CANADIAN THESES

THÈSES CANADIENNES

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

The Impact of a Formative Evaluation Technology on Posttest Responses

Suzanne Daningburg

A Thesis

in

The Department

of

Education

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Arts at
Concordia University
Montréal, Québec, Canada

May 1986

ISBN 0-315-32234-9

Abstract

The Impact of a Formative Evaluation Technology on Posttest Responses

Suzanne Daningburg

Two studies examined the interaction between use or non-use of an electronic formative evaluation tool, the Program Evaluation Analysis Computer (PEAC) system to determine what effect PEAC usage has on the cognitive, affective and evaluative aspects of college students' perception of an ETV program emphasizing cognitive learning. Spatial visualization aptitude and verbal fluency aptitude were assessed as control variables. A specific real-time evaluation question was used for Study 1 and a more global question for Study 2. Cognitive scores for the short answer test in Study 2 increased from pretest to posttest for non-users and decreased for users, indicating that PEAC usage hindered cognitive recall involving comprehension. No other PEAC effects were found. The results suggest that while the PEAC system may be well suited for evaluations involving emotive aspects of ETV programs which elicit greater real-time response rates, it may not provide useful evaluation information when objectives emphasize cognitive change, and may actually interfere with learning requiring comprehension. Recommendations to producers for improved ETV productions are discussed.

ACKNOWLEDGEMENTS

Any success in the endeavor of a thesis or of life is due as much to the dedication of others as to any talent God may have bestowed. I have been privileged to know the friendship, support and kindness of many people, without whose help my efforts would be vacuous. Particularly appreciated for their contributions to this thesis are Dr. Robert Bernard and Mr. Thomas Wilson. Ms. Deena Sacks and Mr. Jay Rubenstein graciously allowed the use of their classes for the experiments, and Ms. Erika Justmann provided steadfast friendship through a trying year. For his consistent efforts which not only kept this thesis on course but made the entire experience one of true learning, I thank my supervisor, Dr. Richard Schmid.

Finally, heartfelt acknowledgement and gratitude go to my mother, Mrs. Rollande Daningburg, for her love and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

Rationale

The general concern of this thesis is how formative evaluation can be used to improve educational television (ETV). Of specific interest is the question of whether the actual process of formative evaluation influences the cognitive and affective aspects of people's viewing experience. Does the act of evaluating alter learners' perceptions such that data are not representative of the non-evaluating target population? There is a substantial literature which suggests that it does. (Coldevin, 1980; Gropper, 1967; Gropper, Lumsdaine & Shipman, 1961; Paluzzi, 1979). However, a new evaluation system has been developed which is said to overcome many of the confounding task demands which have made previous formative evaluation techniques so externally invalid (Corporation for Public Broadcasting [CPB], 1980; Nickerson, 1979; Paluzzi, 1979). The purpose of this thesis is to examine this evaluation system, the Program Evaluation Analysis Computer (PEAC), in order to begin determining what effect it has on learner perceptions.

The PEAC System and Formative Evaluation

The PEAC system was developed in the late 1970s as a formative evaluation tool. It is a portable, microcomputer-based data collection system which records second-by-second responses from viewers of television programs. It consists of about 24 battery-powered response units, each about the

size of a hand-held calculator. Each unit has 16 response keys which can be covered with various templates, such that any number of keys in various arrays are accessible and visible to the subject. Viewers may register continuous or discrete reactions (CPB, 1980). Typical questions would deal with, for example, viewers' like or dislike of various characters in a drama or cartoon, viewers' ongoing level of interest, and so on. Following an instructional warm-up period, the program is shown, and viewers simply respond, unprompted, by pressing the appropriate key. (See Baggaley, 1982a, 1982b, 1982c; Baggaley & Smith, 1982; CPB, 1981; and Radio-Quebec, 1985 for a series of formative evaluation studies employing the PEAC system and Chapter Two of this thesis for a more detailed description.)

Two examples of other methods of real-time recording of student responses during viewing are interpolated questions and conjugate analysis. Gropper (1967) and Gropper, Lumsdaine & Shipman (1961) used the former technique, interrupting the actual program to ask questions. This type of formative evaluation situation, however, is logically different from the natural act of viewing a program, because most educational television does not contain interpolated questions. Ho (1981) and Mitchell (1979) used the conjugate analysis technique to examine television viewing behavior on a moment-by-moment basis and concluded that this technique yielded fruitful ETV evaluation results.

