The purpose of this pilot study was to test the hypothesis that the mean number of correct responses to pictorial items (line drawings) would be significantly higher than the mean number of correct responses to verbal (printed word) items. The items tested were selected from the items included in a posttest administered to 33 first grade students after they viewed a short film entitled "The Three Little Pigs," Subjects, enrolled in special courses for the educable mentally handicapped, had IQ's of from 54 to 88. An ex post facto research design was used to evaluate data collected by the Computer Based Project in the evaluation of media for the handicapped. Analysis revealed that, contrary to the hypothesis, the mean score on word items was greater than the mean on pictorial items. Because of limitations of the study, it is suggested that the exercise was more of a learning situation than an opportunity to contribute useful and meaningful experimental evidence. (WDR)
PICTORIAL VS. VERBAL STIMULI IN POSTTEST ITEMS

Harold C. Lord

ABSTRACT

The purpose of this pilot study was to analyze the difference in mean scores for pictorial (line drawings) and verbal (printed words) stimuli as evoked on a post test administered to educable mentally retarded (EMR) first graders. An ex post facto research design was used to evaluate data collected by the Computer Based Project in the evaluation of media for the handicapped. Statistical analysis of the data revealed that the mean for correct responses on word items was greater than the mean for correct responses on pictorial items. The limitations of the study are discussed, and suggestions are offered for further research.
SPECIAL REPORT No. 7319
COMPUTER-BASED PROJECT for the EVALUATION of MEDIA for the HANDICAPPED

Title: PICTORIAL VS. VERBAL STIMULI IN POSTTEST ITEMS

BY: Harold G. Lord

BACKGROUND

The Computer Based Project for the Evaluation of Media for the Handicapped, based on contract #OEC-9-423617-4357 (616) between the Syracuse (N.Y.) City School District and the Media Services and Captioned Films Branch, Bureau of Education for the Handicapped (United States Office of Education) for the five year period July 1, 1969 through June 30, 1974. The major goal is to improve the instruction of handicapped children through the development and use of an evaluation system to measure the instructional effectiveness of films and other materials with educable mentally handicapped (EMH) children, in-service training and media support for special teachers, and studies related to the evaluation process and the populations used.

The Project has concentrated on the 600 films and 200 filmstrips from the Media Services and Captioned Films (BEE-USOE) depository; however, specific packages from Project LIFE, various elementary math curricula, and selected programs from Children's TV Workshop have also been evaluated. The evaluation model used requires that: 1) objectives of materials be specified and written; 2) instruments be constructed to test and measure effectiveness; and, 3) children be the major sources of evaluation information. A number of instruments and methodologies are employed in the gathering of cognitive and affective data from 900 EMH children and 80 special teachers to make the effectiveness decisions. Over half of the EMH population can neither read or write; therefore, a unique Student Response System (SRS) is employed, consisting of a twenty station G.E.-1000 SRS which can be operated in a group or individual recording mode and is connected to a remote computer system. The computer capabilities consist of remote telephone connections to the Rome (N.Y.) Air Development Command, the Honeywell time-shared network, and the Schenectady (N.Y.) G.E. Research and Development Center; and batch mode capabilities of the Syracuse City Schools, Syracuse University, and various commercial sources.

In-service and media support activities provide on-the-job training for teachers, teacher aides, equipment, and materials to the special teachers in the city schools. The research activities have centered around investigations and special problems related to the development of the evaluation model. The four major areas considered are: 1) testing effects, 2) captioning effects, 3) special student characteristics; and, 4) evaluation procedures validation.

Documentation of the major activities appear in the five annual reports and the 600 evaluations prepared on materials used. Staff members were encouraged to prepare special reports and the attached paper is one of these. The opinions expressed in this publication do not necessarily reflect the position or policy of the Computer Based Project, the United States Office of Education, or the Syracuse City School District, and no official endorsement by any of the agencies should be inferred.
PICTORIAL VS. VERBAL STIMULI IN POSTTEST ITEMS

The practice of using pictures in the presentation of instructional materials is based, at least in part, upon the assumption that pictures clarify verbal descriptions and that they present more realistic stimuli. Yet research has shown that verbal and pictorial presentations of the same stimuli can evoke different, even conflicting responses (e.g., Samuels, 1970).

