After additional exploration of the potential subject-matters, the topic of teeth -- of shape related to function -- was chosen. Through consultation and library search, we learned that among mammals four basic types of teeth are recognized; that they are related to these three functions: biting or nibbling, tearing, and grinding; that in many species the teeth are highly specialized, are related to the habitat and to the preferred food or to a non-eating function; that from even a single tooth specimen a mammologist can deduce considerable information about the habits and environment of the animal.

The topic seemed rich in possibilities for design, for use of multiple media and for expansion into the second-exhibit stage of the project.

Some of the developmental prospects that appealed to us were:

a) children could interact with real materials by touching and comparing real teeth, jaw bones, skulls and mounted animals, all of which were relatively durable, replaceable and inexpensive,

b) filmed close-ups of animals eating in natural habitat might be used, and even the use of live animals might be arranged,

c) tool analogies for each kind of tooth might be made in such a way that a child could operate on different kinds of materials to experience the effects of different tooth structures,

d) topics about animals are appealing to children,

e) because of his own "falling" teeth, the topic is within the child's experience and preoccupation; a clearer picture of his own dental equipment might emerge through activities with mirrors and constructing plastic models.

With this great decision behind us, we began to chart the subject-matter area in detail. These charts showed the characteristics of different types of teeth (number per individual, number of sets per individual, relative size, place in mouth, uses) for each of the major dietary groupings of mammals (herbivores, carnivores, omnivores, etc.). These displays gave a clearer appreciation of the options open to us in choosing clusters of concepts for the first exhibit.

Studies of Learner Characteristics

As we said earlier, the communication model emphasizes that effective message-sending must take into account the characteristics of the receiver. There is obviously no point in sending an elaborate message on a wave length that the receiver is not equipped to receive. It is no less obvious that museum exhibits must be attuned to the characteristics of their target population.

During the early days of the project, in order to find out some of the attributes of our visitors we did two things: 1) we asked those on the Museum staff who were experienced in working with children, and

2) we ran a series of tryouts with dozens of visiting children to find out what they already knew about animals' teeth and how they reacted to the kinds of specimens and techniques that we considered adopting for the exhibit.

In general our visitors came from all economic levels of Metropolitan Boston. Three subgroups would provide the main "customers" for our exhibit: the neighborhood regular, the week-end family group, and the special-event visitors who come either with family members or with social groups (scouts, settlement houses, etc.) for the Museum's special programs during school holiday periods. As with most museum visitors they come to be "entertained" or stimulated by strange objects or activities. Wide differences in reading and number vocabulary skills could be anticipated; attention spans, it was estimated, might be expected to extend from three to five minutes at best (in spite of the fact that estimates for adults traditionally average about thirty seconds per display). Equipment would be subjected to vigorous usage; it should be placed at levels appropriate for the average child's size.

For the exploratory tryouts of how children reacted to tooth-topic materials, a stock of artifacts and specimens was assembled: models of human jaws with detachable teeth; skulls of deer, monkey, bear, woodchucks; separate samples of various teeth; tools that could perform tearing, slicing, grinding functions; samples of foods of differing consistency; and mirrors.

Thus armed, the staff went forth into the Museum to discuss with visiting children the topic of teeth, to explore how much they knew, and to find out how they reacted to different approaches.

We learned that:

a) these children generally knew that teeth differ but they could not verbalize the differences,

b) a few knew that back teeth are "molars," and that there are "eye" teeth below the eyes, but they could not relate either shape differences or function differences to these names,

c) the children were at various stages of losing their own teeth, so that we had to abandon the notion that interaction with the "responsive environment" of food would show how they instinctively used different teeth for different purposes,

d) introducing the subject of teeth through animals was more attractive than by way of their own teeth.

These tryouts showed that the children had only a vague knowledge about the suggested topic. They seemed interested in it and were eager to examine the skulls. Their interest was easily sustained for the three-to-five minute period we had estimated. We reminded ourselves however that these children always seem to relish the staff's attention and that their interest in the topic would not necessarily be as strong when tape recorders and colored lights replaced the staff.

The outcomes of the learner-characteristics studies were these:

- the topic seemed to be a viable one,
- the exhibit should be designed so that reading skills would not be needed,
- verbal material should be aimed at about the level of third-grade vocabulary,
- the exhibit should be built with child-proof mechanisms,
- it should be built at levels compatible with the average size of the age group,
- it should require about three to five minutes.

Objectives and Tryouts Interact

Behavioral objectives were tentatively formulated; ways of utilizing various media to accomplish them were drafted. These were all phrased in terms of the specific things a child would be able to do after visiting the exhibit that he could not do before. For example, if the child understands the concept that incisors have a distinctive shape that is independent of their size, he would be able to do the following: when asked to pick "incisors" out of a box of single teeth of all kinds, he could do this successfully over a range of teeth from the tiny incisors of a mouse to the much larger ones of the beaver.

The first version of the exhibit objectives contained two sections, one dealing with the basic shapes of animal teeth and the second with the function performed by each type, and the action or motion required.

The distinctive shapes of four types of animal teeth were to be experienced by the child through an activity of sorting a stock of real teeth

into piles to match the four key samples; names were to be linked to the shapes by means of a recorded script. Next, the distinctive function of each tooth shape was to be experienced by the child's activating large-scale models of an upper and lower tooth of each type. By feeding ersatz food to these models the child could observe for himself how the different shapes of teeth function in different ways to produce different effects on materials.

One of our persistent worries in those early days was how to say anything significant in the short time period available to us. Consequently we fell into the error of attempting too much. Tryouts of our first version quickly revealed that the material would have to be broken down into much smaller steps with more opportunities for the children to practice the discriminations. We also had the common experience of finding that the names were the most difficult to teach. Although we do not attach importance to the learning of a special vocabulary in this instance, the learning of the names would facilitate our communicating with the child about the teeth and would afford him useful terms for sharing the experience with others after leaving the exhibit.

Procedures and script were revised, new sequences were tried out. In testing the utility of different procedures, we adopted the strategy standard in the programed instruction field: namely, the materials were tried with small samples of half a dozen or so children from the visitor population. Approximately eight different ways of dealing with the subject matter were explored. The information gained from these tests guided the further revisions. During this period, forty-six children between the ages of five and fifteen were tested; the majority (27) were in the 8-12 year range for which the exhibit was primarily intended.

Although some of the five-year olds were able to do some of the activities with the constant help and encouragement of the attendant, the subject matter was too difficult for them and they tired easily. It was doubtful that they could get much out of it later on when tape recordings were to replace the person.

Flexible Construction Materials

As behavioral objectives increased in clarity, the production process gathered speed. With our emphasis on child-partnered tryout-and-revision cycles, we decided to experiment with inexpensive construction materials in modular designs that would give the flexibility our development philosophy required.

The modular concept was carried out in the construction of the exhibit room walls and partitions. It was applied to the construction of "activity stations." By using multiples of a basic spatial unit for stations as well as for walls, we could revise space allotments or the location of stations without expensive or time-consuming structural modification.

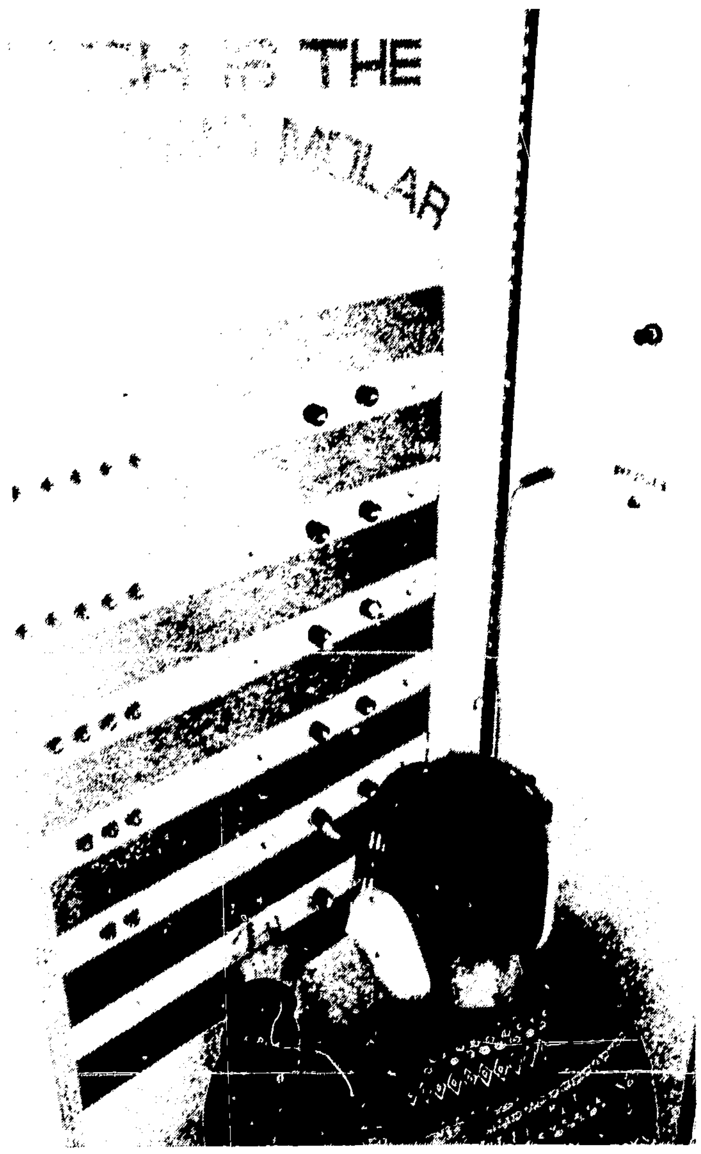
The construction materials were: a) Dexion slotted angle beams which can be bolted together like a giant Erector set, and b) 4' x 8' sheet material, called "Tri-Wall Pak," made of three thicknesses of corrugated board cemented together to form a panel which could be fireproofed, painted, bolted or nailed.[1]

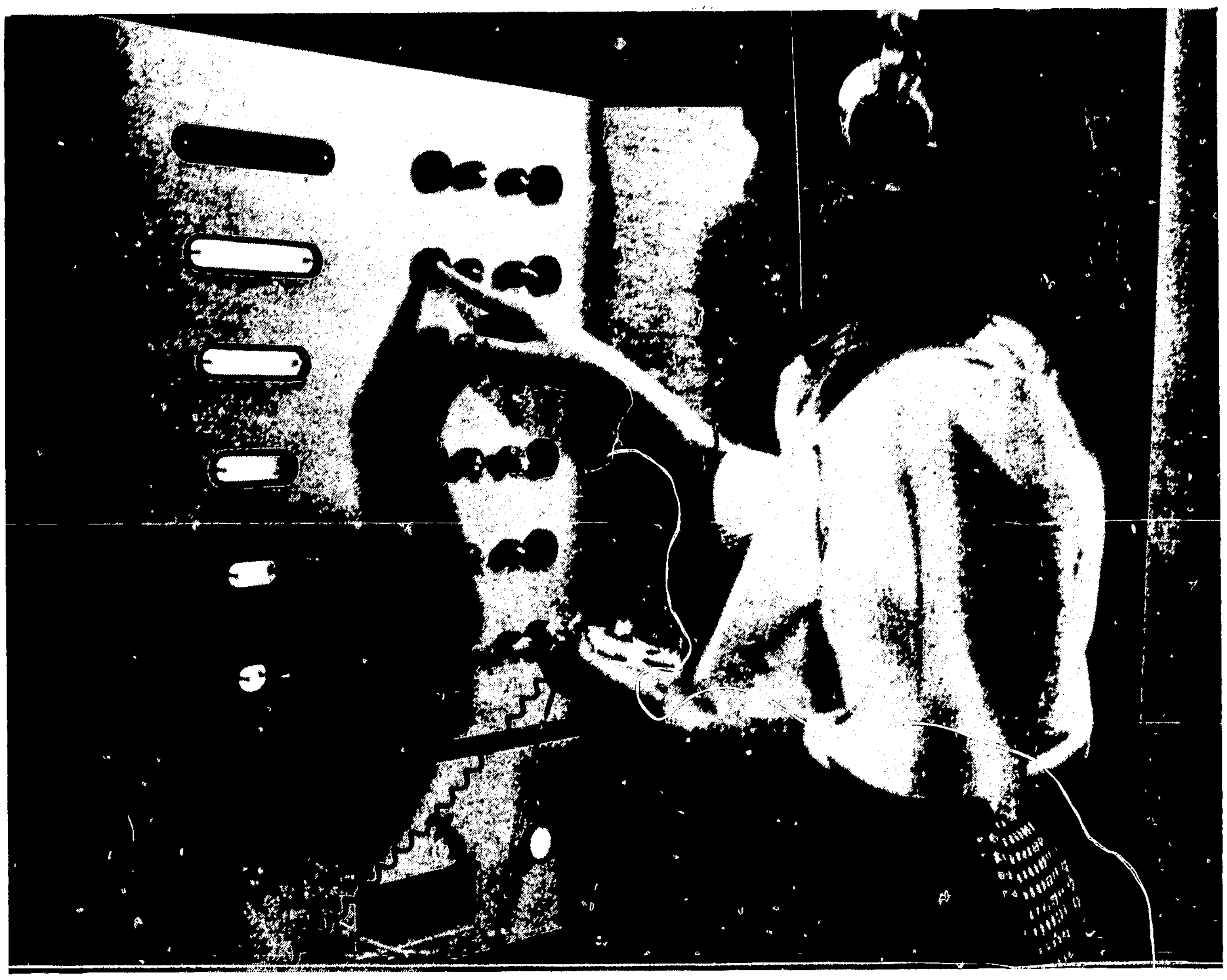
A series of seven instructional stations with automatic recording and response devices was opened to visitors in early August 1967; over a four-month operating period, these were modified and expanded. Their appearance at the time of the first major formal test period is the subject of the next chapter.

Summary

Potential subject matters were weighed against twelve selection criteria with the result that the topic of teeth, their shapes and functions, was chosen for the two related exhibits. Immediate try-outs with child visitors established the extent of their knowledge of and interest in the topic. Then, through children interacting with artifacts, skulls, and models, certain approaches to the topic were evolved, objectives refined and construction started. By August 1967, seven instructional activity stations were in operation.

[1] The Tri-Wall Pak sheets cost about $2 each originally and the fire-proofing treatment adds about $2 apiece more. In places where the fireproofing was applied unevenly, there was a slight problem with paint not adhering. We found that the problem could be solved by either giving the surface a quick washdown first with a damp cloth or later touching up the spotty areas with more paint. After an under-coat of water-base paint, the surface takes any kind of paint well (but local fire regulations should be consulted).


THE



Chapter 5

DEVELOPING THE SHEARING MOLAR SERIES: DESCRIPTION AND FIRST REVISIONS

Preview of the Completed Exhibit Configuration

The project proposal had called for the development initially of a "single concept exhibit that teaches a relatively simple concept or concept cluster through visitor interaction with a walk-in, stand-up or sit-down display." Through several cycles of tryouts of sequences and materials, we repeatedly encountered a problem familiar to teachers and programers -- the need to expand sequences and to add more response opportunities in order to achieve the stated behavioral objectives. As this exhibit threatened to engulf the entire Museum, we narrowed our sights and cut the subject matter to more manageable proportions -- several times.

The first exhibit, instead of being a single display as originally conceived, emerged at last in seven instructional stations -- walk-in ones, stand-up ones and sit-down ones. Furthermore the programed area, where station-to-station sequence was controlled, was supplemented by an equally extensive free-access area where four different units showed some of the curious variety of tooth forms.

While the exhibit stations grew more numerous, the topic contracted. The first series was limited to the shearing molars of carnivorous animals. Later it was combined with a second series of stations about grinding molars in order to bring out the contrast between these two distinctive types of mammalian molars. The final exhibit configuration also included activity stations about canines and incisors.

The formal evaluation periods for the two major configurations are convenient landmarks in the chronology of the project: evaluation of the shearing molar series in December, 1967 and January, 1968, and the evaluation of the combined shearing and grinding molar series in February, 1968. In later chapters discussion of the results will be organized in terms of these two phases.

In addition, some supplementary tests were made of the combined complex in April, 1968 after the free-access area had been rebuilt to include incisor and canine activity stations.

Table 1 summarizes the basic details for these three testing cycles, including the number of subjects concerned. Within each testing phase several different kinds of studies were usually carried out.

TABLE 1

DETAILS FOR THE THREE TESTING CYCLES

I SHEARING MOLAR SERIES: DECEMBER 1967 - JANUARY 1968

Programed area:	seven stations
Free-access area:	four units
Pretests-Posttests:	children
Preliminary Series:	36
Main Series:	83
Special Studies:	145
Observation Records:	
Free-access area:	48
Programed area:	88

II SHEARING AND GRINDING MOLARS COMBINED: FEBRUARY 1968

Programed area:	ten stations
Free-access area:	four units
Pretests-Posttests:	children
Main Series:	86
Special Studies:	33
Observation Records:	
Free-access area:	31
Programed area:	67

III SUPPLEMENTARY TESTS: APRIL 1968

Programed area:	ten stations
Free-access area:	four units (incisors, canines)
Pretests-Posttests:	children
Main Series:	20
Special Studies:	19
Observation Records:	
Free-access area:	17
Programed area:	0

By way of epilog we will describe how the last configuration of try-out stations was translated into permanent materials, how a duplicate set of stations was made and both were installed in a new Museum facility which opened on October 19, 1968.