The PEAC system marked an important change

in the application of formative evaluation in television and in its measurement techniques (see Cambre, 1981 and 1982, for a historical review of formative evaluation, and Weston, 1986, for an overview of approaches to formative evaluation). However, this very change inevitably implies yet other, perhaps more subtle influences on learning and evaluation processes. The question remains as to precisely what are the influences and effects brought about by this new measurement technique. It is clear that the act of responding to an evaluation rating question with one of the more primitive techniques, such as interpolated questions, affected the evaluation and/or the degree of learning and/or the learner's attitude about the material viewed. Has the more sophisticated methodology of the PEAC system eliminated those effects?

The need for basic research specifically investigating the external validity of the PEAC system is stated concisely by Radio-Quebec (1984), "Toutes les fois qu'on utilise le système PEAC il faut enjamber un amoncellement de questions qui sont autant d'incertitudes par rapport à la validité de cette mesure" (p. 1). It is the precision of real-time recorded data which makes the PEAC system worth further research, as described in the following excerpt:

> L'usage d'un instrument du genre de PEAC récupère un aspect de la réalité du téléspectateur, soit la réactivité immédiate des impressions en cours de

> visionnement. Celles-ci correspondent au flot de perceptions qu'une émission produit, lesquelles sont difficiles à récupérer une demie-heure plus tard, et encore plus lorsqu'il faut les traduire verbalement (Radio-Quebec, 1984, pp. 1-2).

Learner Aptitude

At present, the characteristics most often used to define a target audience are overall demographic ones, such as "adolescent" or "white urban middle class", without any reference to such factors as interest, motivation or various abilities. Why is this the case? Demographic variables such as age, sex and educational level are apparently considered to be the only relevant and accessible set of factors. It is known, however, that many factors other than demographic features profoundly influence learning (Bracht, 1970; Cronbach & Snow, 1977). Examples of such factors are individual aptitudes and entry skills.

Since these factors have been shown to influence learning, why aren't they studied and used by educational television producers? Three major reasons for this failure are suspected.

1) Cost. Determining an individual's aptitude for, say, spatial visualization, requires testing, coding and analyzing of the responses, as well as someone to administer and interpret the test. All of these things cost money (not to mention time). In contrast,

obtaining demographic levels is simpler, requiring relatively little extra expenditure and almost no time.

2) Training and orientation of television producers. The modus operandi of most educational television production to date has been "creativity" displayed in the absence of "evidence". In other words, production has been unscientific, relying exclusively on the producer's tacit knowledge of his or her craft. In addition, the relative glamor of television as an instructional medium has distracted educational designers and educational media producers from any attempt to collect empirical evidence about individual aptitudes which may be related to evaluation.

3) Technology. The means for gathering evaluative information which has demonstrable external validity have been lacking. Even with the marked improvement in data gathering techniques by systems such as PEAC, empirical evidence has yet to be gathered to build the theoretical and practical foundations of such methods to give them credibility. Only after evaluation per se makes significant inroads into the production process can variables such as learner differences be accounted for.

The above limitations notwithstanding, the question remains, does the level of formative evaluation in

educational television production need to increase, and if yes, why and how? There are two primary reasons why the answer is indeed yes.

First, it is known that learning is affected by individual differences (Cronbach & Snow, 1977). Since such factors as individual aptitude and level of entry skill influence the results of traditional (e.g., written) instructional material, it seems reasonable to suggest that these same factors affect ETV results. Television is a part of our daily life and a part of our educational system. In order to improve the quality of television as an instructional tool, it is necessary to increase the level of formative evaluation conducted. The results of formative evaluation can influence the production of programs in several ways. In terms of educational cable television, for example, remedial inserts in the form of supplementary program tapes may be made available to learners needing or desiring them. Alternatively, written materials may be created to accompany mass presentation of a program. Productions could also include cues, organizers, reviews and the like to assist those less capable while not seriously imposing on the more capable. Recent technology allows us to go even further. Videodiscs permit random access of information and "separate" programs can actually be made available for each individual learner.

The second reason why both the amount and nature of

formative evaluation must change concerns the nature of ETV. Although commercial television has developed successfully without the benefit of formative evaluation, the same cannot be claimed of ETV. Commercial television has been based on the "numbers game", that is, the greater the number of viewers, the more successful a program is considered. This system works for the lowest common denominator of audience appeal. But ETV cannot be looked upon in the same crude fashion if it is to be used within an educational context. "Popularity" must be coupled with effectiveness. And again, costs associated with superb productions are prohibitive for ETV.

