

A question is voiced at the end of this paper: Is the alteration of established patterns of behavior too much to expect of any single instrument of instruction? The query will echo in the mind of the reader—who, incidentally, will be impressed by the zeal given to making this study technically sound.

A Method for Evaluating the Effectiveness of Health Education Literature*

B. G. GREENBERG, PH.D., F.A.P.H.A., MARY ELLEN HARRIS,
C. FRANCES MACKINNON, F.A.P.H.A., AND SIDNEY S.
CHIPMAN, M.D., F.A.P.H.A.

Professor of Biostatistics, School of Public Health, University of North Carolina; Instructor, Experimental Statistics, Institute of Statistics, Raleigh; Associate Professor of Public Health Nutrition; and Professor of Maternal and Child Health, School of Public Health, University of North Carolina, Chapel Hill, N. C.

HE State of North Carolina, early in 1949, was contemplating a free distribution of a series of mental hygiene pamphlets entitled *Pierre the Pelican*, to all parents of first-born children. The pamphlets and their objective are best described by the author, Loyd W. Rowland¹:

The *Pierre the Pelican* series of pamphlets for the parents of first-born children was especially designed to cover good principles of child rearing—beyond the area of physical care. . . . The pamphlets have the following characteristics: They are simple, are cast at the sixth grade reading level, are illustrated by

sketches, have the continuity made possible by the character Pierre, comprise a series, make use of questions, are of optimal length, and make an effort to cover topics of principal interest to young parents. They are sent monthly during the first year of the child's life.

The characteristics of the pamphlets, as described, fulfill some of the conditions necessary for the effectiveness of educational literature. These conditions, of themselves, are not sufficient evidence of effectiveness. As clearly stated by Knutson,^{2, 3} a final criterion must involve field tests to determine if objectives are met.

Before embarking on a program of distribution, answers to three questions concerning effectiveness seemed highly desirable: Will the pamphlets fulfill in North Carolina the purposes for which they were intended? Will they be equally effective for all demographic groups in the state? Will their effectiveness, if any, justify the cost and effort involved in their distribution?

It was decided to conduct a study †

* Presented before the Maternal and Child Health Section of the American Public Health Association at the Eightieth Annual Meeting in Cleveland, Ohio, October 24, 1952.

† The initial purchase of the materials plus the expense and labor of mailing them would involve an undertaking of known magnitude and expense. To forward the first copy of the series as soon as possible after birth would require an expediting of the usual routine for registration and processing of birth certificates. Moreover, each month's new births would have to be screened monthly throughout the ensuing year to change addresses as well as to remove the names of parents whose infants had died. The latter detail is rather costly but important if the distributing agency is to avoid the possibility of a traumatic effect upon these parents.

TABLE 1

Selection of Sample to Evaluate Pierre the Pelican in North Carolina

Stratum	Basis of Stratification	Number of Counties Eligible	Number of Counties Selected	Terminal Digit of Birth Certificate Number To Be Used in Sample			
				White		Nonwhite	
				Control	Experimental	Control	Experimental
I	Counties with cities over 50,000	7	3	8	3	2, 6, 8	1, 5, 9
II	Counties with cities 10,000-50,000	18	3	0, 4	3, 7	all even	all odd
III	Counties with cities 2,500-10,000	30	5	0, 6, 8	1, 5, 7	all even	all odd
IV	All rural	45	6	all even	all odd	all even	all odd

to obtain information about the first two of these questions since the cost of a program of universal distribution could be readily estimated in advance.† This paper is intended solely to illustrate the method of the study.

PURPOSE OF STUDY

The objective of the investigation was to measure the effectiveness of the Pierre series upon the parents of first-born children in North Carolina. The measure of effectiveness used was the extent to which parents were motivated to adopt the practices recommended by Pierre. (In this instance, "practices" include all actions and attitudes related to approved methods of child-rearing.)

The objective might be more specifically stated by the following questions: Did the pamphlets produce a more healthful attitude toward problems of child-rearing? To what extent were parents motivated by means of this literature to modify their patterns of behavior in bringing up their first child? Inquiry for the latter was directed toward actual performance of the parents and not the verbalization of correct answers.