The present chapter will tell how the stations of the first series worked and what we learned about them before the formal validation period began.

Objectives of the Shearing Molar Series

The stations about shearing molars dealt with:

- their name
- their location (farthest back in the mouth)
- their shapes (with fairly sharp edges and having one, two or three points) and the fact that shape is independent of size (different sized teeth in different sized animals but basic shapes are similar)
- their motion, which is to slice by one another in the manner of scissors' blades
- the fact that they are found in tigers and other cat-family members but not in children.

While these topics were worked into activities at the various stations, as will be seen shortly, our time limitations made it unwise to test for knowledge of all these items. Therefore, as evidence of the exhibit's effect we designed a performance test to deal only with the shape, motion, and name aspects; other items of the information were not evaluated.

The formal statement of behavioral objectives was therefore limited to those actions the child would be tested for at the end of the exhibit visit. As is customary, these objectives are phrased in terms of the specific situations in which he is to be tested:

When asked to pick out teeth that are "shearing molars," the child will be able to select real shearers of different sizes and shapes from a stock of single teeth including non-shearers of the canine, incisor and grinding molar classes.

When presented with a correct and an incorrect movable model of shearing molars, he will be able to point to the one that shows "how shearing molars work."

The pre-exhibit and post-exhibit test materials that incorporate these objectives were in the nature of a game, as a later section will show.[1] The objectives were also translated into the activities of the various stations; these and the other areas of the exhibit will be described next.

Description of the Programed Area

General

Appearance. The exhibit space was about equally divided between the programed area and the free-access area. The exhibit area floor was covered by bright orange carpeting, the walls and instructional stations were off-white. Photo murals were on the walls. The general lighting was subdued while spotlighting was used to point up the important components of the displays. Overhead, some stations were topped by colorful marquees. The work areas were tailored to the child's size; some were low enough to enable the child to sit on the floor. Others were arranged to bring the significant element of the display to the child's eye level.

The floor plan of the entire area is shown in Figure 1.

Audio Equipment. The early tryouts had confirmed the general suspicion that labels would not be read. Audio equipment was installed to guide the child's attention to materials, relations or distinctions and to tell him how to work the devices. The tape recorded messages came through telephones at two stations and through headsets at five stations. At the headset stations the child started the recordings by pushing a button; at telephone stations, lifting the receiver activated the playback units. At all stations the child could repeat the message at will.

The headsets proved to have several advantages: youngsters found them intriguing to wear; they freed the child's hands for other tasks; they somehow gave the wearer a claim over the station space ("territorial imperative") so that he was less likely to be intruded upon by another visitor.

---

[1] Although it is convenient to write in terms of testing the child, we were always conscious of the fact that it was the exhibit that was being tested; this point was often made with the children by way of explaining our "games."

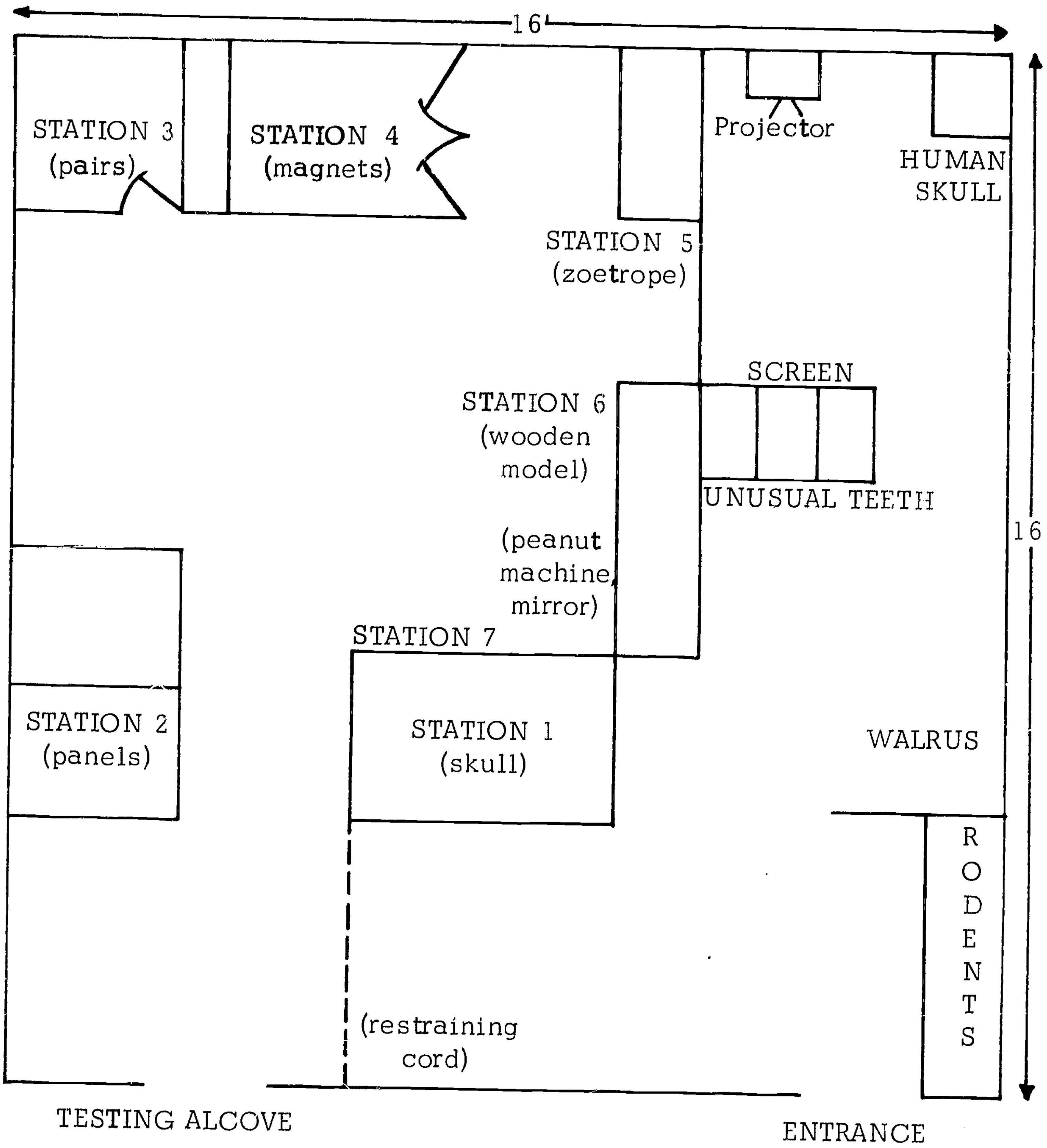


FIGURE 1. Floor plan of the shearing molar series at the time of evaluation in December, 1967 and January, 1968.

Automatic Devices. The stations operated automatically to provide the child with immediate feedback, and all were reset automatically for the next visitor. The equipment will be described in the separate account of each station.

Tooth Specimens. The programed stations required a large stock of single teeth. Many were obtained from the fish and game departments of western states and Maine. Other gifts and purchases filled out the collection. Some of the unusual teeth came from the Museum's own collections.

Within the programed area, all real teeth were embedded in bases of black casting resin so that their roots did not show. The reason for this was that in tryouts the presence of the roots had led to considerable confusion; we decided that when we were making a point about the shape of the tooth above the gum line, we would dispense with the distraction of the roots for the time being. The root topic could be taken up separately later in the free-access area.

Station One

Physical Description: a real tiger skull was mounted on a table top; the jaws were opened and closed automatically by a hidden motor, which included a mechanism to protect wayward fingers from being hurt; overhead, a marquee-type structure supporting two telephone handsets on its outer surface and a giant photomural of a tiger's face on the inner surface.

Objective: to introduce the topic of the teeth in the back of the mouth and to give an overview of the rest of the exhibit.

Activities: listen to recorded message, watch, touch.

Station Two

Physical Description: six lighted pushpanels, each about 4" square, were mounted in a row on horizontal workspace; each panel showed a stylized drawing of either a shearer or a nonshearer. Orange sample block holding six real shearing molars; yellow sample block holding six nonshearing teeth. Headset, start button.

Objective: to call attention to shape of tiger's back teeth, link name "shearing molar" to shape description; contrast this with those of non-examples.

Activities: listen to recorded message; handle and compare real teeth of the shearing and nonshearing types; push panels that match shearing shapes; get feedback from buzzer sound when shearing panels are pushed, no sound from nonshearer panels.

Station Three

Physical Description: enclosed booth with swinging 'barroom' doors; microswitch under floor panel activated mechanism controlling lights and resetting equipment when child left booth.

Vertical panel displayed six pairs of real teeth, one member of each pair being a shearing molar. Each pair mounted in a separate row up the panel; pushbutton beside each tooth. Teeth were attached by weighted wires so that they could be pulled out about 12 inches for closer examination and comparison. On the level with each pair were amber feedback lights -- one light with the lowest pair, two with the second, and so on, up to six lights with the top pair. Pushbuttons connected to a remote bank of counters for recording right and wrong responses for each pair.

Orange sample block like that in Station 2. Headset, start button.

Objective: practice discrimination of shearer shape by selecting the shearing molar from pair. Repetition of name-shape link.

Activities: listen to recorded message; handle real teeth, compare and search for match to sample, push buttons to register choice; get feedback in form of light for correct choice, no light for incorrect (lights had to be turned on in order from bottom to top, thus insuring that every trial ended with a success).

Station Four

Physical Description: enclosed booth with swinging 'barroom' door; microswitch under floor panel activated equipment and reset station when child left.

Six clear plastic tubes, each mounted vertically on a display panel behind which were inset six electromagnets. On the panel beneath the tubes was a container holding a dozen assorted teeth, each of which was mounted in a black plastic base; six teeth were shearing molars varying in size and shape (i.e., one, two or three points); these had steel plugs concealed in the bases. Six nonshearers had non-magnetic weights in their bases.

The task was to select shearers, insert them into the plastic tubes; shearers were held by the magnets, nonshearers fell through. A "finish" button allowed the child to release the shearers from the magnets to prepare for the next visitor. If he did not do this, the mechanism reset automatically when he stepped off the floor panel.

Beneath the mouth of each tube, narrow metal "fingers" were connected to microswitches and activated an electric counter each time a tooth fell through except during the reset process; this gave a running record of the number of nonshearers inserted in the tubes.

Orange sample block for reference. Headset, start button.

Objective: to practice discrimination by selecting shearing molars in a range of sizes and shapes from a stock of assorted teeth.

Activities: listen to recorded message; examine teeth, select, insert choices in tubes; get feedback from fact that shearing molars stick in tubes, nonshearers fall through; continue selecting and testing until all six instances found; by pushing "finish" button, may watch release of shearers to the container in readiness for next visitor.

Station Five

Physical Description: a zoetrope, an old fashioned animation device, was mounted behind a wall panel in such a way that it could be operated by spinning a wheel projecting through a wall slot; inside the device were a series of still images of shearing molars that appeared to move in the characteristic action when the wheel was spun.

Headset, start button.

Objective: to show distinctive action of the shearing teeth in motion, to compare with action of child's own back teeth.

Activities: listen to recorded message; spin the wheel, look through the peephole, feel how own back teeth work.

Station Six

Physical Description: a three-dimensional wooden model of two opposable shearers, each about six inches high, mounted in a recess in the wall; top tooth stationary, lower one could be moved by attached handle (equipped with safety device to prevent finger-chopping).

Peanut machine, operated with provided coins. Mirror. Headset, start button.

Objective: to experience shearing action with oversize model teeth, then to compare with action of own back teeth on peanuts.

Activities: listen to recorded message; move handle to work model, watch teeth slip by each other in action; work peanut machine, eat peanut, watch action of own back teeth in mirror, compare with action of model.

Station Seven

Physical Description: tiger skull revisited -- telephone handset mounted on side of surface displaying the tiger at Station 1. No other equipment.

Objective: to conclude the exhibit message by referring to shearing molars in their real context in the mouth of the tiger and other cat-family members; it contrasted the crushing action of the child's own teeth on the peanut with the slicing action of the moving jaw of the tiger.

Activity: listen to message, feel tiger teeth working.

Description of Free-Access Area

The purpose of this area, just inside the entrance, was to introduce the topic of animal teeth, to arouse the curiosity of the visitor, and to give him a taste of what the programed part was about. From this area he could look across into the other sector and get an idea of the kinds of things he might do if he chose to enter.

At the time the shearing molar stations were formally tested in December, 1967, four units were in the free-access area:

- Rodent skulls. On a black panel were mounted five rodent skulls arranged in order from large to small: beaver, woodchuck, squirrel, rat and mouse. Above each was a button which, when pushed, caused a life-size silhouette of the animal to light up around the skull. The name also appeared.

- Walrus skull. Against one wall, the skull of a walrus was attached where its enormous tusks could be examined and touched. The actual size of the whole animal was indicated by a silhouette on the wall.

- Human skull. In a darkened corner, a life-like model of a human skull was spotlighted. It could be touched and taken apart.

- Unusual teeth. An opaque case had three parallel doors, each 52 inches long but only ten inches wide. Light shone out around the doors inviting further exploration. When one door was opened, one found the saw from a real sawfish; on the inside of the door were photographs of sawfishes. A second door hid the curious spiral 'tusk' of the narwhal; photos lined the inside of the door here too. The third door concealed a very heavy elephant tusk as well as elephant photographs.

Early Modifications

The time from the opening of the first series of stations in early August up to the start of formal tests in December, 1967, was a period of important changes. Even prior to the opening some developmental testing had been possible as stations neared completion; afterwards, we observed and interviewed some of the hundreds of children who made the circuit. The resulting changes that are of general interest will be noted briefly.

Instructional Changes

In the earliest tryouts we had learned that the need for redundancy in the instructional materials was greater than we had imagined. It is well established in learning research that repetition is needed for most learners; that was why several stations dealt with the same concept illustrated in different ways. Tryouts of the automated series suggested that the redundancy needed to be increased for some children especially in the early part of the circuit. Station 2 went through several revisions for this purpose; originally it had only directed attention to differences in samples of shearers and nonshearers. Some children remained here twice as long as the average and, if questioned on leaving, were quite unsure of the message. The script was adjusted to point up the contrast, but this was only partly successful. Finally the pushpanel setup described above was installed to add another chance for overt responding. This has worked out well.

We also learned that many children did not generalize to the extent we had anticipated; they interpreted an instruction to match teeth and drawings very literally. To correct this the messages were adjusted and more examples were added to show the range of shapes that can be called shearing molars.

We also learned that recorded messages have to be short to be effective and that they should call for action from the child early in the sequence. Younger children especially became inattentive when the directions part of the message took up more than fifteen or twenty seconds before involving the child in activity.

The onset of the message, we found, needs to be under the child's control; if he is not focussed to listen, he misses the first phrases. Thus playback units need to be activated by pushbutton rather than by the lifting of a headset from its hook.

Traffic Control

The problem of controlling the flow of visitors through a programed exhibit turned up early as a major problem. In the first place we had to overcome the natural tendency of museum visitors to wander about in any direction. Initially we expected that the recorded message at the strategically placed Station 1 would make the point that the order of visiting stations was important.

In practice, however, large crowds poured into the area and Station 1 was lost to view. New arrivals skirted the group to reach the less congested areas beyond and went about visiting the least crowded stations.

In response to this problem, just beyond Station 1 a restraining cord was installed so that an attendant could control the crowd and prevent waiting lines from forming within the programed area itself. Although this proved effective, we are still seeking an inexpensive method that will let the exhibit operate without an attendant.

The first few days operation of the exhibit convinced us that parents would have to be excluded. Apart from their adding to the congestion, they participated so wholeheartedly in the exhibit that it was impossible to judge how much interest the exhibit held for the children. Parents and preschoolers were thereafter restrained behind the cord. At times a mimeographed handout was available to explain to parents

why we needed the children's unassisted responses to the materials.

Originally the proper path from station to station had been indicated by large numbers about 3 feet square. These were so large the children did not notice them. This problem was solved by smaller signs, 1 foot square, painted in a bright orange and placed lower on the walls.

Stations 3 and 4 with their interesting gadgetry activated by floor panels presented special problems in traffic control: several children would try to crowd in together and then take turns. But the circuitry does not re-set until all weight is momentarily off the floor. This problem was alleviated by installing small swinging half-doors in each booth so that it was difficult for more than one person to enter.