The Corporation for Public Broadcasting (1980) has recognized the need for educational television research in the following statement:

> If future qualitative research efforts extend to the study of cognitive, affective and behavioral impacts of programs, correct interpretation and application of the data will require an understanding of potential interactions among such viewing responses. Varieties of qualitative response may summate, interact and/or directly compete with one another. Appropriate use of such research results will have to be guided by relevant theory and research findings. The qualitative research movement suffers, in general, from the lack of a theoretical base with which to identify key qualitative research variables

and guide analysis of their interrelationships (pp. 27-28).

Having said that the evaluation scenario needs to change, there are three key variables involved in actually implementing a change in educational television production.

1) Learner variables must be identified, assessed for relative importance or usefulness and related with program content.

2) The production system itself must become receptive to utilize any evaluative output. This means:

(a) re-educating the production system;

(b) introducing formative evaluation at the appropriate point(s) in the production sequence to enable it to influence outcome;

(c) establishing the means (the budget) to implement the changes.

(3) An educational/instructional system which fully utilizes the results of these efforts needs to be created. If, for example, certain learner abilities significantly interact with the rate and/or quality of specific content acquisition, the remedial alternatives listed above must be used by both the producers and the learners.

These suggestions deviate significantly from the commonplace view of television as a medium of passive

reception, but consider the contrast. Research on the process of reading or learning math is a vast, socially supported enterprise. Research on learning from television is negligible. One must therefore conclude that a) we are universally capable of properly utilizing the medium, or b) the medium is trivial, and not worth extensive study. To reject both these alternatives necessitates serious, scientifically based study of a medium which probably more profoundly influences more people than reading and math combined.

To summarize, before one can convince producers that evaluation needs to be integrated into the educational production system, one has to identify variables which have some influence over the final audience response. That is the purpose of this thesis. The specific questions which the thesis attempts to answer are:

1) Does the process of evaluating a television program on a moment-by-moment basis, as imposed by the PEAC system, alter the television viewing experience, hence affecting an individual's overall program evaluation rating, attitude toward content and/or cognitive recall?

2) Does a student's familiarity with the topic interact with the formative evaluation responses?

3) Does a specific individual aptitude, such as spatial visualization or verbal fluency, affect either cognitive, affective or evaluative responses

in the formative evaluation of educational television programs?

The Study

In order to begin examining the above three questions, this thesis attempts to explore the relationship between the task demands imposed by the PEAC system and the specific content being evaluated. For example, if the evaluator were to ask learners to assess the visual quality of the program, the increased attentiveness of viewers to that aspect of the content would likely enhance their subsequent recall of it, probably to the detriment of other aspects (Rothkopf, 1970; Watts & Anderson, 1971). While this attention factor is fairly obvious, the process represents a dilemma for the evaluator. The more precise, and thus prescriptive and useful the information, the more obtrusive it becomes, regardless of the measurement technique (PEAC or otherwise). The evaluator is therefore left with measuring either global responses which preserve external validity, or specific questions which by their very nature prime attention to certain aspects while interfering with others. Thus, formative evaluation must both carefully specify precisely what it is that needs to be learned, and recognize that that decision will impact to a greater or lesser extent on the generalizability of the results.

It has been stated above that one of the PEAC system's main contributions to the process of formative

evaluation is that it significantly reduces many of the obtrusive aspects of the evaluative process (Nickerson, 1979), but it is also maintained that certain "cognitive obtrusiveness" (e.g., selective attention) cannot be overcome with any existing technology (Watts & Anderson, 1971). In order to examine this effect, obtrusiveness is assessed by utilizing two levels of PEAC usage, learners using the technique and learners not using it. Assuming that attention is selective, the key to assessing differences between the groups as a result of treatment depends upon the sensitivity of the dependent measures to effects resulting both from evaluation task demands (critical information to which attention is drawn) and from aspects not directly addressed in the evaluation process (incidental information) (Anderson, 1972).

As indicated above, individual differences inevitably impact on the evaluation process, and need to be taken into account. Content familiarity is one of the cornerstones of the instructional design process, and should be a primary consideration when dealing with learning from ETV. A complicating factor in assessing the influence of content familiarity is its suspected interaction with virtually any task demands imposed by evaluation. For example, learners displaying a medium level of familiarity may be better able to implement and accurately respond to production quality evaluation questions than will their high and low familiarity

counterparts because the process of viewing and learning is neither too easy (boring) nor too difficult (distracting or confusing). At the same time, data from low familiarity learners may be equally or more useful for other purposes, such as content revision or remediation.