The parents eligible for this study were those whose first live child was born from July 1, 1949, to June 30,

1950. The birth also had to be legitimate and registered before the baby was three months old. Since the pamphlets are intended for both parents who are bringing up a child under normal circumstances, homes broken by divorce, death, or other separation were excluded. Cases in which the child was physically handicapped or where the mother was working full time and not immediately directing the child's training were also excluded.

DESIGN OF STUDY

The agencies interested in the distribution of Pierre were agreed, that although preliminary evidence by Rowland¹ had indicated usefulness of the pamphlets, they should plan to study the effects in North Carolina from the start. An experimental and a control study group were set up and the pamphlets mailed to all those not in the latter group. The sample for the study group was selected by classifying the 100 counties in North Carolina into four strata according to the degree of urbanization.* A sample of 17 counties (Table

* The selection of the sample was made with the advice and assistance of the following members of the Institute of Statistics: Drs. Richard L. Anderson, Alva L. Finkner, and Emil Jebe.

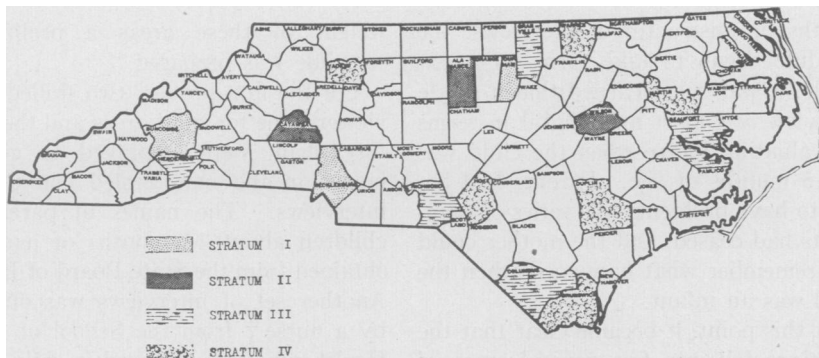


FIGURE 1—Selection of sample to evaluate “Pierre the Pelican” in North Carolina, by stratum.

1 and Figure 1) was selected for the study.

The counties were chosen from each stratum so as to have a high probability of being representative geographically of the three types of regions in the state. Based upon previous birth experience, an expected number of usable first births was calculated in each stratum. This was used to determine a subsampling rate for each race within the various strata so as to have proportional sampling. (These rates are shown in Table 1.) The sampling rates chosen were to provide a number of schedules sufficient for comparisons and tests of significance between the experimental and control groups, by race and residence. The sample was expected to contain around 3,200 families, about 20 per cent of which would be lost because they were either ineligible for reasons stated above or untraceable because of obsolete and incorrect addresses. After the survey was under way, it was found desirable to enlarge the sample by increasing the subsampling rate of the experimental group.

Within each of the 17 counties, the families were assigned at random to either the experimental or control group. The terminal digits of the birth certificate number were used in this process of allocation (Table 1). This reduced the possibility of bias due to differences between counties.

It was hoped that the purposes of the study could be achieved by a schedule of questions which, by leading the respondents to project themselves into a mood or situation, might elicit their attitudes, actions or hypothetical actions in response to those situations. Since this type of information could not be obtained feasibly by a mail questionnaire, personal interviews were used.

FORMULATION OF QUESTIONNAIRE SCHEDULE

The questionnaire was constructed so that an interviewer, with a fair amount of training, could obtain answers to questions which seemed simple to the respondent, although complex in intent and mechanics of asking.

The two senior authors started pilot studies in some of the 83 nonsampled counties which permitted at least one year of experimentation to learn how to ask questions which would give the necessary information, the order of asking them and the respondent's tolerance of the questions. The authors explored means of establishing rapport, of developing suitable probes and determining the threshold of parents' knowledge concerning their probable responses and actions under situations not yet encountered.

Interviews at well-child conferences in some of the nonsampled counties were wire recorded and, as a result of several

months of unstructured interviews, the schedule began to take form. For example, it was soon learned that a single interview was most meaningful in terms of evaluating Pierre when the child was 14–15 months of age. Parents had begun to have problems, the series of pamphlets had ceased, and the mother could still remember what happened when the child was an infant.

At this point, it became clear that the questions fell into five general areas of child training. These were: (1) daily routines of feeding, sleeping and toilet training; (2) learning and development (to walk, talk, use hands, etc.); (3) play habits and social behavior (with toys, other children, strangers); (4) parents' present attitudes and activities (compatibility, social activities, division of labor, etc.); and (5) parents' insight and their plans for the child's future development (answering child's questions, plans for more children, etc.).