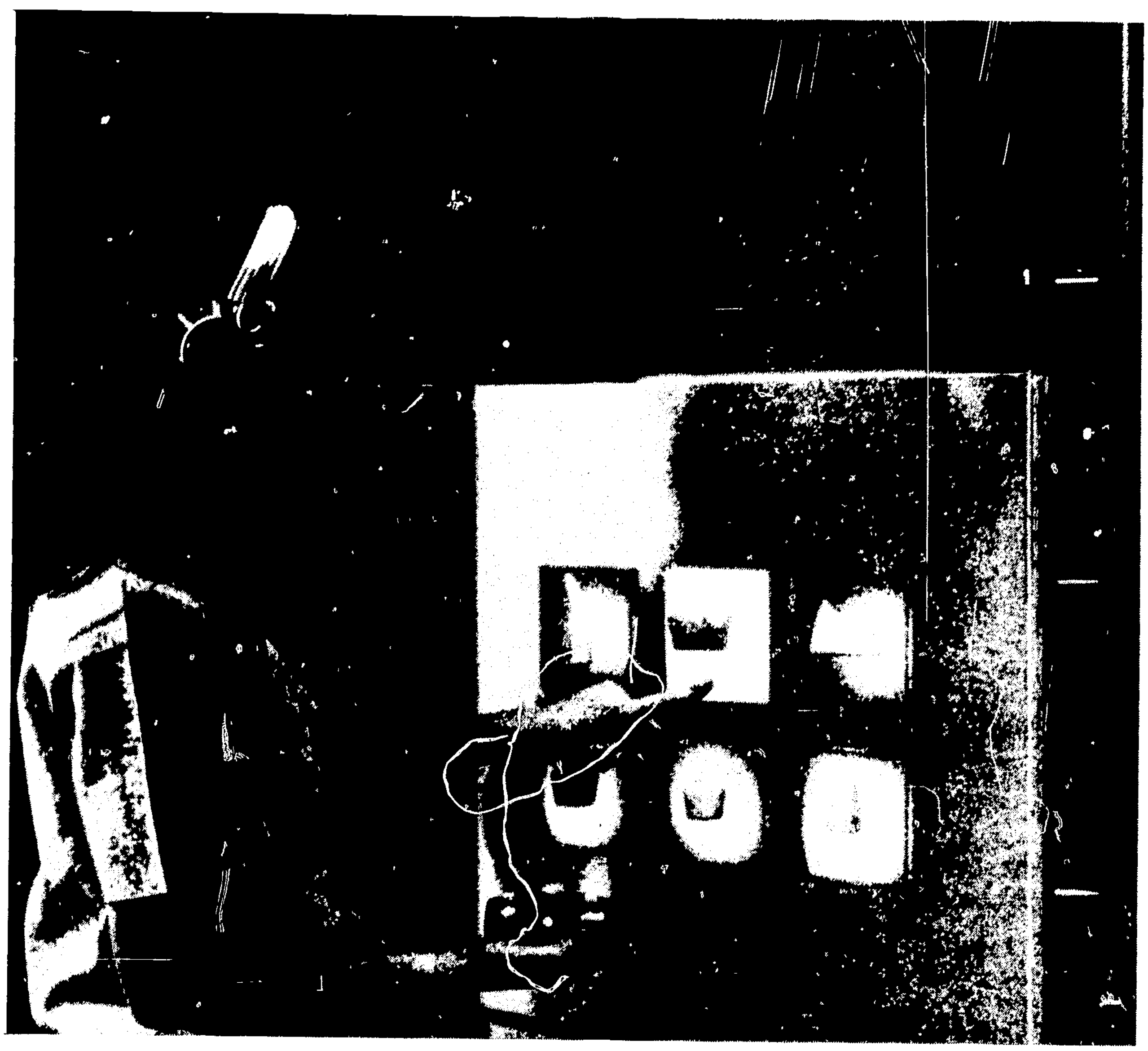
Other Physical Problems

Two planned activities were abandoned because of physical or logistic problems. An X-ray motion picture of a person eating had been part of the human skull unit in the free-access area; it was finally discarded because we lost patience with its constant jamming as a result of poor splicing.

The second activity that did not survive was one in which the child was to feed ersatz food to the oversized wooden teeth in order to see just how they worked. Problems of finding suitable inexpensive material, of dis-pensing it to the child, and of disposing of the debris occupied us for many weeks before we decided we could not afford more time on them.

Summary

The varied activities of the first series of instructional stations show how the shearing molars of meat-eating animals are characterized by distinctive shapes and a specific type of motion. Seven automated stations which had to be followed in order were supplemented by a free-access area where four separate units showed some of the curious variety of teeth. Changes were made to improve the quality of the instructional message and to correct aspects of the physical operation of the system.



Chapter 6

EVALUATING THE SHEARING MOLAR SERIES: METHODS AND RESULTS

After four months of operation in the Museum, the shearing molar series went through a month-long period of formal evaluation. During this time systematic observations of the children's behavior were also recorded.

Methods

Performance Tests

Testboards. To determine whether the exhibit visitor could distinguish shearing molars from other kinds of teeth, six real shearers differing in size and in number of points, were mounted on a testboard along with six nonshearers. Each tooth was embedded up to the gum line in the usual black plastic base.

The teeth were mounted in a 6 x 2 array in a random pattern.[1] Beside each tooth were two holes, labeled "Yes" and "No." (See Figure 2.) Paper inserted into the testboard could be marked through these holes. A rectangular opening near the bottom of the board left room for the child's name and age, the date and initials of the examiner.

Two such testboards were used for the first series evaluation period. On each testing day one board was used for pretest, the other for the posttest. The next day the boards were interchanged, and so on. Different colors of paper were used to distinguish pretest records from posttest records.

Tongs. The defining characteristic of shearing teeth action is that they slide past one another as do scissors' blades. To determine whether or not the exhibit made that clear, two tong-like devices of lucite were made to simulate the basic opening and closing action of a jaw. To the end of each 'arm' of the tongs a row of three real shearing molars was fastened. When the tongs were squeezed and released, the teeth moved together and apart as in a jaw. (See Figure 2.)

[1] The order of the shearers and nonshearers on the testboards was taken from a random numbers table with the following restrictions: that the patterns on the two testboards should be different and that on each half of each board there should be an equal number of shearers and non-shearers.

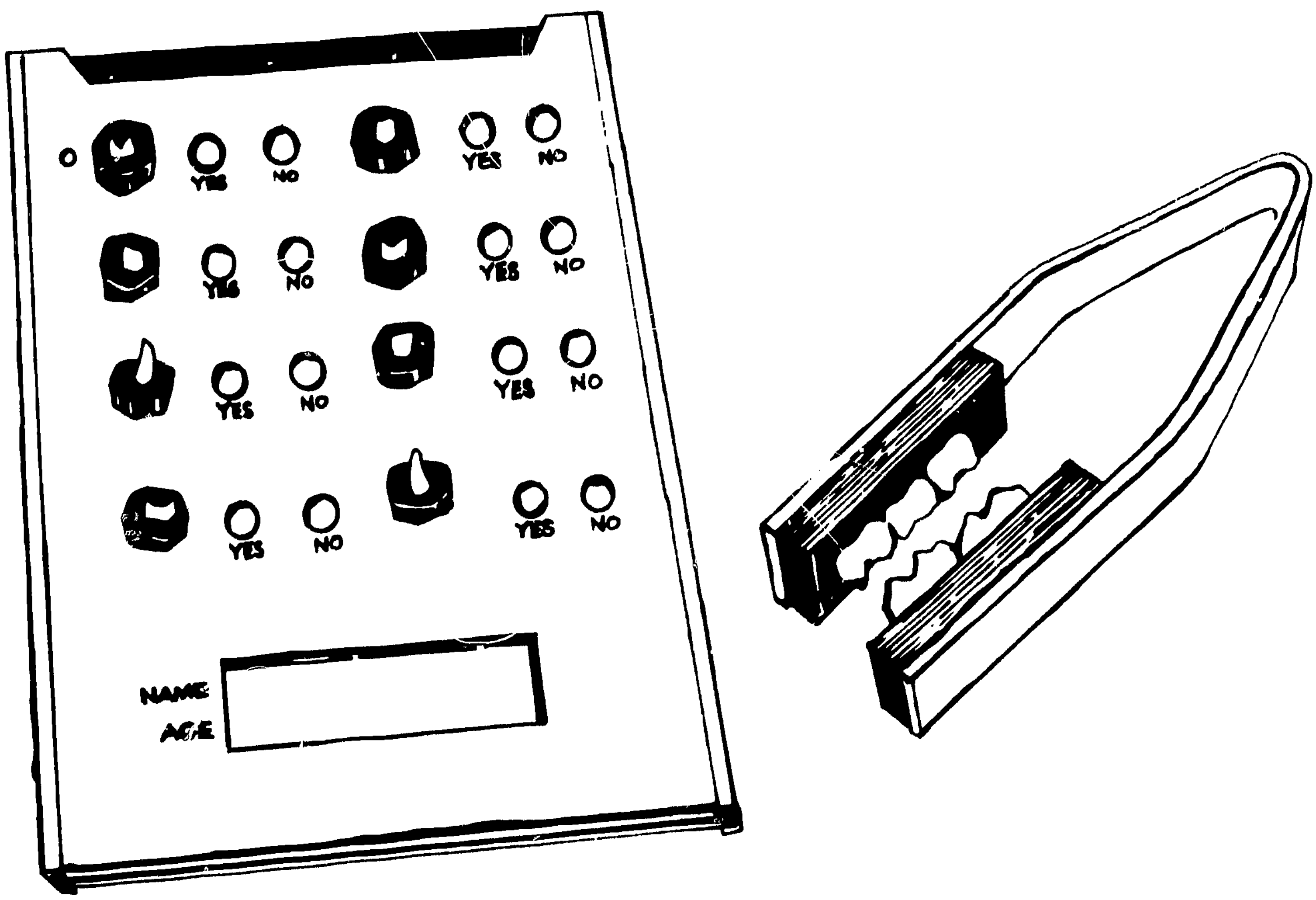


FIGURE 2. Examples of a testboard and a tong device. Eight-trial boards were used to test the combined shearer-grinder complex; similar twelve-trial boards were used in the earlier shearing molar series.

Testing Procedure

A small room just off the path from Station 1 to Station 2 was fitted with a table and stools. Lighted panels in the table permitted quick scoring of the testboards by revealing the correct answers through coded holes in the boards. As in exhibit stations, headsets were mounted here too for instructing the children about the 'guessing game' with the testboards.

Invariably a line of children was waiting to visit the programed area. All had listened to the message at Station 1 (with tiger skull) and many of them had visited other units in the free-access area.

When the staff was ready to begin testing, the first child in line (provided he was at least 7 years old and had not visited the exhibit before) was invited to "try a guessing game" in the testing alcove.

He was seated at the table and shown the testboard. Ordinarily he would put the headset on and listen to the recording. During the early part of this evaluation period the repeater was malfunctioning; therefore, the staff member read the instructions to the child in order to be sure that all children received the same message.

For the pretest the term "shearing molars" was not used since the child, even if he knew about such teeth, might not recognize that phrase. He was asked to consider each tooth carefully, to place a check in the "Yes" hole beside ones like teeth in the back of the tiger's mouth and to place a check in the "No" holes for those not like the tiger's back teeth. The instructions were repeated if the child wished.

When he had finished, the two sets of action tongs were held before him and were opened and closed by the examiner in as impartial a manner as he could. The child was asked (orally) to recall how the tiger's back teeth looked when the jaw was moving and to choose the set of tongs that worked the same way. The child's answer was recorded on his testboard paper.

Then he left to visit the programed stations. When he returned, the posttest followed closely the procedure used before, except that the instructions were all via tape recordings now. The term "shearing molars" was used in asking him to check the testboard this time. When he had completed this and had made his choice of the action tongs, he went on his way with our thanks.

At no time were the response sheets scored in the child's presence and no child was ever told anything about his performance on either pretest or posttest. We tried to minimize as much as possible the test-taking atmosphere, avoided the term "test" (although the children constantly used it in talking with each other outside).

After one child had finished the pretest and started through the exhibit, the next child in the waiting line was usually invited into the testing alcove; on some occasions however when a special study prescribed a certain quota for different age groups, the selection was made on such a basis.

Children were of course going through the exhibit in a steady stream, controlled by a staff member at the restraining cord at Station 1. The number admitted was regulated to prevent lines forming within the programed area. Only a small proportion going through the exhibit could be tested.

At the beginning and end of each testing day, readings of the electric counters for Station 3 were recorded on standard forms. These included the tallies for both tested subjects and the general visitor. (Counters data were not available for Station 4 as had been planned because the microswitch fingers had caused jamming of the teeth in the plastic tubes and were disconnected.)

Results

Pretest-Posttest Comparisons

Preliminary Series: In early December the evaluation crew embarked on a "shakedown cruise" by testing thirty-six children with the 'standard' procedure: a) pretest, b) visit the exhibit with all stations open, c) posttest. This differed from the procedure of the Main Series in that the tongs and one of the testboards were not yet available. Thus both pre- and posttests were made with the same board; memory of previous responses might have influenced the posttest answers.

During this period slight adjustments were also being made to the exhibit: a change of emphasis in a recorded message, installation of an improved version of the wooden models and so on. When these were finished, the setup was then held constant throughout the Main Series.

Let us look briefly, however, at the results of these preliminary tests. Since there were 6 shearers and 6 nonshearers on the testboard and the subjects could respond "Yes" or "No," 50% of the responses would be expected to be right by chance on the average. Our pretests averaged 51.6% correct, a figure which confirms our earlier information that children did not already know much about the topic.

On the posttests, however, the thirty-six children averaged 73.4% correct responses. Such a gain in score is very highly significant in the statistical sense.[2]

Very little difference was apparent in the gains of boys and girls:

	No. of Children	Ave. Age	Per Cent Correct Pretest	Per Cent Correct Posttest
Boys	25	9.32	52.0%	72.3%
Girls	11	9.36	50.8%	75.8%
Total	36	9.33	51.6%	73.4% (*t* = 5.88***)

When scores were segregated by ages and their averages plotted in Figure 3, we see a fairly even effect. Older age ranges may show a very slight edge in entering knowledge but the amount of gain shows no clear relationship to age.

---

[2]The gains in score from pretest to posttest were evaluated by a standard statistical procedure, the *t* test for paired deviates. The resulting statistic, *t*, enables the investigator to judge whether his obtained difference is within the range expected by chance or whether it is so far beyond this that he is justified in concluding his effect is "significant."

By convention, statisticians award 'stars' to *t*'s in the manner of the Guide Michelin:

one asterisk:	the *t* merits the term "significant," is a result that would occur on a chance basis about 5 times in 100;
two asterisks:	the *t* is called "highly significant," occurs by chance only about once in 100 times;
three asterisks:	the *t* is "very highly significant," occurs by chance only once in 1000 times.

We will follow the practice of using these terms and attaching the appropriate number of asterisks to the *t*'s cited.

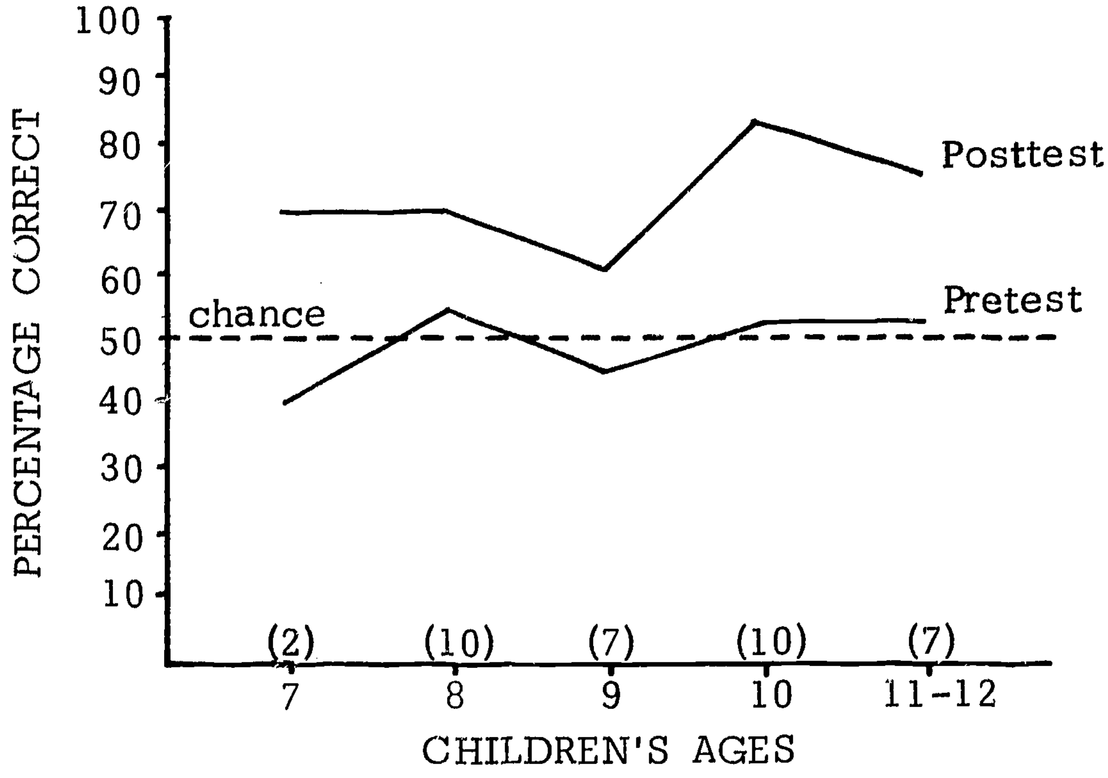


FIGURE 3. Preliminary Series: graph of pretest and posttest scores for pooled results of boys and girls in the shearing molar series, December, 1967. (The number of cases at each age level is given in parentheses just above each age.)

Main Series: With the evaluation crew now practiced and the final pieces of testing equipment on hand, the Main Series of formal tests to validate the shearing molar complex began in mid-December, 1967. The only differences now in procedure were: (1) the tongs choice-task was inserted just after the testboard task for both pretest and posttest, and (2) with two testboards available now, their use for pretest and posttest was alternated by days.

A total of 83 children between the ages of 7 and 12 years were tested; they averaged 9 years of age. This time we had more girls, 51 in all, and only 32 boys. The boys averaged about half a year older than the girls.