Given that the PEAC system necessitates a high degree of learner involvement, learner aptitudes might be critical in information acquisition. To examine the nature of these relationships, two aptitudes, spatial visualization and verbal fluency, are employed as control variables.

Spatial visualization has been defined as involving imaginary movements, transformation, or other changes in three-dimensional objects (French, 1951; Guilford & Zimmerman, 1956). Winer (1981) noted that, "This ability seems to be required in various perceptual-cognitive tasks involving the mental transformations of visual images" (p. 18). Given the reliance of television on visual images for its message design, it seems appropriate to investigate spatial visualization ability in relation to the evaluation of ETV.

Verbal fluency has been shown to be a critical factor in an individual's ability to learn from a given instructional treatment (Anderson & Kulhavy, 1971). Aptitude in verbal fluency was thus employed as a control variable in order to explore its relation to the evaluation of ETV.

Summary of Hypotheses

The study is exploratory in that it attempts to integrate several areas of inquiry which seem to be logically related. The major effect expected was that use of the PEAC system would influence cognitive recall in certain conditions. Specifically, an interaction between the Familiarity and Usage factors was expected to indicate that Usage facilitates recall for subjects who have a moderate degree of familiarity with the program topic, but decreases recall for students with high and low familiarity. This is because students whose scores indicate moderate familiarity were expected to be more conscientious while using the PEAC system and thus pay more attention. Students with high or low familiarity, on the other hand, were expected to be either bored or overwhelmed, respectively, by the combination of the cognitive task demand and their level of prior knowledge. The main effect for the presence or absence of the PEAC system was expected to produce superior incidental recall for those undistracted by the evaluation process.

Spatial visualization and verbal fluency were considered important examples of individual aptitude, both related to a student's ability to benefit from certain types of instruction. To the author's knowledge, no previous research has suggested specific interactions or effects in relation to an evaluation tool such as the PEAC system. It was, however, hypothesized that students low

in either one or both of these abilities would show greater recall when the PEAC system was used than when it was not used, because students would be compelled to pay more attention than they normally would. Students with either moderate or high ability levels were not expected to differ.

CHAPTER TWO

Review of the Literature

In the late 1930s, the emerging field of mass communications and its accompanying marketing strategies made two major contributions to media evaluation: a) an understanding of instructional media in the light of communications theory and b) new methodologies for data collection (Cambre, 1981). Despite this historical footnote, program evaluation and the development of research on formative evaluation in education is often mistakenly viewed as a recent phenomenon (Madaus, Stufflebeam, & Scriven, 1983, p. 3). In fact, it was only in 1967 that Michael Scriven coined the term formative evaluation. Evaluation research has existed far longer than it has had a name. Like most other areas of research, perspective has been dependent on the zeitgeist current at the time. A look at the historical progression of formative evaluation research shows that work has ranged from the examination of military training films in the 1920s to the analysis of children's television programs in the 1980s. Formative evaluation methodologies have often been by-products of creative experimental research techniques, and methodological patterns as well as theoretical strands have emerged.

The main concern of this review is with formative evaluation of ETV. The first section defines what formative evaluation is and the context in which it is best

used. The second section looks at the development of the PEAC system. The third section discusses educational television while the fourth section looks at individual aptitude. Finally, the relationship between the evaluation methodology, content familiarity and individual aptitude is examined.

Formative Evaluation

Although the term formative evaluation was only recently coined, concepts leading to its development began as early as the 1920s with World War I training films. What is formative evaluation? Its dynamic nature has given rise to several working definitions. Cambre (1981) defines evaluation as the act of determining the value of something, and educational evaluation as the process of determining the value of educational programs/materials. Gooler (1980) distinguishes between formative evaluation techniques and strategies. The former term includes those methods or procedures used to gather, analyze and report evaluation data, such as rating scales, interviews and regression models. Strategies, on the other hand, refer to more generic notions of the conceptualization of evaluation issues. The results of formative evaluation techniques are not generalizable in nature but are rather of direct and immediate assistance to those who have to make crucial decisions in a given situation. Nickerson (1979) has written that formative evaluation is best defined as that process whereby information representing feedback from

members of a potential target audience is acquired during the production process to help shape the program into its most effective form. A rationale for the use of formative evaluation has developed from the fact that all major models of instructional design require identification of desired outcomes and the vast majority also require the try out and revision of instruction (Dick, 1980).