Although it was hoped to analyze the combined results from each question by methods of scaling, it was apparent that improper emphasis on one or more of these areas might easily distort the final results. A panel of experts* in the fields of pediatrics and psychiatry, available as consultants to the agencies involved, was questioned on this problem. There was almost unanimity of opinion that the greatest weight should be given to feeding habits and parents' attitudes and activities, particularly as they concern the mother-child relationship. This opinion reinforced the investigators' convictions of the necessity for interviewing the mother, preferably alone. Having obtained a general notion of the relative

weight of these areas a preliminary schedule was prepared.

For the first pretest, two skilled interviewers, one for rural areas and the other for urban, were employed to go into homes in the nonsampled counties for interviews. The names of parents of children about 14 months of age were obtained from the State Board of Health. Another set of interviews was obtained by a nurse † from the School of Public Health who was engaged in field training in Cleveland, Ohio. The three interviewers were apprised of the purpose of the questionnaire and by trial and error they continued to improve the questions. A crude item analysis was then made to eliminate those questions which did not help to distinguish between urban-rural or white-nonwhite patterns of child training so that only sensitive and discriminatory questions would be used.

The true purpose of the questionnaire was revealed to neither the respondent nor the persons later employed for the interviewing. Neither group knew about experimental and control families. These conditions we consider essential for a valid determination of effectiveness in studies of this kind. The Pierre series entered into the interview only near the end in what appeared to be a casual manner. There was no reference to the State Board of Health and the interviewer mentioned only the School of Public Health.

All questions and their sequence were designed so as to promote an informal interview. If the respondent's reply was indefinite, probing questions of tested value were available for use. Several buffer questions without relation to the evaluation phase but with health significance, such as the duration of breast feeding, were also included.

The final pretest of the questionnaire was made by 15 interviewers in their home and neighboring counties, using June, 1949, births, and the completed

* The authors gratefully acknowledge the helpful and kind replies received from the following individuals: Drs. David A. Young, Benjamin Spock, Robert B. Lawson, Maurice H. Greenhill, Jasper Hunt, Robert J. Murphy, Lloyd Thompson, and Vernon Kinross-Wright.

† Miss C. Margaret Johnson.

form was ready by September, 1950, for the July, 1949, births.

INTERVIEWERS

To obtain highly competent interviewers, area supervisors of the 1950 federal population census were consulted early about the future employment of their outstanding enumerators. Fifteen interviewers were selected from this group. All were housewives, most of them with college training, and, with few exceptions, in the age range from 30 to 50. It was part-time work for all of them, and their patience and enthusiasm for collecting the material contributed greatly to the quality of the interviews.

A three-day training period with regular classes was held in a central place. Owing to the fact that census enumeration differs in technic from attitude surveys, the first day was devoted to general instructions about this type of interviewing and a detailed review of the entire questionnaire. Before termination of the formal training, each interviewer was tested. As a final effort to improve the interviewing technics, each interviewer completed about ten schedules from among the June, 1949, births, and all errors and omissions were reviewed.

The interviewers' work on the actual survey was spot-checked for completeness and accuracy by a supervisor who visited some of the interviewed families. Every schedule was also checked against specific information available from the birth certificate but which had been withheld from the interviewer.*

* This information included: child's name (interviewer knew only if it was a boy or a girl), mother's and father's ages, and father's occupation. Minor discrepancies in these categories were permitted when they seemed reasonable. Consistent deviation from correct birth certificate data was the means of discovering and discharging one unscrupulous interviewer. All her interviews were rechecked and verified by her replacement or the supervisor.

† Birth statistics for nonwhite include Negro, Indian, and Chinese with 97 per cent Negro, and the remainder predominantly Indian.

NUMBER OF SCHEDULES COLLECTED

In North Carolina from July 1, 1949, to June 30, 1950, there were 21,012 white and 4,381 nonwhite † resident births which could be included in this study since they were properly registered, legitimate first-births. Among this group of children it was learned that infant deaths later removed 612 whites and 229 nonwhites. From the total of 24,552 surviving infants, the sample which was chosen consisted of 3,150 study cases (2,309 whites and 841 nonwhites). There was a loss of 30 per cent of the original sample due to untraceable addresses, the requirement of an intact home, and a few other causes. Reasons for the failure to get complete interviews are presented by race and by study group in Table 2. There were only three refusals to be interviewed.