On the pretests, 57.2% of the answers were correct, a somewhat higher result than in the earlier series. This raised in our minds the possibility that the increase might be due to the presence of children who had visited the exhibit before. There was no way to tell, and as we mentioned earlier, we never found a feasible way to prevent their inclusion in test data.

For the post-exhibit task, the children were right on about 78.6% of their choices. Such a gain over the pretest average is very highly significant.

The results for boys and girls separately are these:

	No. of Children	Per Cent Correct Pretest	Per Cent Correct Posttest	t
Boys	32	59.4%	77.8%	4.12***
Girls	51	55.9%	79.8%	7.82***
Total	83	57.2%	78.6%	8.57***

All the t's being 3-star ones, it is well established that the exhibit stations are having a marked effect on our visitors' behavior.

The kind of analysis we have made so far reflects an increase in the number of successful trials, but it does not tell us specifically how many individual children were affected by the exhibit. We can get this by counting the number who improved, made the same score or a worse one afterwards. The result of such a count is:

Girls (51 cases)	80.4% Improved
Boys (32 cases)	68.8% Improved
Total (83 cases)	76 % Improved

By another statistic (chi square), the distribution of "better-worse" cases is very highly significant. This then serves to reinforce the earlier conclusion and reassures us about the generality of the effect.

On the pre- and posttests, the possible scores range from 0 to 12. It is interesting to see how many children made each score on the pretest and to compare that trend with the figures for the posttest scores. The graph in Figure 4 shows the two score distributions. There we note that the most frequent pretest score was 6 right out of the 12 trials, which is the score expected by chance alone. For the posttests, however, the most frequent score was the perfect one, 12 right out of 12 tries. A distribution shaped like that of the posttest scores commonly indicates that the test exerted a "ceiling" effect, preventing the better students from showing their competency. It also suggests that with a more sensitive test the difference between pre- and posttests would have been even more marked.

Next, in order to see how the exhibit affected children of different ages, we plotted the pre-exhibit and post-exhibit scores separately for each age group for both boys and girls. These are shown in Figure 5. For the girls, both pretests and posttests show a slight steady incline over ages, the difference between the two lines remaining virtually constant. This reflects the fact that at each succeeding age children know a little more on entering the exhibit (or are more skillful test-takers) but the amount they learn within the exhibit is about the same regardless of their age.

The picture of the boys' results is not quite so neat, perhaps because the numbers of cases at each age are smaller and extreme cases can exert a more noticeable effect. For 7 and 8 year olds, the exhibit effect is trivial. A closer look at the individuals in this age range shows that only 45% of them improved in scores after the exhibit visit, while for girls at the same ages, 80% improved. From age 9 up, the boy's posttest scores are above the 80% success level and are just slightly superior to the girls' scores. The trend at the younger age then is most likely attributable to random variations.

In general, both the Preliminary and Main Series show that the exhibit had a marked effect on the children's ability to perform the shape discrimination task. The graphs of the results suggest that the amount of improvement does not vary much over ages. The latter is a pleasing result for it indicates that the topic treatment in the exhibit did not apparently assume concepts or skills that are acquired with age.

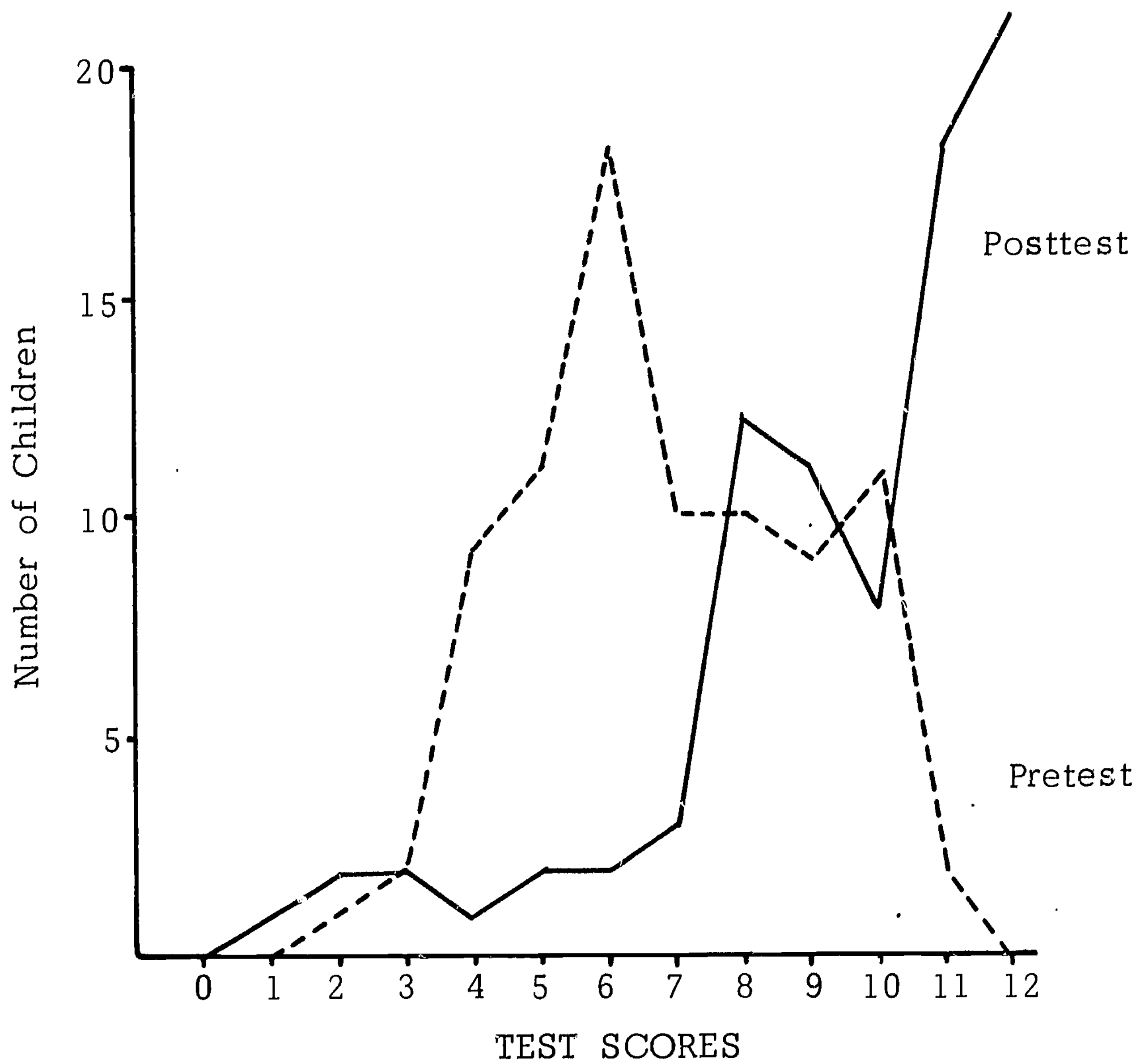


FIGURE 4. Main Series: graph showing the number of children who obtained the various scores on the testboards before and after visiting the exhibit. (Chance score is 6.)

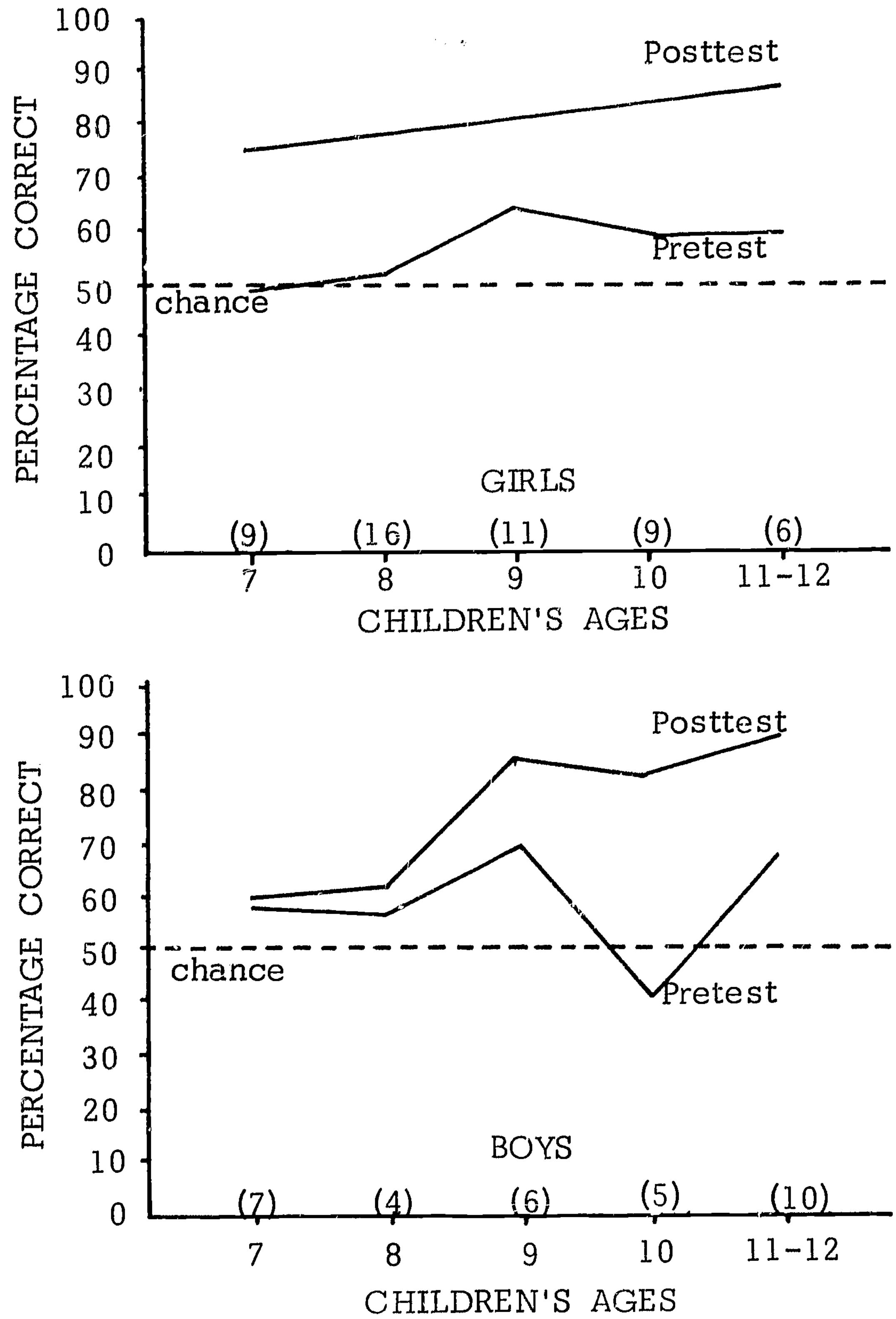


FIGURE 5. Main Series: graphs of pretest and posttest scores for boys and girls in the shearing molar series, December 1967 and January 1968. (The number of cases at each age level is given in parentheses just above the age.)

Tongs: The two tong-like devices, it will be recalled, show a correct and an incorrect version of the way shearing molars move by each other. It was a quick and inelegant interim method for finding out whether the children had noticed the distinctive operation of the shearers.

Of the subjects in the Main Series, 78 made the tongs choice before and after visiting the exhibit. (In no case were they told their result.) For the pre-exhibit choice 71.8% of them made the correct selection and 84.6% were correct afterwards. By chance alone one would expect 50% of the choices to be correct. The high rate of success on the pretest choices may have been due to the unconscious influence of the experimenter who was handling the tongs. A more attractive explanation would be that the dramatic automated tiger skull (beside which the children waited their turn) made the point of shearing most effectively. At any rate, the increase to 84.6% after visiting the exhibit is not an improvement significantly greater than might be expected from chance variations.

If we look into these results more closely, we find that 55 of the 78 children did not change their minds after going through the exhibit -- that is, they made the same choice both times (50 of the 55 were right both times). But the ones who did change their minds are interesting: 74% of these changed in the 'right' direction, and this is a significant effect. It does suggest that the exhibit made this point to those who had not already got it at Station 1.

Observations and Counter Records

During the shearing-molar validation period, systematic observations of child behavior in both the free-access and programed areas were recorded by means of a checklist. The nature of such a checklist may be seen by referring to the Appendix where a similar but longer record for the later combined series is reproduced.

The common procedure was for one staff member to observe in one area while another operated in the other area. To try to follow a given child through both areas was usually impractical because of the delay at the line waiting to enter the programed sector.

The information gained from these records was important for three reasons: (1) it gave us a general picture of how children respond in this kind of controlled educational situation, (2) it provided supplementary data to be cross-checked with test data, and (3) it showed us how the materials were being used and how the equipment was operating. These data had an immediate impact of course on our

revisions since they clearly pinpointed trouble spots or points of confusion. The details are too numerous and often too equipment-specific to report now, but the findings of more general interest will be discussed here.

In the free-access area, the four units (rodent series, walrus skull, human skull, and 3 large "teeth" specimens) were about equally popular. The rodents, just inside the entrance, were often missed until the child was on the way out. The transition point between free-access and programed areas, Station 1, was the most frequently visited spot in the 'un-controlled' area -- in fact the waiting line there signalled an important attraction and many children rushed to line up without first exploring the other units in the area.

Two aspects of the children's behavior interested us especially: one was the amount of time they spent at the exhibit units and the other was their 'choice behavior' -- panel-pushing, button-pushing and the like.

For the total time spent in the free-access area, we have records of 17 girls and 31 boys between the ages of 7 and 13. The boys spent 4.32 minutes and the girls, 3.32 minutes on the average. Some spent only a few seconds before lining up at Station 1; one stayed sixteen minutes and had his visit cut prematurely short by a general power failure. It is interesting to note in passing that the records show that fewer than 25% touched the openly mounted specimens.

For the programed area, time records are based on 41 girls and 48 boys above the age of 6. At the six programed stations (Stations 2 through 7, Station 1 having been counted in the free area), the boys averaged a little over 13 minutes; the girls spent a little over 12 minutes on the average. The shortest time was 5 1/2 minutes and the longest was 30 minutes. Stations 2, 3 and 4 generally required 2 1/2 to 3 minutes each; the other three took less than 2 minutes each. Younger children stayed longer, especially at the early stations.

If the average times spent in both areas are combined, we find that boys took about 17.5 minutes and the girls, 15.5 minutes in the total exhibit, overlooking the fact that the waiting-line time was often in excess of ten minutes. It was frequently noticed also that children who had finished the circuit of the programed area returned to the line to wait for another turn.

One of the findings that surprised us had to do with "button pushing. It had been generally thought that children would push every available

button indiscriminately. The several tasks requiring discrimination of shearers from nonshearers were planned as learning situations, not test setups, and we had expected that children would explore the effects of right and wrong choices alike. This expectation was not confirmed. Table 2 summarizes the results for the three stations where right and wrong choices could be observed.

The general picture is not one of uninhibited button-pushing. Many children made no incorrect response at all in these choice situations, and most others apparently tried to make only correct responses. The number who made "perfect scores" is especially high for Station 3 (Pairs), being 33% for boys and 48% for girls. The records also suggest that the boys were more venturesome in their approach. More of them pushed buttons or panels in an exploring fashion. The girls' responses are predominantly correct, perhaps reflecting more docile or more achievement-oriented natures.[3]

The counters' records for Station 3 bear out these observational data. The counters give a running record of right and wrong choices for each of the six pairs of teeth. During the validation period, the records show that from 650 to 700 responses were made for <u>each</u> pair (these data are from all visitors of course, not just those who were tested or observed). The percentages of correct choices varied from a low of 67.7 on the third pair to a high of 75.0 on first pair, a trend which again refutes the idea that children generally push all buttons. (Counters were disconnected from Station 4 because of equipment problems, so we have no supporting data for that station.)

---

[3]Wherever percentages of correct choices are cited for Station 3, they are based only on <u>first</u> choices for each pair. This is true whether the data are derived from observation records or from the electric counters. It may be recalled that the manner in which this station operated required a correct selection at each successive pair before the next pair could be tackled; thus if the first button pushed was wrong, the remaining button had to be pushed in order to turn on the correct-signal light. These second choices, which had to be right, are of no interest in our analyses and therefore do not enter into any percentages cited for this station. By chance alone, about 50% of the first choices are expected to be right.

TABLE 2

CHILDREN'S BEHAVIOR IN CHOICE TASKS,
SHEARING MOLAR STATIONS, DECEMBER PERIOD

STATION 2	(6 Panels: 3 shearers, 3 nonshearers)	
	46 Boys	38 Girls
Percentage of Children Who:		
Pushed Every Panel	47.8%	29.0%
Pushed Only Correct Panels	15.2%	29.0%
Made One Error or Less*	34.8%	52.6%
Percentage of Correct Panels Pushed:	58.7%	65.6%

STATION 3	(6 shearer-nonshearer pairs)	
	39 Boys	31 Girls
Percentage of Children Who:		
Pushed All Buttons	7.7%	.0%
Made Perfect 'Score'	33.3%	48.4%
Made One Error or Less*	53.9%	64.5%
Percentage of Correct First Choices:	69.2%	78.9%

STATION 4	(Magnets; 6 shearers, 6 nonshearers)	
	36 Boys	35 Girls
Percentage of Children Who:		
Put All Teeth in Tubes	30.6%	14.3%
Made Perfect 'Score'	19.4%	31.4%
Made One Error or Less*	41.7%	65.7%
Average Number of Trials to find the 6 shearers:	10.06	8.18

*Includes cases in row above.