One of the first large scale evaluations of an instructional film was conducted in 1921 by Lashley and Watson. Their effort was a noteworthy forerunner of formative evaluation even though carried out on a completed film because (a) authors expressed the belief that findings would be of value in planning future productions and (b) the methodologies employed greatly resembled those an evaluator would use in pretesting, such as expert judgment, testing on samples of target audience and so on (Cambre, 1981).

In terms of theoretical developments, three academic movements of the 1940s and 1950s had an impact on audio visual education and evaluation: psychometrics; the growth of psychologists' interest in evaluation activities as a form of applied research; and the emergence of the field of mass communication (Cambre, 1981). Two of the major figures in the audio visual education and evaluation field in the late 1940s were A. A. Lumsdaine and R. W. Tyler. Lumsdaine (1947) suggested that appropriate measuring instruments for determining whether a teaching film is

successful in accomplishing its intended purpose were factual tests, attitude scales and/or indices for measuring interest. He incorporated these ideas into a pretest/posttest experimental design. Tyler (1942) focussed attention on the importance of clear educational objectives in the development process. The works of both men gave some public exposure to the concept of evaluation as a continuous process and hinted at the need for a systematic approach to product development. Nevertheless, the results of evaluation research from the 1920s to the 1950s were not well publicized. For the most part they appeared in relatively obscure technical reports (Cambre, 1981).

In 1956, Hoban, then editor of the AV Communication Review, listed six reasons for the failure of evaluation research to take hold. Evaluation research, he noted, was: neither understood nor accepted as necessary; expensive and time consuming; restrictive on the researcher; not professionally rewarding; demanding in terms of skills such as human relations and group processes which were not taught in graduate training programs; and a process often leading to unwelcome and threatening negative findings. Despite these determents, however, the rapid growth of the mass communications field, with its emphasis on the audience, established a need for data collection strategies that would improve the communication process by improving methods of feedback in evaluation.

One such improvement in feedback was the Program

Analyzer, developed in 1940 to evaluate radio programs. It was the first major mechanical device to influence evaluation and it allowed for simultaneous and continuous data collections from 10 to several hundred respondents during the course of one program. Profiles were based on a statement of like/dislike, learning/not learning, or some variation thereof (Cambre, 1981). Of specific interest to this thesis is Twyford's (1954) attempt to establish the reliability and validity of these profiles by developing an audience rating profile that correlated significantly with the objective test results. He compared the effectiveness of six different rating scales for measuring learning: subjective indication; like/dislike; prediction of learning by classmates; clarity; goodness; and terminal knowledge. The most effective was found to be a scaled subjective indication such as "I am learning"; which produced a degree of accuracy comparable to testing (Cambre, 1982). Forty years later, however, in reviewing research on formative evaluation advocated for materials development, Weston (1986, p. 6) noted that, "In general, there does not seem to be an indication that any one method of gathering student feedback is superior to another, only that the materials that had undergone formative evaluation with students were superior to unrevised versions."

In 1967, Markle noted that consumers of educational products were exerting no pressure for empirical validation. She wrote, "A brilliant presentation is

defined by its impressiveness to colleagues, who can appreciate its artistry, and not by its effects on students, who may not only be unable to see why it is brilliant but, indeed, may have failed to acquire even lesser behaviors from it" (p. 105).

Geis, Burt & Weston (1984) reviewed the literature in an attempt to delineate overall, standardized guidelines for formative evaluation. Amidst a myriad of inconsistent findings, they managed to group the guidelines into a set of organizing categories: the type of learner to use in testing; the role of the learner during testing; the role of the developer during testing; and the kind of data to collect. It is the latter category which is of interest here. When discussing the kinds of data to collect, however, reference is made to records of learner/developer interactions, pretest and posttest scores, attitude questionnaires and debriefing sessions. No mention is made of continuous, real-time data, or data gathered by any electronic means.

One might well reflect today, as did Baker & Alkin in 1973, that while the number of good examples of formative evaluation is expanding, the level of research into the process is relatively limited. Expanding on the same theme, Dick (1980, p. 4) asked, "If formative evaluation procedures and techniques have been in use for almost 20 years and are included in almost every model for the design of instruction, why then is there not more research on the

effectiveness of these techniques?" Dick goes on (1980, p. 4) to list five factors which limit formative evaluation research: a) lack of funds; b) generic problems associated with conducting experimental research on the components of a total system; c) lack of a large pool of designers needed to integrate various approaches; d) perceived lack of time on the part of organizations which use systematic design; and e) personal factors related to receiving negative feedback from formative evaluation results.