Within each race, the pattern of the distribution for no interview was almost identical between the experimental and control groups. One slight discrepancy occurred in infant deaths. The names of the control group were not on the mailing schedule for Pierre and infant deaths had not been removed from their list. The remaining children's deaths listed were either missed in the matching process or occurred in the interim between 12 months and the time of interview.

In both the experimental and control groups, 13.5 per cent of the white mothers were working full time whereas only 8 per cent of the nonwhite mothers were doing so. This was possibly due to the fact that relatively more nonwhite mothers were involved with a second child at the time of the interview. The rate of separations of parents among the nonwhite group was almost four times greater than that among the white.

The almost perfect similarity of the experimental and control groups (with exception of infant deaths) gave encouraging evidence of the validity of the controls. There was also additional

TABLE 2

Composition of Original Sample, Classified by Reasons for not Obtaining Complete Interviews by Race and Study Group

Results of Interview Efforts	White		Nonwhite		Total	Per cent of Sample
	Experimental	Control	Experimental	Control		
Complete Interviews Not Obtained	357	319	116	159	951	30.2
Mother working full time	180	128	29	40	377	12.0
Family had moved out-of-state or with no forwarding address	107	97	34	27	265	8.4
Parents separated or dead	30	31	34	41	136	4.3
Family unknown at address	22	31	5	19	77	2.4
Infant died or abnormal	5	21	4	13	43	1.4
Not first child	10	6	7	12	35	1.1
Illegitimate	0	0	3	5	8	0.3
Mother ill	0	1	0	2	3	0.1
Refusals	1	2	0	0	3	0.1
Other *	2	2	0	0	4	0.1
Complete interviews obtained	868	765	288	278	2,199	69.8
Total sample of eligible births	1,225	1,084	404	437	3,150	100.0

* Mother could not speak English, or impossible to reach home for months due to roads and weather.

confirmation obtainable from other variables, such as household composition, age and education of parents, home-ownership, and father's occupation.

The question arose whether control families had read Pierre pamphlets through copies from friends, or some other means, thus tending to dilute any differences between groups and invalidating their use as controls. If differences appeared despite this potential dilution, it would indicate a value of the pamphlets greater than that measured. Fifty (less than 5 per cent) of the control mothers replied that they had seen a monthly pamphlet on child care, and an attempt was made to ascertain if the reference was to Pierre or some other literature. Pierre was not referred to by name to avoid creating the impression of an undue and detailed interest in the series. It was verified that one-half of the 50 mothers actually had seen one or more of the Pierre pamphlets.

SOME RESULTS OF THE SURVEY REGARDING FEEDING HABITS

As an illustration of the method of evaluation used in this study, this pre-

liminary report will partially examine the results of the interviews concerning feeding habits. Obviously, the over-all effectiveness of the Pierre pamphlets cannot be judged from an examination of this limited material. It is hoped that a complete evaluation will be published at a later date.

1. *The first group of questions concerning feeding was designed to learn about the mother's course of action when a child refused to eat certain foods.* Having engaged the mother in conversation about feeding in general, the interviewer tried to find out if there was any food for which the child had shown a dislike. It was noted that 28 per cent of the 2,199 mothers said that their children had never refused a new food when it was first offered. Within this group, there were no significant differences* between the experimental and control or between the white and nonwhite groups.

The remaining 72 per cent of the mothers whose children disliked some food were queried to learn what they did when the child refused to eat that

* Tests of statistical significance used throughout were at the 5 per cent level.