It seems generally assumed that, in terms of direct sensory appeal, stimuli presented pictorially are superior to those read in printed words or heard in spoken words. This assumption has led to numerous current practices and procedures in utilization of audiovisual materials, not only in education but also in business, industry, and mass communication. People are constantly bombarded with materials pictorially presented in books, magazines, newspapers, posters, movies, and over television. Pictures of one kind or another, either presented alone or together with written and/or spoken words, are abundantly utilized in textbooks and instructional materials under the assumption that they facilitate learning by appealing directly to sensory and cognitive reactions of the learners.
According to Bourisseau, Davis, and Yamamoto (1965), "the assumption that pictures are indeed more effective than printed words in both the presentation of information and the evocation of sensory responses is a conclusion supported more by faith than by facts [p. 250]." Most related research findings are not really relevant, for their emphasis has been on preferences of one or the other mode of presentation (Allen, 1960). Yet, on the basis of preferences and belief in the idea of pictorial superiority over words, instructional materials have been revolutionized to reflect this assumption. This idea has persisted in the absence of solid research evidence, and has achieved a prominent and seldom-disputed position in general audiovisual theory.

The purpose of this pilot study is to analyze the difference in mean scores for pictorial (line drawings) and verbal (printed words) stimuli as evoked on a post test administered to mentally retarded first graders. Hopefully from this analysis, empirical evidence will be provided to determine if, in fact, there is support for the assumption that pictorial stimuli are superior to verbal stimuli.
Review of Related Literature

Words and pictures are stimuli. What is learned from them depends on the kinds of responses that they elicit. According to May (1965), for the convenience of exposition, responses can be put into three major categories. In the first category are all preparatory responses of attention, selective discrimination, and perception which presumably have been learned previous to the presentation. Second are the responses involved in learning what the presentation purports to teach. These are called acquisition responses. In the third category are the responses involved in anchoring what has been learned so that it can be used on subsequent occasions. These are called consolidating responses. For the purpose of this study, only responses involved in the acquiring of new information, skills, concepts, and understandings will be dealt with.

Acquisition responses are specified by learning tasks (May, 1965). If the task is to reproduce the material verbatim from memory (recall), the responses required to perform it are different from those that would be required to produce evidence that the substance of the material had been
comprehended. The same is true for mere recognition responses.

Lumsdaine (1958) prepared a large number of pairs of unrelated words, pairs of unrelated pictures, pairs of pictures and words, and pairs of words and pictures. The pairs were presented in random order, 16 pairs to each list, to intermediate grade school children and to college students. The experimental conditions were varied in respect to rate of presentation and in other respects. The members of each pair were selected to minimize previously learned associations. The learning task was mainly that of rote memorizing. It was found that under all conditions and for different grade levels of students, picture-word pairs were learned best, word-picture pairs least, with word-word and picture-picture in between. Lumsdaine interpreted these findings as supporting his hypothesis that pictures are superior stimulus information for association and inferior response information, while words are superior response information and inferior stimulus information. Words are poorer stimulus terms than pictures mainly because they are apt to have more than one meaning; words are better response terms than pictures partly because most people have had a great deal
more experience in seeing an object or picture, before hearing or seeing its name (May, 1965).

Bern (1958) changed the criterion task from recall to discrimination, presenting both words and pictures in the test situation, he found that for the performance of this task, picture-picture and word-picture were superior to word-word and to picture-word. Thus it would appear that when the criterion task is changed the associations that mediate the correct responses are different. Bern's finding of better performance for conditions involving pictorial response with the introduction of pictorial testing is in agreement with the stimulus and response generalization hypothesis for presentation and testing conditions (see Kerlinger, 1964, pp. 323-324).

According to Hartman (1961) who reviewed the literature on multiple channel association learning:

To derive their hypotheses, both Lumsdaine and Bern have postulated chains of intervening sub-associations between the stimulus and the response information to be associated. Lumsdaine cites a study by Pan (1926) which indicated that better association between unrelated words could be achieved by providing words that mediated between stimulus and response. Lumsdaine's derivation is quite complex, and the following simplification is perhaps an injustice. Pictures make good stimuli because they do not evoke competing
verbal associations. Words, on the other hand, do evoke competing verbal associations, and for this reason, are poor in the stimulus position. Words, according to Lumsdaine, make better responses than pictures because they possess a greater variety of associations which can be attached to the stimulus. Second, verbal associations are less specific than pictorial associations and, therefore, are easier to attach to a given stimulus [p. 251].