The "feedback" notion of communication models greatly influenced the programmed instruction movement which was closely tied to Skinnerian learning theory. The notion of feedback was integral to the process. Commonalities between the programmed instruction movement and earlier film research studies included the guidance of student responses through appropriate stimulus techniques. The rules that promised to improve instructional materials so impressively were intricately tied to the behavioral learning theories of Skinner (1957), who believed that educational subject matter could be analyzed logically into a number of small steps. Much of the formative evaluation research was conducted with paper and pencil instruction and was considered an outgrowth of the early research on programmed instruction (Dick, 1980, p. 5).

Concurrent with the development of formative evaluation methodologies and theories was the rapid development of computer-based evaluation technologies, and

the increasingly widespread use of non-traditional educational media such as television, video and computer. This in turn led to evaluation questions which were more focussed. How, for example, can the cognitive and attitudinal impact of educational television upon specific learners be improved through formative evaluation? What factors would influence a specific formative evaluation rating? The need for formative evaluation of educational products dramatically increased, and with it rose the need for more sophisticated formative evaluation tools.

The PEAC System

The development of computer-based evaluation methodologies led to greater emphasis on scientifically rigorous formative research and evaluation. In the late 1970s a new electronic technique, known as the Program Evaluation Analysis Computer (PEAC) system, was commissioned for development jointly by the Ontario Educational Communication Authority and the Children's Television Workshop. The system, consisting of a series of manual response units programmed to record momentary changes in audience response, marked a significant change in the application of formative evaluation and in its measurement and quantification techniques. Produced as a formative research tool, the main objectives of its development were to meet both the time and the cost restrictions of the production process (Nickerson, 1979). In addition, the tool "overcomes perhaps the most

intransigent problem of all - that of taking communication research methods out of the laboratory and into the field, where all can understand them" (Baggaley, 1982a, p. 70). Coupled with these advantages is a problem common to all untried new discoveries which is the unknown aspect of the methodology itself. The paramount importance of the need for basic research about the PEAC system is stated by Radio-Quebec (1984), "Ces questions auxquelles nous apportons une réponse pratique sur la base d'intuitions réfléchies nous laisseront des interrogations aussi longtemps qu'elles n'auront été vérifiées systématiquement" (p. 2) and again, "Ainsi, à nos incertitudes par rapport à la mesure elle-meme et à son caractère dérangeant s'ajoutent des incertitudes au niveau de l'analyse quantitative des données. Cette situation invite donc à des vérifications méthodologiques de base" (p. 4).

Certainly the PEAC system allows a quality and degree of evaluative detail never before possible. Color graphics data reporting techniques present viewer reactions time-locked to the programming which evoked those reactions. This is important because, as Baggaley (1982c, p. 70) notes, "Audience reaction can shift and change from one moment to the next, and overall reactions to a programme can be due to an isolated moment within it which even its producer cannot predict." Fine-grain analyses are possible, with visual representations of viewer involvement, comprehension and/or enjoyment directly linked

to program dynamics (e.g., specific script lines, appearances of a given personality, scene or format change). Data are stored and accrued by a central microcomputer on floppy diskettes and are retrievable through hard copy or color video display. These data reporting methods, including the superimposition of real-time viewer responses over an entire program or segment, are specially designed for meaningful input to program production decisions (Nickerson, 1979).

For formative research purposes, the device is usually used in a group setting, with a group leader overseeing the viewing and simultaneous evaluation of programming material. However, with appropriate panel instruction, PEAC units can be geographically distributed, with viewers taking units home for program evaluation. Respondents would have an electronic qualitative diary, entering program evaluation judgments on this device, instead of in a traditional paper and pencil diary. At the end of the diary period, PEAC devices could be picked up by field interviewers and data from each device aggregated in a central computer for analysis and reporting.