TABLE 3

Food Refusals and Mother's Actions When Children Refused to Eat Particular Foods, by Race and Study Group

<i>Food Refusals and Mother's Action When Refused</i>	<i>White</i>				<i>Nonwhite</i>			
	<i>Experimental</i>		<i>Control</i>		<i>Experimental</i>		<i>Control</i>	
	<i>No.</i>	<i>Per cent of Total Disliking Food</i>	<i>No.</i>	<i>Per cent of Total Disliking Food</i>	<i>No.</i>	<i>Per cent of Total Disliking Food</i>	<i>No.</i>	<i>Per cent of Total Disliking Food</i>
Try again (after one week)	130	20.8	103	18.2	15	7.4	24	13.5
Try again (within one week)	147	23.5	124	21.9	44	21.7	24	13.5
Combine with other foods	83	13.3	81	14.3	23	11.3	22	12.4
Discontinue food altogether	120	19.2	131	23.1	70	34.5	55	30.9
Force or coaxing	142	22.7	124	21.9	51	25.1	52	29.2
Unknown	4	0.6	4	0.7	0	0.0	1	0.6
Total disliking some foods	626	100.0	567	100.0	203	100.0	178	100.0
Liked all foods	242	...	198	...	85	...	100	...
Total mothers interviewed	868	...	765	...	288	...	278	...

particular food. The mothers whose children never refused a food were asked the same question but on a supposition basis—that is, what *would* the mother do if the child refused a food.

Among the mothers for whom this was a hypothetical situation there was no difference between study groups. The results for the remaining mothers, for whom this constituted an actual problem are presented in Table 3.

In both the control and experimental groups, the white mothers followed an almost identical pattern of handling refusals, indicating that no effect could be ascribed to having received the pamphlets. For the nonwhite mothers, there was a nonsignificant trace in the experimental group toward using less forceful means and coaxing the child to eat. On the other hand, there was a greater tendency among them to discontinue the food altogether.

That the question was discriminatory in itself is evidenced by the fact that white mothers definitely tried the food again after one week rather than discontinuing it altogether, as a large share of the nonwhite mothers had done. The white mothers also used less force. Although the data on residence will not be shown here, they indicate that both

white and nonwhite urban mothers used significantly less force and tended to try the food again after a week or more.

2. The aim of the *second group of questions* was to determine if the mothers were unduly concerned about feeding, as manifested by rigid adherence to a time schedule. The interviewer not only asked the mother what time schedule she had used in feeding, but attempted to draw out this information in an indirect manner. This was necessary because many mothers vary in their interpretation of "demand feeding" or "four-hour feeding." The purpose of the questions was disguised by phrasing them as follows:

"Did your child sometimes cry between meals because he was hungry? (If 'yes') What did you do about this? Why? (If 'no') Did you ever awaken your child for a feeding? Why?"

White mothers showed no difference between the experimental and control groups, and there was a greater rate of demand feeding among nonwhite control mothers. That the questions were discriminatory is indicated by the finding that nonwhites did 50 per cent more demand feeding than whites, and rural groups more than urban ones.

3. The *third set of questions* was con-

TABLE 4

Feeding Independence at Time of Interview, by Race and Study Group

<i>Independence of Feeding</i>	<i>White</i>				<i>Nonwhite</i>			
	<i>Experimental</i>		<i>Control</i>		<i>Experimental</i>		<i>Control</i>	
	<i>No.</i>	<i>Per cent of Total Sample</i>	<i>No.</i>	<i>Per cent of Total Sample</i>	<i>No.</i>	<i>Per cent of Total Sample</i>	<i>No.</i>	<i>Per cent of Total Sample</i>
Tries spoon at most meals	328	37.8	278	36.3	100	34.7	71	25.5
Tries spoon once in a while	197	22.7	139	18.2	64	22.2	63	22.7
No spoon, but with hands	193	22.2	199	26.0	73	25.3	76	27.3
Never by himself, no interest	39	4.5	29	3.8	16	5.6	18	6.5
Never by himself, not practical *	111	12.8	120	15.7	35	12.2	50	18.0
Total	868	100.0	765	100.0	288	100.0	278	100.0

* Those mothers who said it was not practical meant that the child did not eat enough, wasted food, dirtied himself or the surrounding area.

cerned with measuring the speed of transition from one method of feeding milk to another, where the change involved any sequence of breast, bottle, and glass or cup. (The pamphlets stress a gradual change-over.) The results were tabulated according to the method of taking milk at the time of interview. There was no significant difference between study groups in the proportion of children drinking milk from a glass or cup at 14-15 months of age. Of those children already drinking some milk from a glass or cup, about 92 per cent had made the transfer gradually, with no difference between the two study groups or the two races.