The task at Station 4 was to find the 6 shearing molars from the stock of 12 teeth. Here it is interesting to ask how many trials it took the children to find all 6 shearers. As Table 2 shows, the girls averaged just slightly over 8 attempts to get the required six teeth, while the boys took an average of about 10 tries. If the objects were picked out at random without any knowledge entering into the choice, approximately 11 trials on the average would be needed to find the 6 shearers.[4]

The high number of tries for the boys is a reflection of the fact that over 30% of them put all of the teeth in the tubes, as noted in the table. And as a matter of fact, inspection of records reveals that these boys put the <u>same</u> wrong teeth in the tubes repeatedly; one boy, for instance, put the 6 nonshearers through the tubes a total of 17 times. Here again is the suggestion that boys were more exploring in their approach to the situation. At this station it was obviously fun to play with the equipment and try to figure out how it worked. The boys who did not do this averaged 8.08 trials to get the shearers, the same rate the girls showed.

A few observations about the audio equipment should be recorded. The children took to the headsets most enthusiastically; in fact for many, these were the most attractive part of the exhibit. One child when asked what she liked best about the Museum, answered

[4]The task of computing the probabilities of being right by chance for this station is complicated. For the first sample tested, the chances of being right are 6 out of 12, or 1/2. For the second choice, the probability will depend on whether the first choice was right or wrong. If it was right, then there are 5 shearers and 6 non-shearers left, and the chance probability of being right is now 5/11. But if the first choice was wrong, 6 shearers and 5 nonshearers are left, and the probability of picking a shearer at random is 6/11. For each successive choice, the probability depends on the outcomes of all the previous choices.

I was fortunate in having a statistician friend, Mr. Joel Kleinman, Cambridge, Mass., who sorted out this sticky problem for me. With the aid of two simplifying assumptions (that a wrong tooth will not be tried repeatedly and that a child will continue until all 6 teeth are found), he computed that if only chance were operating in the choices, it would require an average of 11.1 trials to locate the 6 shearers.

"the earmuffs that talk!" For the most part, the children were able to handle the headsets without help, except for a few who needed guidance at the first station. Here 21% of the boys and 25% of the girls repeated the recorded message, but for subsequent stations, this repeat capability was seldom used.

Special Studies

At times tests were made with one station excluded from the series. These tests were tried as time permitted in order to see if different stations contributed equally to the total learning experience. Originally considerable redundancy had been built into the series as a way of allowing for individual differences. When the plan to install the grinding molar series was adopted, the question arose whether the redundancy in the shearer series could be reduced without appreciable learning loss; if it could, we would gain much needed space for the grinding molar stations.

The omission of a station might have two effects: an adverse one due to the loss of information or practice, and a favorable one because the shortened interval between shape learning and posttest puts less strain on memory. The two might offset one another. In spite of the ambiguity in interpretation, we decided to run tests as time permitted to see whether any unexpectedly large effects might occur.

Primarily we experimented with Stations 3 and 4, both being practice of shape discrimination learned at Station 2. (The plan to omit one unit for test purposes was facilitated by the fact that each was closed for repairs at one time or another.)

The general trends, shown in the table below, suggest that the early station, Station 3, makes an important contribution, but that omitting 4 does not detract from shape-discrimination learning.

RESULTS WHEN ONE STATION OMITTED FROM SERIES

Station Omitted	Ave. Age	No. Cases	Per Cent Success Pretest	Posttest	Diff.	Per Cent Cases Improving
3	9.35	52	62.5	77.6	15.1	67.0
4	9.96	51	57.2	80.6	23.4	76.5
5	9.21	28	58.9	83.6	24.7	92.9
6	9.07	14	54.2	75.6	21.4	78.6

Since Stations 5 and 6 dealt with function, their omission would not be expected to detract and might even enhance the shape posttest score. No statistical tests were made on these data but the general picture is compatible with this expectation.

Younger children, however, seemed to be adversely affected by the omission of a station. If we look at the records of the 7 and 8 year olds for Stations 3, 4 and 5 (too few cases for Station 6), we find:

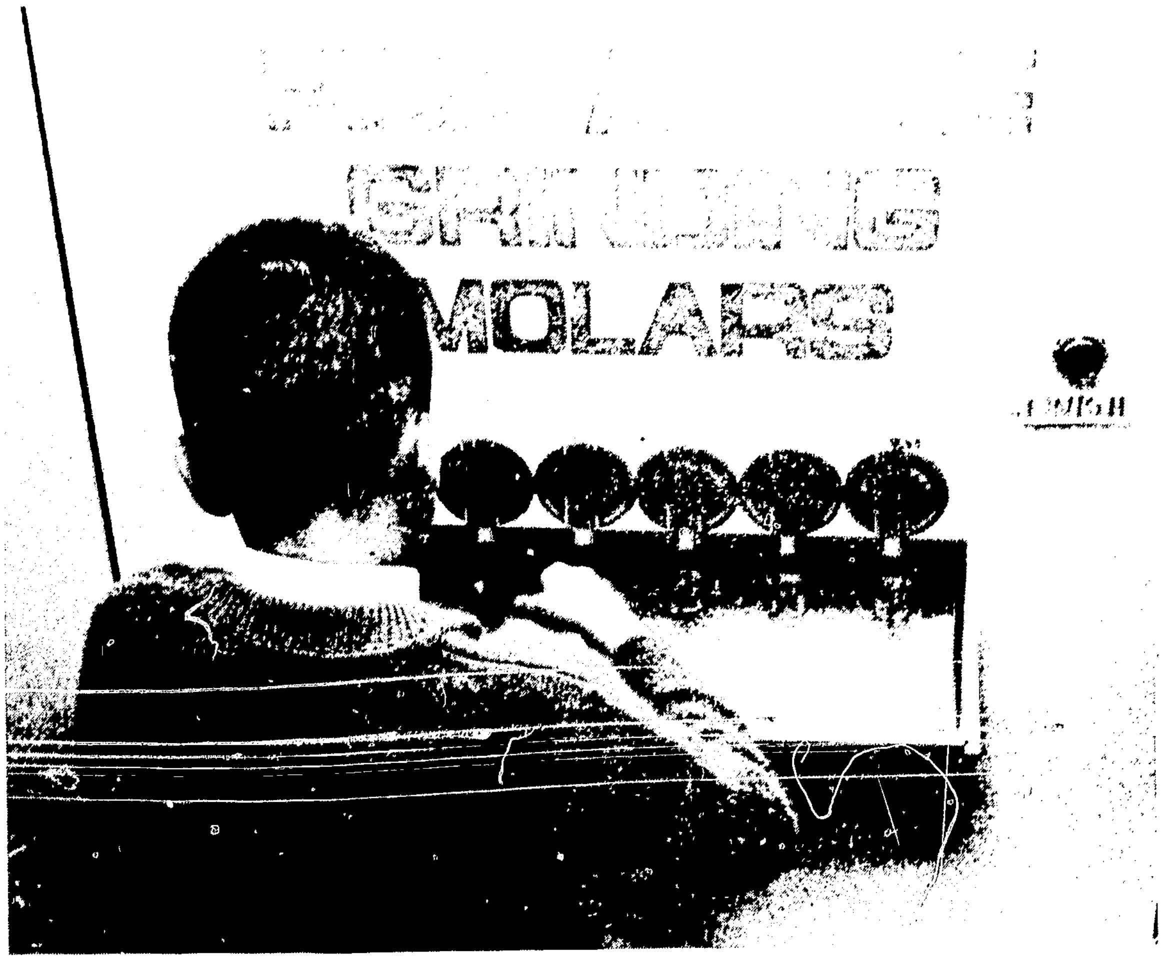
Station Omitted	No. of Children	Per Cent Success Pretest	Posttest	Difference
3	19	51.75	63.16	11.41
4	12	56.94	77.78	20.84
5	7	60.71	78.57	17.86

The gains are all smaller than those for the total sample. This fact suggests that the initial redundancy was serving a useful function for the younger visitors.

In spite of this, we decided that Station 4 (the magnets one) could probably be converted to a grinding molar station without seriously damaging the learning about shearing molars.

Summary

After visiting the seven instructional stations, children were significantly more successful at performing the discriminations taught by the exhibit than they had been beforehand. In all, 36 children in the Preliminary Series and 83 in the Main Series participated in these validation tests. Records of their behavior in the exhibit area are summarized; two main findings were that children averaged from 16 to 18 minutes in the exhibit (omitting waiting time) and that in choice activities, a large majority try to make only correct responses and do not engage in general "button-pushing" behavior.


MOLARS

Chapter 7

THE COMBINED SHEARER-GRINDER COMPLEX

Introduction

It is by the contrast between shearing molars and grinding molars that the exhibit gains significance as an introduction to adaptive relations of structure and function. These two kinds of teeth with their distinctively different shapes, movements and purposes afford the clearest clue to the life style of a given mammal. The shearers with their sharp edges and scissors-like action are well suited to cutting apart meat; the grinders with their flat, bumpy surfaces and circular motion work efficiently to crush and grind a mixture of meat and plant matter.

Through the exhibit the tiger serves as the representative of the cat family with their shearing molars, and the bear represents the omnivorous group with grinding molars.

The rationale for the combined exhibit as presented to the child at Station 1 is this: while most people are familiar with the obvious canines and incisors in the front of the animals' mouths, not many know the action of the powerful back teeth which are considerably less conspicuous; therefore the programed stations show how these seldom-seen important teeth do the heavy work of preparing food to be swallowed.

So that the picture of the basic shapes of mammals' teeth would be complete, incisors and canines were later installed in the free-access area. The concepts of the entire exhibit are summarized schematically in Figure 6.

At the close of the validation testing of the shearing molar series, the installation of the grinding stations began. Since these were closely patterned after the shearer units, only a short debugging process was needed.

The new configuration in the exhibit will be described briefly and then the validation results for the combined complex presented.

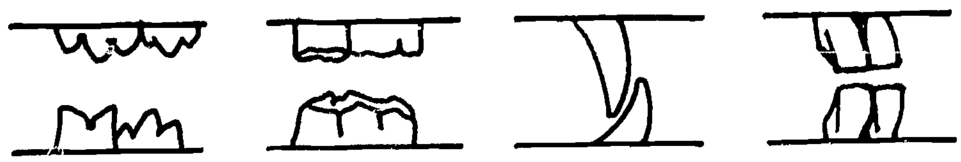
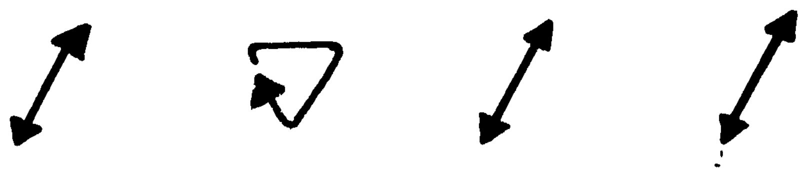
NAME	SHEARER	GRINDER	CANINE	INCISOR
SHAPE				
MOTION				
FUNCTION	cut tear	crush grind	slash hold	nibble hold

FIGURE 6. Schematic summary of exhibit concepts.

Description

To make room for the grinding-molar materials, the programed area was enlarged at the expense of the free-access area as Figure 7 shows; otherwise the free-access area remained unchanged.

The ten stations of the combined series presented the following picture:

At Station 1 a bear skull shared the spotlight with the tiger skull. The script called attention to the back teeth of the two animals. The tiger skull continued to move automatically but the automation of the bear skull was not accomplished until many months later.

Station 2 remained the same as before, calling attention to the shapes of shearing molars and affording feedback through the six pushpanels.

Station 3 remained as an opportunity to exercise the newly learned discrimination by choosing the shearing molar in six successive pairs of teeth.

Station 4, a new station, was a copy of Station 2 except that the new one dealt with the different shapes of the grinding molars.

Station 5 was the magnet mechanism that formerly was Station 4. Now it tested for recognition of grinding molars instead of shearers.

Station 6 consisted of an oversized wooden model of the grinding molars of a bear. The child could move these to see how they work with an almost circular motion during grinding.

Station 7 consisted of the wooden model of the gigantic opposable shearers (formerly Station 5), redesigned to approximate more closely the way the upper and lower teeth move in relation to one another.

Station 8 consisted of a double zoetrope, automatically revolving to present side-by-side the actions of the two kinds of teeth. The child could look through first one peephole and then another to compare the moving teeth.

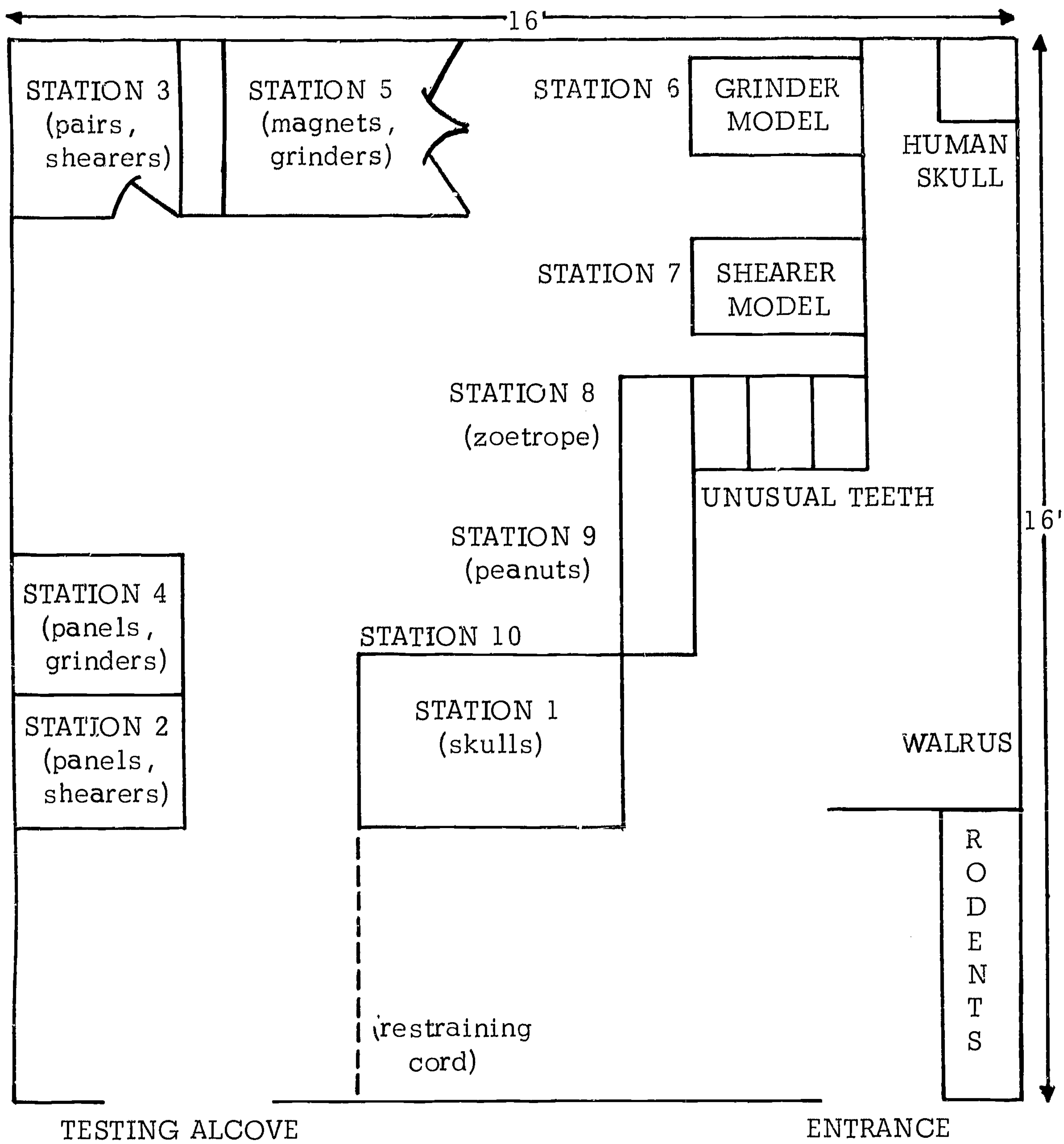


FIGURE 7. Floor plan of the combined shearing and grinding molar complex in February, 1968.

At Station 9, the peanut machine and the mirror were installed so that the child could see how the action of his own back teeth compared with that of the tiger's and the bear's. Pennies to operate the machine were furnished; a telephone handset carried the recorded message.

Station 10, by means of a telephone handset at the back of Station 1, summarized the exhibit by relating the concepts of shearing and grinding molars back to the context of the actual skulls.

Objectives and Test Procedures

The behavioral objectives for the combined series consisted of the two previous statements dealing with the shape and motion of shearing molars, plus two similar statements for the grinding molars; namely,

> when asked to pick out teeth that are "grinding molars," the child will be able to select real grinders of different sizes and shapes from a stock of single teeth including shearers, incisors and canines; and

> when presented with a correct and an incorrect movable model of grinding molars, he will be able to select the set that shows the molars closing against each other firmly and not the set where they slide by each other in a scissors-like action.

For practical reasons, the tong-like devices were made to demonstrate the occlusion but not the typical circular motion of the grinding molars in action; in one set the molars closed together with the lowers firmly against the uppers, but in the other set, they slid by each other more as shearing molars do.