Conflicting hypotheses may be made regarding the influence of the PEAC system on learning and attitude. Using Festinger's & Maccoby's (1964) work on the theory of cognitive dissonance, it could be argued that the PEAC system reduces comprehension and recall. Inversely, it is possible that the task, by forcing participants to evaluate

the content continuously, acts as a reinforcement and increases the attention of the respondent. The problem is recognized in this statement by Radio-Quebec (1984, p. 2):

> La décision d'utiliser le PEAC s'appuie également sur le fait que le comportement de visionnement de la télévision implique un réactivité très immédiate. Ce comportement demande une participation réduite des processus cognitifs et en particulier de ces processus auxquels on fera appel dans la séance de recherche: formuler ses impressions verbalement, remplir un questionnaire, analyser une émission en fonction de certains aspects précis, etc. Il y a donc intéret à capter les réactions selon un mode qui requiert peu de participation des processus cognitifs et expressifs, ce que permet dans une certaine mesure le système PEAC.

The degree of influence of the PEAC system on the type and level of learning involved seems logically to be related to the task demands imposed by the formative evaluator. For example, rote learning would be improved by a real-time question dealing with verbatim recall of material presented, while comprehension would be facilitated by a real-time question requiring comprehension. The task demands are cognitive. This is an extension of the work of Watts & Anderson (1971).

Regarding attitude, several psychological theories may explain either a positive or a negative response bias. For

example, Festinger's (1957, 1962) theory of cognitive dissonance and the work of Zimbardo, Cohen, Weisenberg, Dworkin & Firestone (1965) on the role of distractors in communication situations would predict an increase of content appreciation for PEAC users. The respondent, who must devote more energy to comprehend the message, justifies this effort by attributing more value to it and giving the program a more positive evaluation. On the other hand, it is possible that the more difficult task entails a certain amount of frustration that results in a more negative evaluation.

Caron, Giroux & Mayrand (1984, p. 3) note two factors exemplifying the simplicity of the PEAC methodology as compared to the complexity of the stimulus. First, real-time data garnered is based on only one overall scaled evaluation question, due to the structure and technology of the system, even though the object of evaluation is multidimensional. Second, traditional data gathering techniques such as questionnaires have a discrete number of questions which are determined by the researcher. This means that, for each dimension studied, there are an equal number of responses. With the PEAC system, on the other hand, the measured dimension is continuous. Although the researcher preprograms the system to accept responses at particular time intervals, it is the respondent who ultimately decides the moment and frequency of his or her response. It is impossible to distinguish between the

viewer whose reaction is stable from one whose attention wanders and/or one who simply forgets to respond.

Theoretically, "any measure of audience response may be collected via the system, from interest value and clarity, to complex factual responses prompted via multiple-choice routines" (Baggaley, 1982b, p. 20). The possibility of the task demand (the overall evaluation question) influencing the data collected, however, has led most researchers using the PEAC system to simplify the task demand as much as possible. Thus, most evaluations collect only responses based on global real-time questions. Even this simplification is not without problems. A global judgment on a like/dislike scale must be considered a multidimensional variable. Yet what factors underlie a particular judgment? Caron, Giroux & Mayrand (1984, p. 7) provide the example of two subjects expressing the same response at the same time point. How is one to know, they ask, whether each subject's judgment is elicited by the same content elements? An individual's response may be in reaction to a character, the decor, the values, and so on. The considerable loss of information makes the opinions expressed difficult to interpret. To understand the effects of television, Baggaley (1985b, p. 7) stresses the usefulness of attribution theory, which "examines man's desire to attribute causes to the events and phenomena he encounters, and the bases on which his attributions may be predicted." In reference to attribution theory in an

earlier work, Baggaley (1980, pp. 110-111 and pp. 162-165) lists four factors to be examined in an evaluation of television's effects: 1) individual differences, due to psychological and social factors, between the responses of audience members; 2) differences over time, in the responses of individual viewers; 3) technical effects, such as differences in audience response due to the manner in which televised material is mediated; and 4) parallel influences upon audience responses by the wide range of alternative factors thought capable of such effects.

In formative evaluation research with the PEAC system, Baggaley has used both general (1982a, 1982b) and specific (1985a) questions to collect real-time data. In all cases only one overall evaluation question was used. The reason for this latter choice is at least partly technical. At the present time, the response analyzers are not capable of aggregating real-time data on more than one question at a time. Of the 16 keys on each hand-unit, only 5 -- the letter keys A through D -- are meant for continuous data. The remaining keys, numbered 1 through 11, may be used for responses to questionnaires. Due to the fact that only one real-time question may be asked of a viewer, Baggaley stresses the importance, for most purposes, of collecting data with as broad and far-reaching an effect as possible, that is, information based on a global question. Exceptions to this occur when the purpose of the formative evaluation study is the evaluation of a specific factor,

such as the comparison of a range of characters used in current educational media programs for children which Bagaley recently examined (1985a).