4. *The fourth area of questioning concerned the procedure of leaving a baby unattended while nursing from a bottle supported by a holder or a pillow.* (The Pierre pamphlets strongly discouraged this practice.) The question was limited to its use during the first eight months of life, since many children have already started to drink from a glass or cup by that time. Moreover, this practice may be less objectionable as the baby gets older. Excluding those who were still nursing at eight months, there was no difference in the proportion of the two study groups in both races who never engaged in this practice.

5. *The final questions were directed to learning if the mother fostered the development of the child's independence, as manifested by encouraging him to feed himself as he develops.* A series of questions were used to ascertain if the mother actually encouraged the recommended practice or only gave lip service to the recommendation.

The children were rated according to their independence in feeding and the results are shown in Table 4. For white children, the degree of independence in feeding appears similar in both groups. Among nonwhite children, however, there is a slight increase in self-feeding from 76 per cent in the controls to 82 per cent in the experimental group, a difference bordering upon statistical significance. This increase in self-feeding was primarily noted in the rural children since urban nonwhite mothers were inclined to discourage it for so-called reasons of practicality.

DISCUSSION AND SUMMARY

Every health agency has an obligation to evaluate the effectiveness of its educational literature. No matter how satisfactorily a piece of literature might appear to fulfill its intended goals, there still remains the need for adequate testing. A true evaluation must inevitably

answer the question, "Does the material fulfill the objective for which it is being used?"

It seemed to the agencies concerned that a planned distribution of *Pierre the Pelican* pamphlets in North Carolina afforded an excellent opportunity to study their effect upon child-rearing practices. Accordingly, a controlled study was devised and is described in this paper. The method employed is not an original one but simply an application of sound statistical and survey methods.

The results in this preliminary report concerning feeding practices show that *Pierre's* recommendations, of themselves, had relatively little influence upon parents in North Carolina during the period of study. The only evidence in favor of *Pierre* appeared among nonwhite mothers in connection with the fostering of independent feeding habits. It is well recognized that further analysis of specific groups by age of mother, residence, or educational level may modify these conclusions.

The authors realize that the effectiveness of *Pierre*, when distributed as in this study, should not be judged solely on the basis of changes in feeding practices.

Furthermore, it is possible that this type of survey is not a true measure of different patterns of behavior between study groups, although this possibility is less likely, in view of the differences discerned between urban and rural, and white and nonwhite groups. It may be that a more complete evaluation of *Pierre* would require a series of visits to ascertain changes in patterns of behavior.

There was no question that *Pierre* was favorably received by parents. After the first three months of distribution,

self-addressed return post cards were mailed to recipients in all of the 83 nonsampled counties. Of 2,496 cards mailed, 1,265 or 50.7 per cent were returned, and every answer was in favor of continuing the series. Of those returning the cards, about 90 per cent answered that they were saving the pamphlets, and without being asked, 82 per cent volunteered the information that they had found *Pierre* helpful.

This evaluation shows that in order to affect attitudes and practices, more must be known about how to utilize the complex psychological factors which stimulate persons to action. Also, the alteration of established patterns of behavior may be too much to expect from a single instrument.

ACKNOWLEDGMENTS: It would have been impossible to undertake this study without the consultation, encouragement, and assistance of a large group of professional workers. Dr. David A. Young and Miss Dorothea L. Dolan, North Carolina Hospitals Board of Control, were responsible for making the distribution of *Pierre* possible and for suggesting the evaluation project. Dr. John A. Clausen, National Institute for Mental Health, was instrumental in both promoting the evaluation project and in designing the plan of study. He also advised on the construction of the questionnaire. The personnel of the Public Health Statistics Section, State Board of Health, who were responsible for maintaining an accurate mailing list and distributing *Pierre* were: C. R. Council, P. M. Atkins, and Glenn A. Flinchum. Gratitude is also expressed to the conscientious interviewers whose enthusiasm made this study feasible.

REFERENCES

1. Rowland, Loyd W. A First Evaluation of the *Pierre the Pelican* Health Pamphlets. *Louisiana Mental Health Studies* 1:1-23 (Nov.), 1948.
2. Knutson, Andie L. Evaluating Health Education. *Pub. Health Rep.* 67:73-77 (Jan.), 1952.
3. Knutson, Andie L. Pretesting: A Positive Approach to Evaluation. *Pub. Health Rep.* 67:699-703 (July), 1952.