The role of mediators in paired-associate learning was investigated in an experiment by Davidson (1964). The materials were pairs of pictures of objects familiar to sixty second grade school children who served as subjects. They were divided into high and low ability groups based on the results of a preliminary experiment in which they were tested for initial abilities to form associations between pairs of pictures. Each of the two groups was assigned at random to five experimental conditions. For each condition twenty pairs of pictures were exposed at the rate of a pair every five seconds. There followed two test trials during which only the stimulus picture was exposed; the task was to choose from four pictures in a workbook the one that belonged to the stimulus picture. On the second trial the order of presentation was changed. For condition A no mediators were employed. For condition B the experimenter pronounced the names of the two pictures (e.g., "chair"--"shoe") on the
presentation trials, but not on the test trials. Condition C was the same except the experimenter inserted a preposition between the pairs (e.g., "chair under shoe"). For condition D a mediating sentence was used such as "the chair doesn't look large under a big shoe." For condition E the picture showed a big shoe on a chair. The mean scores took a large jump beginning with condition C where connective mediators were introduced. For the high ability groups the jump was from a mean of 13 to 33, and for the low ability groups it was from 10 to 26. At the level of conditions D and E where sentences and pictures were the mediators the low ability groups had combined mean score almost as high as the combined means of the high ability groups. When pictures of relations (such as "a big shoe on a chair") were added to sentences, the combined mean score of both groups went up four points. The study showed that there can be highly significant facilitation as a result of introducing mediating links.

Other experimenters have been concerned with different aspects of pictorial and verbal presentations of stimuli. Karwoski, Gramlich, and Arnott (1944) found longer reaction time to pictures than to words or objects. They suggested that picture stimuli require more interpretation.
Otto (1962) found that verbal and pictorial presentations of stimuli will result in significant differences in both the number and nature of responses evoked. The response of eighty fourth graders to pictorial (line drawings) and to verbal (words) stimuli failed to support the expectation that good readers would give more responses to verbal stimuli and poor readers more responses to pictorial stimuli. It was clear that the particular stimuli involved determined the magnitude of difference between pictorial and verbal items, and not a student characteristic. Otto suggested that experimenters should be able to predict with accuracy the extent to which verbal, pictorial, or verbal-pictorial stimuli will facilitate the attainment of particular concepts. Naturally much more must be known before such prescription can be applied.

A picture of an object and the word-name of the same object do not necessarily elicit the same responses. Yet, in practice it is almost routinely assumed that a picture reinforces a verbal description and vice versa. Otto (1964) presented data which implied that the contrary may be true: a picture may in fact present clues that tend to detract from and to contradict verbal descriptions. He suggested
further studies to clarify the role of pictures in concept attainment was needed. In another study, Bourisseau, Davis, and Yamamoto (1965) concluded that pictures may well be more restrictive than words in a person's information-processing system.

The contradictory nature of existing knowledge in regard to pictorial and verbal stimuli suggests that serious further study and experimentation in this area is needed.

Objective of the Study

After a group of mentally retarded first grade students view a film and are administered a posttest, the mean of correct responses for pictorial items will be significantly higher than the mean of correct responses for word items as measured on selectively sampled items. This pilot study is to be conducted as an ex post facto research project using data made available through the facilities of the Computer Based Project of the Syracuse (New York) Public Schools.
Methods

A description of the procedures used in this *ex post facto* pilot study follows:

**Subjects**

The subjects were thirty-three (N=33) first grade students enrolled in the mentally retarded special education courses of the Syracuse (New York) Public School during the school year 1972-1973. Intelligence quotients ranged from 54 to 88 as measured on the full score composite Stanford-Binet Intelligence Quotient Test. No effort was made to include further demographic information in the study (e.g., sex, age, reading ability, etc.).

**Procedures and materials**

The subjects in a group were shown a ten (10) minute, color, 16mm film, entitled "The Three Little Pigs." The production of the film was done by Coronet Films. The content was the fable about the efforts of three little pigs to outwit a hungry wolf. The animal characters were portrayed in animation and live action. The following annotation for the film was taken from *The 16mm Film Catalog, Computer Based Project.*
The film presents the familiar story of the three little pigs, how they go out to seek their fortune, and how the third little pig cleverly outwits the big bad wolf. Real animals and people are used to portray the story along with some pictures from the story book [p. 5].

Immediately after the showing of the film a written test of ten (10) items was administered to each subject. Five of the test items were pictorial and five were verbal. Items #1, 2, 3, 4, and 6 were pictorial (line drawings) and items #5, 7, 8, 9, and 10 were verbal (recognition or recall).

The subjects responded to the test items by answering through the electronic response system available at the Computer Based Project. The responses were recorded on computer and the scores were subsequently made available through a hard copy printout.

Data

The data collected by the Computer Based Project was used in this ex post facto study. The film files were searched and reviewed to obtain those media for which pictorial test items had been created. After the media were identified (i.e., the film "The Three Little Pigs"), the student response sheet printouts for the posttest for that particular
Title were obtained and the data transcribed from them. The data were separated into pictorial and verbal items and the response scores logged for each student.