The validating test procedures were like those used in the shearer series. The testboards had to be changed to accommodate the grinding molars; whereas the two previous testboards had held twelve real teeth each, now four new boards were made to hold eight teeth each (see Figure 2). Two boards were color coded red to signify the shearing-molars test and two were marked with blue for the grinding molars. One blue board and one red board were used for pretests, the others for the posttests. Thus each child made 16 choices in the pretest and 16 in the posttest, as compared to 12 each in the earlier series.

After the pretest with both testboards came the question about the tongs, the shearing pair first and then the grinding ones. The same sequence was followed in the posttests.

The instructions for the testboard task were given by recorded messages over earphones for both pre- and posttests, but for the tongs they were given orally.

Validation Results

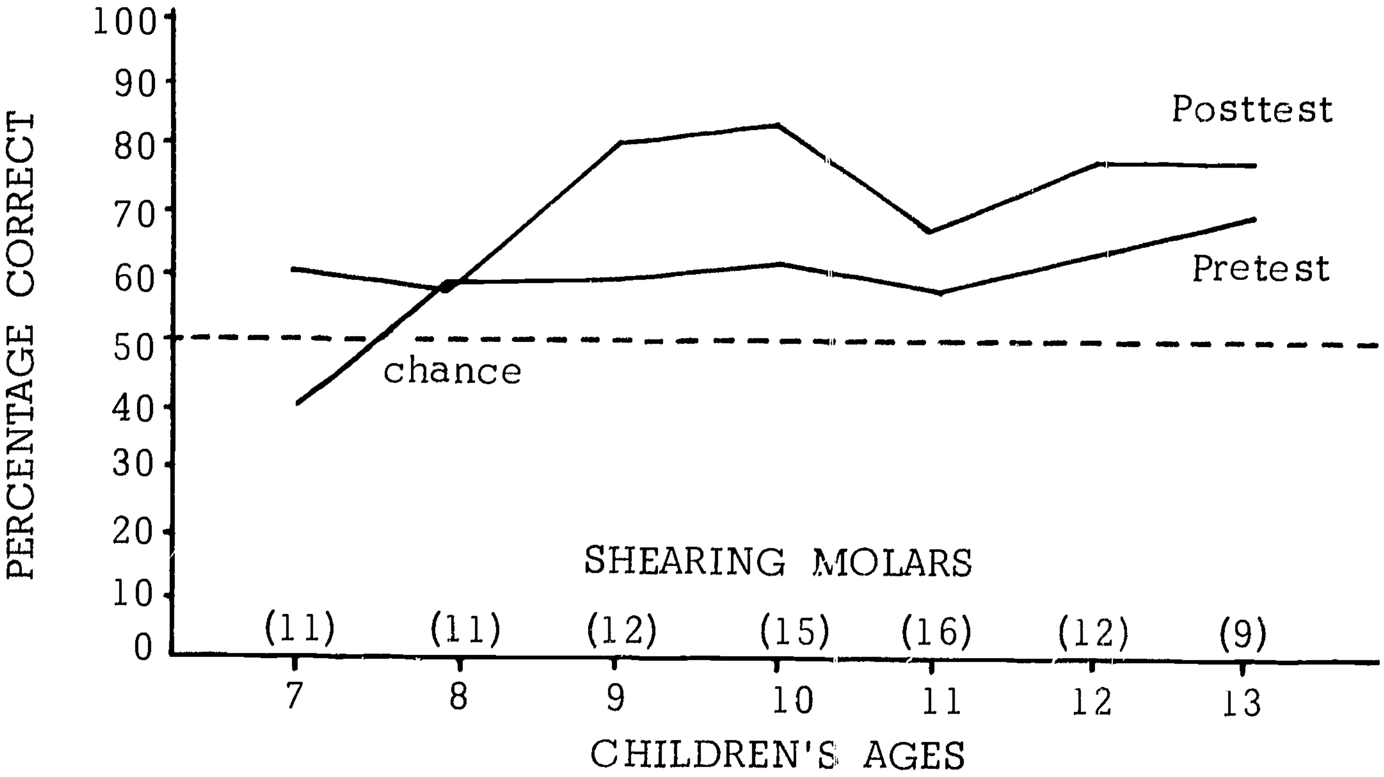
For standard tests with all stations operating, 47 girls and 39 boys, seven years of age or older, took part in the formal evaluation during February, 1968.

The total results for this large a sample are very highly significant in the statistical sense. For boys and girls considered separately, and for shearing and grinding molars treated separately, the results are unmistakably significant:

	Ave. Age	No. of Children	Per Cent Correct Pretest	Posttest	Difference	t
Girls	9.66	47	62.0	77.3	15.3	5.74***
Boys	10.41	39	63.8	77.4	13.6	4.39***
Total	10.00	86	62.8	77.3	14.5	7.21***

	No. of Cases	Per Cent Correct Pretest	Posttest	Difference	t
Shearing Molars	86	57.9	71.3	13.4	4.80***
Grinding Molars	86	67.8	83.4	15.6	4.64***

The graphs in Figure 8 show a steady effect for all ages. Again, there is a slight upward trend with age for pretest score; and posttest scores tend to follow the same direction -- in other words, the amount of gain is fairly steady regardless of age.



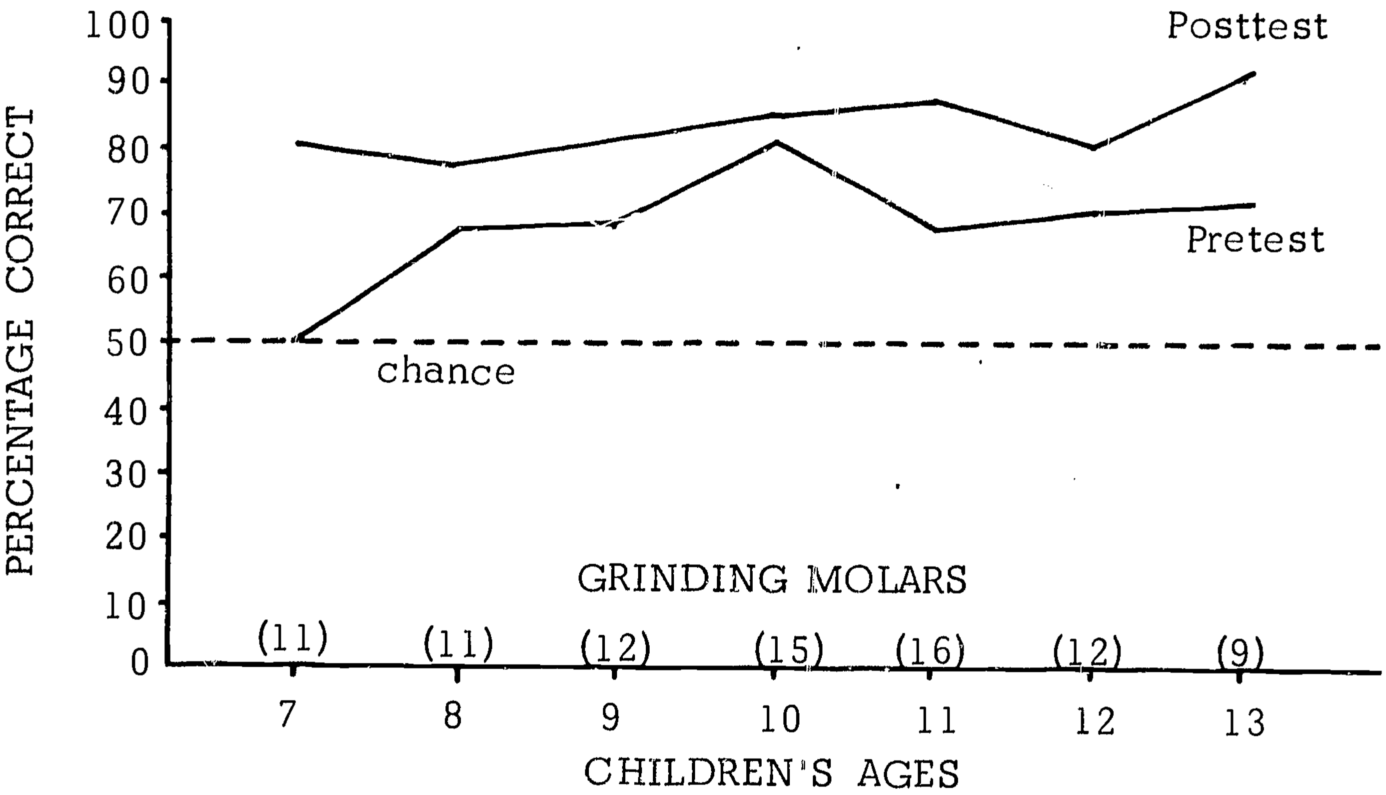


FIGURE 8. Graphs of pretest and posttest scores for the shearing molar and grinding molar sections of the February evaluation period. (The number of cases at each age level is given in parentheses just above the age.)

The most apparent fact about these results in terms of averages is that the posttest-pretest difference was considerably less than that when only the shearing molar series was operating. In the December series the posttests averaged 78.6% success, while here for the comparable shearing molar tests the average was only 71.3%. The reduced rate of gain may be due to the fact that the amount of information in the exhibit has doubled or it may be because the shearing molar series lost some redundancy when the magnet station was transferred to the grinders topic.

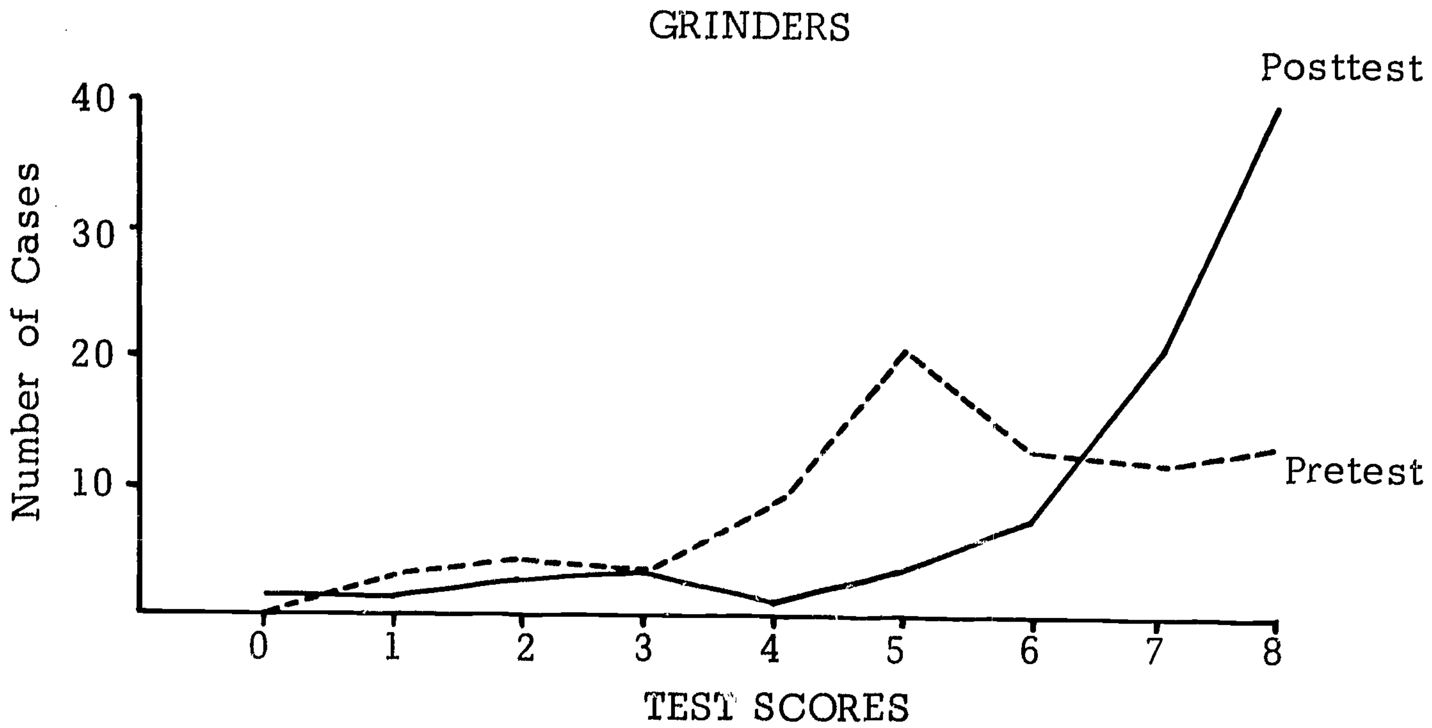
For the grinding molar tests, we see that children already knew something about the topic before visiting the exhibit as evidenced by the fact that they could select grinding molars with almost 68% accuracy.

If we turn to individual performance scores to see about the generality of the gain trends for shearers and grinders tests combined, we find:

Girls:	78.7% Improved
Boys:	76.9% Improved
Total:	77.9% Improved

Overall, these figures are about at the same level as for the earlier shearing series alone, but the number of boys showing improvement is almost ten per cent higher than before. In the statistical sense, the per cents for boys, girls and total are each highly significant. These per cents are important in indicating the generality of the exhibit effects. It would have been possible for the gain scores cited first to be significantly different as the result of extreme scores by only a few children. These significant per cents, however, indicate that actually a large number of individual children did contribute to the gain-score effects.

For the December-January series for shearing molars alone, we found that the most frequent score on the pretests was the chance-expected score but that on the posttests was the perfect score. Now for the combined shearer-grinder series, we made the same sort of analysis of the score distributions. These are shown in Figure 9 with the results of the shearers and grinders plotted separately. Because the testboards now have fewer trials than before, a score of 8 is perfect and a score of 4 is the expected chance average.



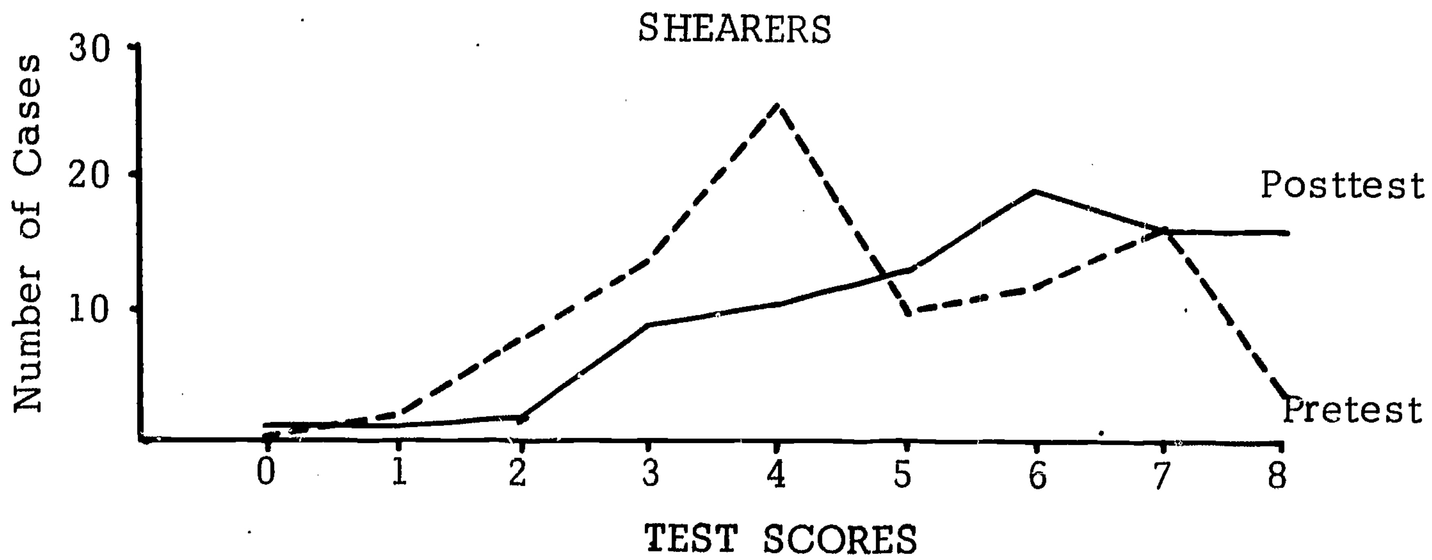


FIGURE 9. February Series: graph for the combined shearer-grinder complex showing the number of children who obtained the various scores on the testboards before and after visiting the exhibit. (Chance score is 4.)

On the pretest the scores for shearers are distributed very much as they were for the earlier series; the most frequent score is the chance score. The posttest scores are heavily grouped toward the upper end, and while scores of 7 and 8 were strongly represented, the most frequent score was 6 rather than the perfect score.

The graph for the grinding molar tests shows the children's familiarity with grinders since high pretest scores are fairly common. The effect of visiting the exhibit is dramatically shown, however, in the swift rise of the posttest-score distribution - the perfect score of 8 was obtained by 46.5% of the children.

The tongs results were:

	No. of Children	Per Cent Correct Pretest	Posttest
Shearing Molars	75	62.7	76.0
Grinding Molars	75	73.3	80.0

The gains for the shearers are greater than they were in the previous series but the effect is not significant. Neither is that for the grinding molars where the children showed considerable skill in selecting the correct pair of tongs before visiting the programed part of the exhibit. Whether they acquired this information from Station 1 or before visiting the exhibit we cannot tell.

Observation and Counter Records

During the validation period children were observed as they took part in the exhibit activities. In all, checklist records were made for 14 girls and 17 boys in the free-access area. The results are quite similar to those from the earlier period. The girls stayed on the average 3 minutes while the boys stayed a little longer, 3.4 minutes, as they did in the other series. The shortest stay recorded was 47 seconds and the longest was ten minutes.

In terms of popularity of the four units in this area, the walrus lost out this time; he received only 12 visitors from our sample while the rodent series had 26. The general picture is that the boys were more exploratory in their behavior, pushed more buttons and touched more specimens here.

A total of 67 children were observed in the programed area, Stations 2 through 10 (Station 1 having been recorded as part of the free area). For Stations 2, 3 (pairs) and 5 (now magnets), the average times for boys and for girls were generally a fourth to a third shorter than they had been for the same stations in December and January. This is puzzling because the recorded messages at these stations were actually a little longer than they were earlier. It is possible that they were more efficient; it is also possible that an increased number of repeaters were included in the sample.

For the total times in the programed area, the boys averaged about 17 minutes and the girls about 16. These times are both just about 4 minutes longer than the shearing series alone took. The addition of 3 new stations and the inclusion of grinding-molar information in some of the old stations did not therefore drastically increase the required time. But the children spent less time on the earlier more basic stations.

The "button-pushing" phenomenon is even more striking for this period. For Station 2 in December, 29% of the girls and 48% of the boys had pushed all six panels; but now in February, only 8% of the girls and 10% of the boys did that. Station 4 in the February series is just like Station 2 except that it concerns the shape of grinding molars instead of shearing ones. The data here are quite similar to those of Station 2 for February. These and other results are summarized in Table 3.

Station 3, the pairs task, shows a general improvement too in the number of correct choices made. The only exception is that fewer girls made a perfect score in this station -- only 27% now compared to 48% in the December sample.

The magnet station, now number 5, cannot be compared with the December station because it has been changed from shearers to grinders. The per cent of correct responses now is very high, as the table shows. The boys only required an average of 6.34 trials to find the 6 grinding molars, while the girls averaged slightly more with 6.97 trials required for the task. This of course reflects the fact, noted from other data, that these children already knew something about grinding molars. None of the children were observed to play with repeatedly inserting the teeth in the tubes as some had done in the early series.

TABLE 3

CHILDREN'S BEHAVIOR IN CHOICE TASKS,
SHEARER-GRINDER COMBINED SERIES, FEBRUARY PERIOD

STATION 2	(6 panels: 3 shearers, 3 nonshearers)	
	30 Boys	37 Girls
Percentage of Children Who:		
Pushed Every Panel	10.0%	8.1%
Pushed Only Correct Panels	33.3%	16.2%
Made One Error or Less*	80.0%	73.0%
Percentage of Correct Panels Pushed:	76.8%	76.1%
STATION 3	(6 shearer-nonshearer pairs)	
	30 Boys	37 Girls
Percentage of Children Who:		
Pushed All Buttons	3.3%	2.7%
Made Perfect 'Score'	66.7%	27.0%
Made One Error or Less*	76.7%	64.9%
Percentage of Correct First Choices:	84.4%	75.0%
STATION 4	(6 panels: 3 grinders, 3 nongrinders)	
	30 Boys	37 Girls
Percentage of Children Who:		
Pushed Every Panel	10.0%	10.8%
Pushed Only Correct Panels	43.4%	21.6%
Made One Error or Less*	73.3%	62.2%
Percentage of Correct Panels Pushed:	74.8%	70.8%
STATION 5	(magnets; 6 grinders, 6 nongrinders)	
	30 Boys	37 Girls
Percentage of Children Who:		
Put All Teeth in Tubes	0.0%	2.7%
Made Perfect 'Score'	76.7%	51.4%
Made One Error or Less*	86.7%	81.1%
Average Number of Trials to Find the 6 Shearers:	6.34	6.97

*Includes cases in row above.

The records kept by the counters attached to the pushbuttons of Station 3 support the observational data. Regular visitors to the exhibit during this period made between 837 and 914 responses to each pair. Of these 77% to 81% were correct. These results for the general population are remarkably close to those recorded by the observers for the test sample.

Various findings certainly show that the children, and especially the boys, did not engage in as much exploratory behavior as before. The difference does not appear to be attributable to changes in the script for the latter were primarily additions of context and did not involve any difference in instructions. The data do fit in with the hypothesis that many of the visitors may have gone through the exhibit before.

Special Studies

In order to explore the possibility that the tests themselves were informative, we collected two batches of data from control subjects during the February period. In the first batch, 13 children were given pretests, then posttests immediately afterwards without the usual exposure to the exhibit in between. The percentages of success on these consecutive tests were:

	Pretest	Posttest
Shearing Molars Tests:	67.3%	66.4%
Grinding Molars Tests:	68.3%	74.0%
Total:	67.8%	70.2%

The differences are trivial.

Later in the validation period there was an opportunity to collect the second small series. According to a set balanced pattern determined in advance, children were assigned to be tested either by the standard procedure (pretest, exhibit, posttest) or by the control procedure (pretest, posttest, exhibit). Ten subjects were in each group. The percentages of successful trials were:

	Pretest	Posttest
Standard Tests:	64.4%	75.0%
Control Tests:	51.9%	51.3%

The rate of gain for the standard tests falls just short of statistical significance, a result that is not surprising with so few cases. The control tests are pleasingly unimpressive.

But there are some peculiar aspects to the data. The difference in pretest scores for the two samples is a case in point: the children taking the standard procedure started out with a decided advantage over those in the control group. A closer look at the standard data reveals that 4 of the 10 subjects made perfect scores on the grinders' pretest, while only one of the control children did. There is little profit in speculating much over peculiarities in such small data samples. Unfortunately time did not permit a longer series. We can only conclude that both batches of control data show nothing to make one ascribe instructional value to the test instruments themselves.

Summary

The instructional system of ten stations contrasting shearing and grinding molars was visited by 86 test subjects above the age of six. The children showed a very highly significant gain in performance tests after the visit. The rate of gain was not as strong as for the earlier series, but the combined shearer-grinder complex dealt with approximately twice as much information.

The test results provide concrete evidence that a visit to the exhibit is followed by a significant improvement in distinguishing shearing and grinding molars from other types of teeth.

Children spent an average of 19 to 20 minutes in the exhibit (not counting the time in the waiting line). They refrained from indiscriminate "button pushing" to an even greater extent than in the first series.

Chapter 8

FINAL PHASE OF THE PROJECT

The evaluation period in February provided the validation data for the combined shearer-grinder complex. The exhibit did achieve the goals set for it. The rate of success for a heterogeneous population is fairly respectable. Yet there remains room for improvement, and adjustments to the system will continue.

Two aspects of the project remain to be discussed: a short series of tests made in April after the free-access area had been changed and the translation of the entire system into permanent materials.

April Short Series

To round out the treatment of the main tooth forms of mammals, activity stations for the front teeth -- incisors and canines -- were installed in the free-access area. They replaced existing units as a temporary measure to enable us to collect a few test data and observation records with them in the system. The Museum was preparing to close for five months while moving into an adjoining facility which was undergoing remodelling. Since this would mean a long interval before we could resume testing, we decided to use the final vacation week (Easter) to gather some supplementary information about how the entire complex might function with the incisor and canine stations included. The latter were quickly mocked up and installed in the free-access area although they were more like the programed section. The big teeth specimens were removed to make room for them.

The incisor unit incorporated the human skull, and displayed two photos of a child biting into an apple and of a child biting into an ear of corn. The human skull was mounted on the station panel at eye level, and adjacent to it was a mirror so visitors could compare their own teeth with the skull's.

Five real teeth were attached to the panel in the vertical row along with five large pushbuttons. Those buttons that were next to incisors lighted up with a red lamp when pushed, while others showed no light. A telephone carried the recorded explanation.

The canine station presented a leopard skull and a striking photo of a leopard head. A telephone and a sample block with various sizes of canines completed the equipment here. In an expansive moment we had the leopard's canines fitted out with concealed lamps so that when a button was pushed, they lighted up with what we hoped would be a droll effect. In practice it turned out to be more corny than humorous, and frightened as many children as it amused.

The rodent series and the walrus were now provided with telephone sets to carry a short message about the animals. The programed area remained unchanged.

Two kinds of data were collected: 1) a short series of pretests-posttests that turned out to be uninformative, and 2) a batch of observational records that were quite useful in revising the free-access area.

The pretests and posttests were for 20 children in the standard procedure and for 19 in the control procedure. Children were assigned to one or the other group on the basis of a pre-arranged pattern. The number of cases we were able to gather was too small to provide much information but they are given here for completeness. The percentages of successful trials for the two sections were:

		Pretest	Posttest
Shearers	Standard	53.5%	64.2%
	Control	54.6%	66.5%
Grinders	Standard	62.9%	72.8%
	Control	77.0%	75.7%

For the shearing molars, control subjects gain as much as those exposed to the exhibit; for the grinding molars the results are in the expected direction but the control subjects knew a great deal more to begin with.

The control group was on the average half a year older than the standard group. Whether that might account for the difference, we cannot tell. The data show wide and unaccountable fluctuations; about all we can conclude is that they tell us little.

One analysis was made to see whether the testboard data gave any indirect evidence of the effect of the new canine and incisor stations. The testboards each contain one or two samples of each of these front teeth. The question the child was always responding to was whether these teeth

were either shearers or grinders (depending on the series). Thus a correct response for a canine or an incisor tooth was a "no" response.

The rate of success for canines and incisors in April could be compared with that from the February series when no canines and incisor stations were present.

For this analysis we pooled pretest and posttest data, standard and control series; this was done because the canine and incisor stations, being outside the programed area, were supposed to have been visited before the pretest by all test children.

Percentages of success in April are given below along with comparable figures from the February series:

	Canine	Incisor
April	77.9% (190 trls)	80.7% (228 trls)
February	66.3% (380 trls)	76.5% (456 trls)

The data suggest that the canine station may have been having some impact. The effect with the incisors is small.

A formal test of these stations would have had to include a proper pretest before the child entered the exhibit area and some method of insuring that these stations were actually visited by the child. These stations were, however, not well enough developed yet to warrant that much trouble. The present figures, taken only as rough estimates, point to the need for more work on the incisor station at least.

The same conclusion was suggested by the April observation data which showed that the message at the incisor station was both too long (1.65 min.) and unclear so that children became impatient and resorted to random button-pushing.

The observation records, covering 17 cases in the free-access area only, were especially informative on the topic of recorded messages and feedback buttons in a free situation.

In the programed area where the flow of traffic was regulated, children could listen to the messages undisturbed and make responses at their own pace. But in the free-access area where crowds often gathered, a child trying to listen on the phone would be surrounded by others who

impatiently pushed buttons while waiting for a turn at the phone. And of course these button-pushers had no notion what the station was about or why certain lights were coming on; and for the child on the phone, the effect was spoiled too. These problems were especially apparent at the canine and incisor stations where the messages were in excess of one minute. The rodent series was judged most successful; it was the shortest (0.85 min.) and required the simplest response.

Study of these records suggests that:

1) in an uncontrolled area, the messages should be no longer than 30 seconds,

2) the responses required should be simple, requiring little deliberation or study, and

3) duplicate phones would be useful on stations where no feed-back is involved or where button-pushing would not spoil the experience for another who was watching.

New Construction

With the closing of the Museum in May, the programed stations were redesigned, retaining the essence of the validated task, but translating it into permanent materials and a new configuration in keeping with the structure of the new facility.

The modular concept was retained but now the main construction materials consist of molded plywood units bolted to metal posts. These can be combined in varying patterns to form the instructional stations. This type of construction retains some flexibility so that a certain amount of change is still economically feasible.

The pairs of photographs that appear through this report show the developmental and permanent versions of many of the stations. A number of modifications were made during the reconstruction process. In a further effort to relieve the traffic problem, duplicate copies of six stations were made. They were arranged in a chain-like configuration as shown in Figure 10. The free-access area and Station 1 with the skulls are now in the center of the programed "ring." The exhibit now covers an area about 24 feet wide and 28 feet long - over 2 1/2 times the area of the developmental setup.

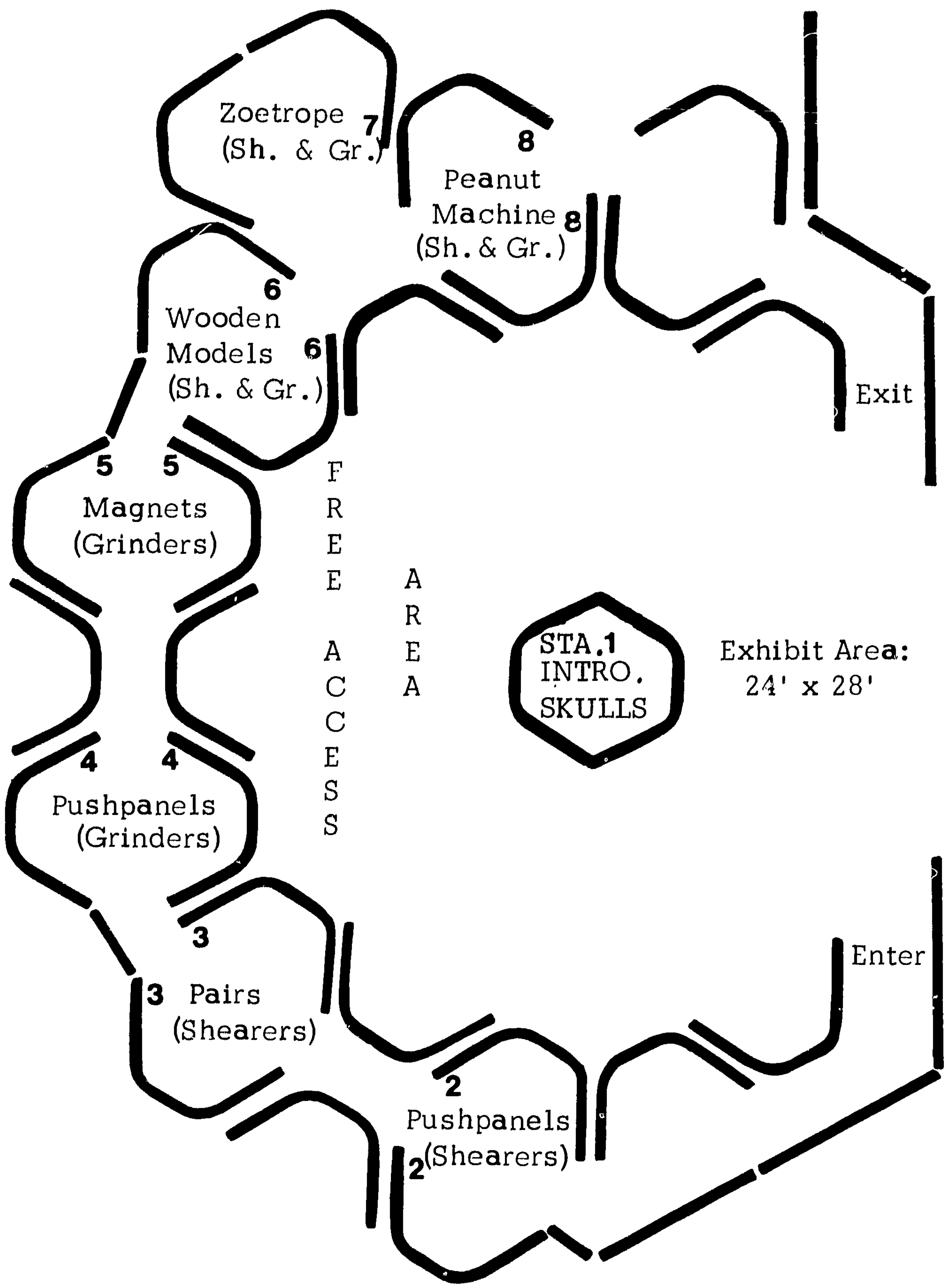


FIGURE 10. Floor plan of the exhibit in the new facility. Pairs of duplicate programed stations form a ring around the central free-access area.

The lower jaw of the bear moves automatically in the slow rotary motion typical of animals with grinding molars. Both skulls have now had to be protected behind a plexiglas shield. This was because, with increasing age, the skulls became more fragile and consequently took less kindly to having their jaws slammed shut by impulsive visitors. The mechanism providing the realistic jaw motion did not fare well either in an unrestricted environment and so reluctantly we had to provide the plexiglas protection.

Stations 2 and 4 had made their point about tooth shapes through pushpanels bearing schematic drawings of the various tooth forms. In the new facility these drawings have been replaced by handsome three-dimensional models of the teeth which are mounted on the colorful lighted pushpanels.

The magnet station, it may be recalled, had a dozen loose teeth in a container under the magnets and tubes. In its developmental form it had operated for ten months without a single tooth being lost. In the new facility, however, the entire stock of teeth disappeared all at once several times within a few weeks and none were recovered. Consequently the station was modified so that the teeth are individually tethered now on enough nylon cord to permit a child to test them against the magnet in his search for the grinding molars.

Another change was that the large wooden models of shearing and grinding molars which formerly occupied two stations, were now reduced in size and combined into one station. Whereas originally each tooth was roughly about six inches high and eight or so inches long, they are now only about one inch or so in height and from one to three inches in length, depending upon the kind of tooth. And instead of having only a single upper and lower tooth, each model now consists of a set of two or three uppers and lowers. By decreasing their size and presenting them as sets in close proximity, we felt that the shapes and distinctive motions could compared and contrasted more easily.

Formerly, in the zoetrope station, children looked through two peepholes to see the animated strips of shearers and grinders in action. Now the zoetrope has been changed to display two strips of drawings simultaneously; and it is mounted out in the open where children may spin it themselves to view the different actions of the shearers and grinders.