A comparison of the mean correct responses for each of the categories was computed.

<table>
<thead>
<tr>
<th>Correct Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pictorial</td>
</tr>
<tr>
<td>$\bar{X}$</td>
</tr>
</tbody>
</table>

The data was analyzed by descriptive statistics utilizing Biomedical Computer Programs for correlation with trans-generation description and tabulation (BMD02D) and stepwise regression analysis (BMD02R). Inferential statistics were used to determine the parameters of the study and the degree of error involved. The degrees of freedom ($df$) was 2-1 ($N-1$) or one (1). The analysis of variance (ANOVA) employed F-ratios at the .05 level of significance. The null hypothesis statement that there is no actual relationship between variables was stated as follows:

The mean for correct responses on pictorial items ($P$) will be equal to the mean for correct responses on word items ($W$).

$$\bar{X}_p = \bar{X}_w \quad \text{(non-directional)}$$
Limitations

This study, since it was conducted as an ex post facto pilot research project, was limited to gathering data directly involved in pictorial-word responses. Some bias may exist since only less than 10% of the data available from the Computer Based Project involved responses to pictorial test items.

No effort was made for more complete demographic information on the students. Some additional information was available in the computer files of the Computer Based Project, but this data was not readily able to be retrieved for the thirty-three (33) students used as subjects for this pilot study. Such further information would be valuable in designing aptitude-treatment interaction (ATI) studies in the future.
Results

The results from the study are presented in the following charts.

Means

<table>
<thead>
<tr>
<th>Mode</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.Q.</td>
<td>67.69696</td>
<td>6.65430</td>
</tr>
<tr>
<td>Pictorial</td>
<td>2.48485</td>
<td>1.12141</td>
</tr>
<tr>
<td>Word</td>
<td>3.15152</td>
<td>1.27772</td>
</tr>
<tr>
<td>Total</td>
<td>5.63636</td>
<td>2.14794</td>
</tr>
</tbody>
</table>

Contrary to the hypothesis the mean for correct responses on word items (3.15152) was greater than the mean for correct responses on pictorial items (2.48485).

The mean for intelligence quotient was 67.69696 with a standard deviation of 6.6543. The range of I.Q. was 54 to 88.

<table>
<thead>
<tr>
<th>Mode</th>
<th>Sum of squares</th>
<th>Mean Square</th>
<th>F Ratio</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pictorial</td>
<td>1 21.117</td>
<td>21.117</td>
<td>0.469</td>
<td>6.7102</td>
</tr>
<tr>
<td>Word</td>
<td>1 43.370</td>
<td>21.685</td>
<td>0.474</td>
<td>6.7666</td>
</tr>
</tbody>
</table>
The F-ratios of 0.469 and 0.474 were not significant at the .05 level. Therefore the null hypothesis was accepted in this two-tailed test of significance, and the research hypothesis was not supported.

Variance-Covariance Matrix

<table>
<thead>
<tr>
<th></th>
<th>I.Q.</th>
<th>Pictorial</th>
<th>Word</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.Q.</td>
<td>44.2802</td>
<td>-0.9110</td>
<td>-0.5152</td>
<td>-1.4261</td>
</tr>
<tr>
<td>Pictorial</td>
<td>-0.9110</td>
<td>1.2576</td>
<td>0.8617</td>
<td>2.1193</td>
</tr>
<tr>
<td>Word</td>
<td>-0.5152</td>
<td>0.8617</td>
<td>1.6326</td>
<td>2.4943</td>
</tr>
<tr>
<td>Total</td>
<td>-1.4261</td>
<td>2.1193</td>
<td>2.4943</td>
<td>4.6136</td>
</tr>
</tbody>
</table>

The scatter plots for pictorial-I.Q., word-I.Q., and total-I.Q. yielded no discernible pattern.

Correlation Matrix

<table>
<thead>
<tr>
<th>Mode</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pic to word</td>
<td>.6014</td>
</tr>
<tr>
<td>Pic to total</td>
<td>.8798</td>
</tr>
<tr>
<td>Word to total</td>
<td>.9089</td>
</tr>
<tr>
<td>Pic to I.Q.</td>
<td>-.0122</td>
</tr>
<tr>
<td>Word to I.Q.</td>
<td>-.0606</td>
</tr>
<tr>
<td>Total to I.Q.</td>
<td>-.0998</td>
</tr>
</tbody>
</table>
Discussion

Since this study was a pilot study only, no attempt was made to hypothesize the myriad of other plausible rival causal factors. The subjects were restricted to two mentally retarded special education classes; they had been pre-selected in this grouping by their intelligence quotient score. This limitation would render this study very questionable in terms of external validity.

This pilot study had a number of other limitations, some of which have already been stated. Because of time constraints, the experimenter chose to accept alternatives which in some instances, compromised the design of the experiment. These decisions pertained mostly to the type and quantity of data collected. Coupling the weaknesses of the pilot study with the weaknesses of *ex post facto* research in general, this exercise was more of a learning situation than an opportunity to contribute useful and meaningful experimental evidence.

Although the analysis of the data of this study refuted the hypothesis, the results tended to support the find-
ing of Otto (1964) and Samuels (1970). Much more must be discovered about the relative contributions of pictorial and verbal stimuli before the contradictory nature of existing research can be resolved.

Hanes (1973), in his summary of an article entitled "The research on how children learn from pictures," made the following comment:

A great amount of diversity exists in the research concerned with the developmental aspects of children learning from pictures. The diversity of the research findings as well as the complexity of the perceptual learning process severely limits the number of generalizations that can be made. However, the initial importance of visual processing for learning in the young child cannot be overlooked. Although the young child is limited with respect to short-term memory storage capacity and visual processing strategies, it appears that pictorial representations are extremely useful means of information communication. Increased efficiency in visual processing may be attributed to the acquisition of processing strategies, increased memory capacities, and the increased ability to utilize subtle dimensional and spatial cues in pictures. Even though developmental changes in the learner contribute to the increased learning performance, attributes of the stimulus are also influential on learning performance within developmental levels. This conclusion is most strongly supported by the research on stimulus sequencing and on processing of complex pictures that utilize a number of embedded cues [pp. 19-20].

In an attempt to summarize pictorial research, Levie (1973) contributed these comments regarding picture-word comparisons:
Pictures are superior to words as cues for eliciting specific responses. Research has rather consistently shown that pictorial recognition memory is superior to recognition memory for words, that free recall of pictures is superior to words (even though recall is demonstrated verbally rather than pictorially), and that pictures are superior stimulus items in paired-associate learning. The advantages of pictures as cues are limited by two conditions: when abstract rather than concrete subject matter is concerned, and when the learner is lacking in visual literacy as in the case of young children. The tautological nature of these two frequently mentioned limitations is apparent if they are restated thusly: Words are better than pictures for referring to things that are not readily picturable, and words are better than pictures when the learner is unable to learn from pictures. Words also appear to be superior to pictures in tasks involving sequential memory.

Picture-word comparisons of affective responses are complicated by the fact that people respond emotionally to the physical parameters of pictures as well as to their referential meaning components. That is, while affective reactions to the words "flower" are likely to be the same no matter what the visual appearance of the printed word, different pictures of different flowers may evoke quite different emotional responses, just as different artistic renderings of the "same" flower may evoke different emotional responses. The little empirical work done in this area indicates only that affective reactions to words and pictures are different, without shedding much light on the nature of the differences [pp. 38-39].

Bracht (1970) pointed out that psychologists have been too optimistic in their expectations of formulating general laws of learning and have not given sufficient attention to individual differences. He suggested "that no single instructional process provides optimal learning for all students"
He stated that investigations in aptitude-treatment interaction (ATI) are needed. Snow (1970), Salomon (1970), and Salomon (1971) concurred and offered a paradigm for future research. More recently, Davis (1973) and Schwen (1973) suggested the same conclusion, but they called the investigations aptitude-treatment analysis (ATA).

Future research should consider what different results, if any, might be obtained when the pictorial stimulus is varied systematically (e.g., addition of color, complexity and detail, photographs in lieu of drawings) as is the verbal stimulus (printed in various type styles and sizes and in color) and the picture with the accompanying word. Efforts to explore hierarchical dimensions of sensory responses to such stimuli should provide interesting data, as would studies of the range of responses to the stimuli. In addition to characteristics of the stimulus mode, attention must also be given to receiver characteristics. Useful information would include the effects of individual differences such as age, ethnic group membership, socioeconomic class, reading ability, and sex. Further investigation is needed in order to resolve the apparent contradictory nature of existing research in pictorial and verbal stimuli and subject interaction with them.
References


Computer Based Project. 16mm Film Catalog: Computer Based Project. Syracuse, New York: Computer Based Project, Prescott School, Syracuse Public Schools, November 1971. mimeograph.


Lumsdaine, A. A. Cue and response functions of pictures and words. In M. A. May and A. A. Lumsdaine (Eds.), Learning from films. New Haven: Yale University Press, 1958, 123-149.