The activities and displays that were in the free-access area of the developmental version of the exhibits have not yet been installed in the new facility. The intention is to add these to the center area gradually as time and funds permit.

Since the opening of the new Visitor Center, the number of visitors has increased beyond the capacity of the facility; considerable attention has had to be given to the management of crowds. In our area, one or more attendants are needed to regulate the flow of traffic in the programed area. Between five and six hundred children now complete the exhibit circuit daily.

The task of adjusting the series to its new construction and its new setting continues. When the need for refinements to the equipment has dwindled, another test series will be run to see whether the rate of gain in learning has held up to the level achieved in the old quarters.


10

Chapter 9

RETROSPECT AND PROSPECT

In this chapter our aim is first to consider those findings and observations of the project that may have general relevance for museums and then, in the final section, to present some thoughts on exhibit research prospects.

Discussion of the Findings

Developmental Strategy

The cycles of evaluation and revision to which our thinking and materials were subjected were of critical importance, we believe, in the achievement of an exhibit with demonstrated effectiveness. Without the changes forced on us by empirical data, the exhibit stations could not have reached so large a proportion of the visiting public. There is no doubt in our minds that the exhibit is immeasurably different as a result of the iterative process.

We realize also that our thinking was markedly changed by the developmental philosophy. We were conscious of a continuous refinement of ideas and a pressure to analyze situations more deeply and to make explicit the effects we expected from them. Only gradually, we realized, did the formulations of the exhibit objectives move toward clarity.

Thinking of the exhibit as a communication form made us conscious of the need to understand our audience, to speak the same language and to be responsive to their characteristics and interests. By studying them and by taking them into partnership in the development cycles, we believe the exhibit was significantly improved.

The evaluative methods not only guided us when we were uncertain, but more importantly, they identified gaps in our presentation, points of confusion, ineffective graphics, and equipment that was not functioning as we intended. On the other hand, they also told us when we were on the right track with an approach or a technique. Thus by helping us identify both good and bad features early in development, we believe the evaluative tools were in the end economical of time and funds. Time was saved because maladaptive

solutions were caught before they were heavily built on; errors were caught before they had been translated into expensive permanent structures.

Evaluation used as a tool throughout the development process need not be an expensive proposition. As an earlier chapter mentions, it does not need to be sophisticated and formal, especially in the early stages. Important information can be gathered through observational studies and interviews conducted by nonprofessional staff. For more advanced procedures, staff can be trained to handle these with a minimum of supervision. Equipment does not need to be elaborate, as the methods of this project illustrate. While it is true that evaluation can be conducted on a modest scale, we must not leave the impression that validated exhibits are no more expensive than unvalidated ones. Even though the construction materials are low cost, it still remains that the time that production and project staffs spend in revising exhibits increases appreciably the labor component in the budget. Thus if a simple question of comparative economics is raised, validated exhibits will come out relatively expensive. If the question is phrased, however, in terms of effectiveness per dollar spent, the practice of validation through successive approximations should emerge as the wiser use of resources.

In the summer of 1968 we had the opportunity to apply the developmental procedure on a larger scale in a series of studies of new exhibits being prepared for the new Museum facility referred to in the last chapter. Although the Museum was closed to the public during this period, a series of Guinea Pig Days was instituted to let children try out exhibits in their unfinished states; the objective was to let developers study the effects their partially developed materials achieved with the visitors. Observational studies, interviewing and visitor-characteristic studies were pursued for many of these exhibits.

Apart from the informative nature of these efforts, several other outcomes emerged. One was that the tryout philosophy has to be "sold" -- some designer-developers were not only unconvinced of its utility but viewed it as a test of their plans and projects rather than as a tool for development. To inquire into objectives was a delicate and often unrewarding experience for the evaluator.

On balance, we must add that some designers immediately recognized the potential value of the tryouts and devised lists of specific questions on which they sought observers' data. They learned to take advantage

of the chance to get quick answers to design problems rather than spend time in deliberation or argument.

To return to our own project, one consequence of the development practice was the psychological problem the staff had in sustaining interest through the long evolutionary period. In a project involving as many separate stations as this one did, it had seemed wise to confine the construction to sturdy temporary materials until the outlines of the entire series were clarified and the system debugged. For us, the fact that everything was open to change over almost a two year period meant that we did not have the normal satisfactions of achieving subgoals and of obtaining final closure on problems that had engrossed us. It was fortunate that periods of psychological flatness did not hit everyone at the same time, but it should be recorded that it did take conscious effort to maintain momentum.

It should be recorded as well that tryout and evaluation cycles were exciting and surprising; it was fun to find out how our ideas fared under the critical onslaught of the small people. There was always the rewarding element of the unexpected.

In summary then, we believe that our developmental procedure was economically feasible -- that it saved time and money; and that it is practical to attempt this process with inexpensive materials and nonprofessional personnel under professional guidance. But most importantly, we are convinced that it was absolutely critical in achieving effectiveness for the exhibit system we built.

Instructional Exhibits

The project has demonstrated the educational value of a series of sequential exhibits in a museum setting. It is not our purpose here to argue the question of whether extended concept-building with complex displays is a valuable or desirable enterprise for museums to adopt. Our purpose has been primarily to demonstrate an exhibit design philosophy and to show that contemporary learning research has some ideas that can make exhibits for children exciting and instructive.

In the education world, those who prepare instructional materials know that no matter how skilled they are, an appreciable amount of their work will consist in filling in gaps identified in the course of developmental testing. We had the same experience with our exhibit

series. The finding that one must take smaller steps or give a wider range of examples is an important confirmation of the need for tryouts. It is quite clear that without these additions many children would have missed the message.

The implication for instructional exhibit building is that more materials and perhaps more space may be required to convey a message than one originally estimates. Closely related to the gap-identification issue is the need to build in redundancy. Longer involvement with a concept or a procedure is required for some learners than for others. Usually this can be achieved by phrasing things in alternate ways or by preparing different activities for the same teaching objective.

It is important to note that the need for small steps and frequent response opportunities may be partly a function of the setting. Distractions in the vicinity would create a different learning situation than would a quiet tranquil area.

The involvement of the child in active participation with the exhibit materials and the feedback they provided him were successful aspects of the system, we felt. The children not only enjoyed the series, but their interest was sustained throughout. They returned over and over even though a visit represented a commitment of 15 or 20 minutes of their time beyond the waiting-line time.

For instructional effectiveness we learned that recorded messages had to be short and to call forth an active response early on. It seems better to have the onset of the message under the child's control -- otherwise the message starts before he has his ears turned on.

Headsets were helpful in focussing the child's attention, in cutting down distracting noise, and in giving the child a claim to the activity area.

The system of stations we evolved, however, cannot instruct unless the flow of traffic is controlled. Each station needs to be operated by a single child or at most by a cooperative pair. On occasions when we have experimented with a relaxation of controls, the result was nothing short of bedlam.

When traffic is regulated so that no more than one or two children are waiting for each station, the system operates well, the children respect each other's turns and make their way without help in proper sequence through the circuit.

From an early tryout we learned that signs can be too large to be seen. When the huge numbers identifying the stations were replaced with small ones, visitors followed them easily. A related finding was that labels can be too stylized for children to read easily.

At times we have experimented with allowing parents into the programed area. The experience has been uniformly depressing. It is a rare parent that allows their child to proceed on his own without interference.

Future Research

Thinking about what makes one exhibit more effective than another always leads to thinking about research. The evaluative methods of the present project, with only slight modifications to include systematic controls, can easily be put to work in a program of research on exhibit effectiveness.

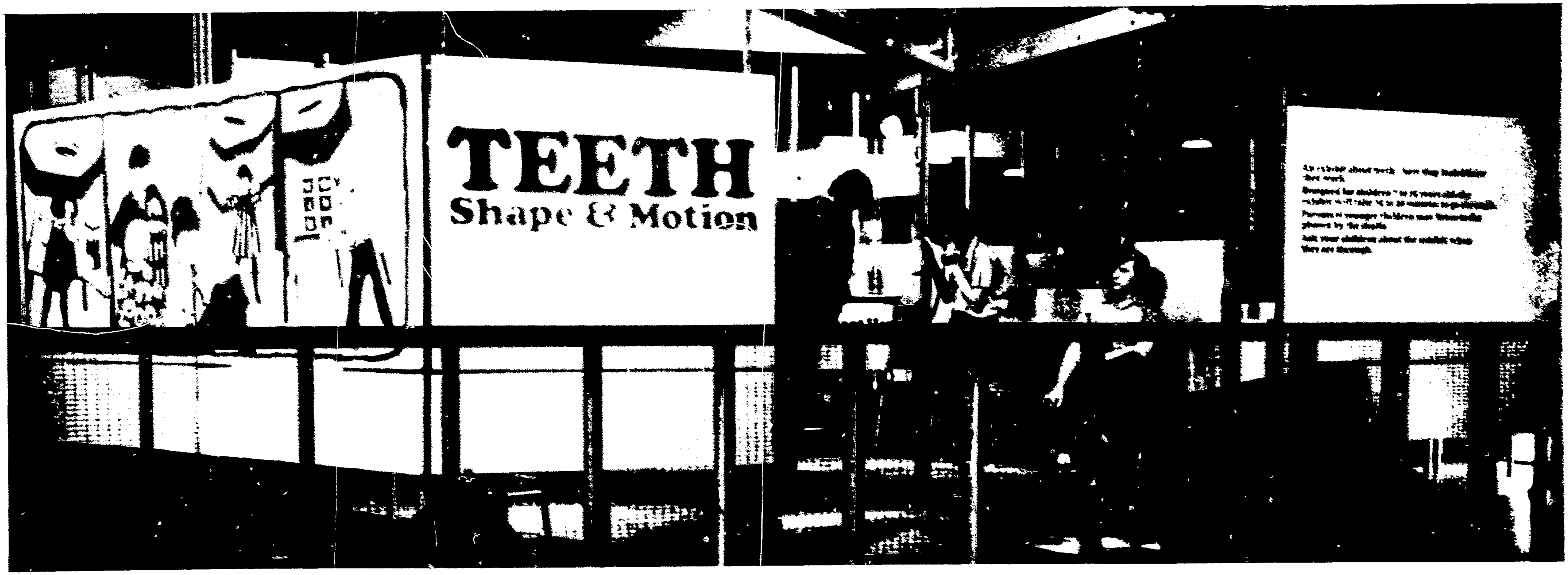
In considering the qualities of good exhibits, we become involved in questions of judging what effectiveness means. In our project we took as a working definition the performance of the child immediately after leaving the exhibit. This was an easy practical solution but we would not want to define effectiveness solely in terms of immediate impact. Probably the most important exhibits are those which exert long-term effects. Long-term effects of museums, like long-term effects of teaching methods, are not easy topics to tackle with research, but we believe a start can be made. We now have an information bank about our visitors which could serve as a base for inquiring into the characteristics of older memories of museum experiences. At the moment we can tap experiences only a few months old, but this could be a foothold on a difficult problem.

The role of real objects in exhibits is a topic that concerns museums vitally since it is often tied to questions of the role of museums in the modern world. Many are convinced that the impact of real objects is much more powerful than is that of representations, however clever or dynamic. This conviction, widely shared, has led to the inclusion of real objects in various innovative school programs. Yet little research on the question exists. Museums especially, it would seem, would profit from a better understanding of their important assets and how they affect the visitor. Systematic research with proper evaluative methods could help to clarify the role of real objects in effective exhibits.

In the museum setting, problems of attracting and holding the visitor's attention in the midst of distractions are familiar concerns of designers, and yet the topic is one about which little is known. The museum setting would be an attractive place for study of methods of overcoming the effects of distractions. Research findings in this area would have implications as well for the classroom scene.

The prospect of turning our series of duplicate instructional stations into a research vehicle for comparative studies is most enticing. Hypotheses about the properties that make for good exhibits could be put to the test by making the two copies of the exhibit differ only along the hypothesized dimension. For example, the value of different media for giving information to children could be studied in these twin circuits; or the feasibility of different instructional strategies for the museum setting could be investigated.

Research of this kind is not easy, however; the many variables exerting an influence on the experiences of the museum visitor make it especially hard to define experimental comparisons incisively enough to yield useful, informative results. Yet a beginning can be made to develop techniques for research in this area and to build up a stock of research findings that are realistically applicable to exhibit development.


TEETH
Shape & Motion

References

Ausubel, David P. The psychology of meaningful verbal learning: an introduction to school learning. New York: Grune and Stratton, 1963.

Gagné, Robert M. Contributions of learning to human development. Psychological Review, 1968, 75, 177-191.

Mager, Robert F. Preparing objectives for programmed instruction. San Francisco: Fearon Publishers, 1962.

APPENDIX

Example of Checklist for Recording Behavioral Observations

in Both Free-Access and Programed Areas

The Children's Museum
Val-Ex Observation Record
Free-Access Area
February 1968

Observer's Name ______________________ Date __________ Time ______

Child's Age approx. ____________________ Sex ___________

Came with: alone adult (parent, teacher, can't tell)
other children (1, 2, ____)

INTRODUCTORY ROOM: (start timing)

RODENT SKULLS

Came here: first 2nd 3rd 4th last not at all

	Beaver	Woodchuck	Squirrel	Rat	Mouse	None
Order of Pushing Buttons:						

Pushed more than one button at a time _______

____
TIME

BIG TEETH BEHIND DOORS

Came here: first 2nd 3rd 4th 5th not at all revisited

	Sawfish	Narwhale	Elephant	None
Order of opening doors				
Looked at pictures				
Touched				

____
TIME

WALRUS

Came here:	first	2nd	3rd	4th	last	not at all	revisited
Touched it							
Moved jaw							

____
TIME

HUMAN SKULL

Came here -	first	2nd	3rd	4th	last	not at all	revisited
Touched it							

____
TIME

STATION I

Came here:	first	2nd	3rd	4th	last	not at all	revisited
Listened to message:	all	part	none	repeated	can't tell		
Watched skull	yes	no					
Touched skull	yes	no					
Noticed Mural Inside	yes	no					
Other							

____
TIME

CHILD LEAVES INTRODUCTORY ROOM

Leaves the exhibit area
Hesitates about entering programmed area
Expresses interest in entering programmed area
Waits in line
Other: (e.g. parents takes child away.....)

TOTAL TIME IN FREE-ACCESS AREA (before let into prog. section or he leaves) ____________

GENERAL BEHAVIOR

General Attentiveness:	Careful Examination	Average Observance	Quick Glance
Progression	Orderly	Random	
Attention Directed by:	Adult	Children	Self
Adult Interference	Yes	No	
Told not to touch by adult	Yes	No	

The Children's Museum
Val-Ex Observer's Record
Programed Area
February 1968

Observer's Name ____________________ Date ________ Time ________

Child's Age approx. ____________ Sex ________

Came with: alone adult (parent, teacher, can't tell)
other children (1, 2, _____)

Came from: Pretest Pretest not available

STATION 2

Uncertain what to do	Discovers for self	Assisted	Understood OK
Earphones handled	correctly	wrongly	
Buttons pushed:	ok	has to be shown	
Refers to red block:	first	at wrong time	not at all
Refers to yellow block:	second	at wrong time	not at all

Panels pushed

1	2	3	4	5	6

Message repeated
Other

Time

STATION 3

Came in right sequence yes no guided
Earphones handled correctly wrongly assisted
Buttons pushed: ok assisted

Refers to sample block

Order of Pushing
Buttons (include Repeats)

	Left Button	Right Button
Pair 6		
Pair 5		
Pair 4		
Pair 3		
Pair 2		
Pair 1		

Repeated message
Tried to redo procedure
Other

Time

STATION 4

Came in right sequence:	yes no	guided	
Had to wait turn			
Earphones handled:	correctly	wrongly	
Buttons pushed:	ok		
Refers to blue block:	first	at wrong time	not at all
Refers to gray block:	second	at wrong time	not at all

Panels pushed

1	2	3	4	5	6

Message repeated
Other

______
Time

STATION 5

Came in right sequence:	yes no	guided	
Waited to get in			
Earphones handled:	correctly	wrongly	assisted
Buttons pushed:	ok	assisted	
Refers to sample block:	yes	no	

Record sequence of rights (R) and wrongs (W):

Tried to repeat procedure
Other

______
Time

STATION 6

Came in right sequence:	yes no	guided	
Had to wait turn			
Earphones handled:	correctly	wrongly	assisted
Button pushed:	ok		

Worked mode[1]
Touched teeth or inspected carefully
Other

______
Time

STATION 7

Came in right sequence:	yes	no	guided
Had to wait turn			
Earphones handled:	correctly	wrongly	assisted
Button pushed:	ok		
Worked model			
Touched teeth or inspected carefully			
Other			

Time

STATION 8

Came in right sequence:	yes	no	guided
Had to wait turn			
Earphones handled:	correctly	wrongly	assisted
Pushed button:	ok		
Peek inside	red hole	blue hole	
Repeated message			
Other			

Time

STATION 9

Listens to message	
Takes peanuts	eats peanuts
Looks in mirror	
Looks in mirror while eating	
Repeats message	
Other	

Time

STATION 10

Listens to message		
Touches:	bear skull	tiger skull

Time

TERMINAL BEHAVIOR

Revisits any stations
Wants to revisit stations but sent out
Leaves quickly
Any attention to mural: yes no
Taken to post-test post-test not available
Other

____
Time

_______ TIME LEFT EXHIBIT AREA