One solution to the limitations imposed by a single overall evaluation question would be to have subgroups of the sample respond to different and more focussed evaluation questions. This, however, raises the issue not only of whether a viewer could effectively concentrate on only one particular, predetermined aspect of the program, but also of the results such a concentration would have on the cognitive and affective responses. Even the evaluation of a specific program aspect might be contaminated by, for example, a general attitude toward the whole program. In addition, the artificially imposed focus enforced on the viewer might modify his or her general comprehension.

Since only one real-time question is possible, an ideal representation of real-time questions would necessitate a separate sample for each of several questions, including, for example, questions of a general nature as well as questions focussing on particular details of the program. While in theory this is a good idea, in practical terms it is often difficult to supply one appropriate sample group, let alone the many that would be needed for a design of this nature. The most parsimonious solution would seem to lie in determining precisely what should be evaluated and constructing the task demand(s) as well as the post-viewing questionnaire(s) to reflect this.

The evaluation scenario represents a strange circumstance of research design. The experimental factor -- the independent variable -- is usually known, with the dependent variable being the subject's response. With the PEAC system, on the contrary, the experimental factor must be inferred after the collection of data. That is, the researcher must determine what in the program provoked the responses. The attention thus is shifted back to the educational program.

Educational Television

The idea that student feedback could be used to improve the effectiveness of educational television programs was investigated as far back as 1961, when Gropper, Lumsdaine & Shipman showed that the effectiveness of televised instruction could be improved based on student responses to post-viewing achievement tests. This study did not deal with evaluation rating by viewers per se but rather focussed on an evaluation derived from cognitive recall of televised material. This is important in the context of this thesis because it suggested that the effects of educational television could in fact be improved through formative evaluation. The authors argued that post-viewing achievement tests were inherently suitable as a formative evaluation measure because they did not destroy the continuity of the program. Essentially, the study exemplified the try out and revision process. In a study several years later, Gropper (1967) investigated whether

cued recall questions during the program itself increased the program's effectiveness. The study demonstrated that the success of programmed television instruction was influenced not only by program design, but by active responding (during viewing) on the part of the audience. In this now dated study, however, active responding consisted of the subject's answering of questions interpolated throughout the viewing sequence. Scores on a post-viewing achievement test were higher for students who had had to respond to questions at various points during the program than for students who did not have interpolated questions. In other words, although questions interpolated into the program altered the continuity of the program, they also increased learning effectiveness.

Requiring a response of the learner is frequently an attribute of programmed instruction and may take the form of overt, covert, or reading modes, or a combination of the three. The claim is made that overt responses require greater attentiveness on the part of the learner. Other studies indicate that covert responses may yield comparable learning gains with considerable time savings (Jurgemeyer, 1982, p. 20). Although the situation described in Gropper's study cannot be considered typical of educational television as it exists today, it did raise questions which to date are not uncontentiously answered. Does any type of learner response during viewing affect cognitive recall? Does any response affect learner evaluation ratings or

learner attitude toward the content?

Much of the work that has been done on the evaluation of educational television has tended to focus on user surveys, a priori analyses and recommendations, and occasional studies employing performance on a standardized achievement test as the dependent variable (see Chu & Schramm, 1979, for an overview of research on ETV). Little emphasis has been placed by most investigators on the heart of the matter -- the behavior of the learner as it relates to the attainment of specified educational objectives. (Orr, 1966, p. 364). Although written 20 years ago, Orr's (1966) advocation of the necessity of theory building still applies today:

> Once standard and dependable evaluation techiques have been developed to a sufficient degree, it will then, and only then, become possible to subject educational telecasts to scientific study involving the experimental determination of the effectiveness of various program content and techniques in achieving specific educational goals. Only then will it be possible to resolve the conflicting conclusions regarding the value of educational television which presently abound in the literature (p. 369).

Formative evaluation is used to determine whether or not the educational objectives of a program are being accomplished. In order to provide an accurate indication of the degree of attainment of them, therefore, the

educational objectives must be clearly defined and criteria for their assessment agreed upon. The importance of precisely defining the terms used cannot be overemphasized. For example, Romiszowski (1981, p. 43) distinguishes between goals and objectives in the following manner. An educational aim, goal or purpose may be stated in terms of inputs or processes, such as teaching or solving problems, while an educational objective is more precisely stated in terms of outputs or products (p. 43). Formative evaluation may be concerned with both processes and products. The present study explores the processes involved in the evaluation of a given product. The educational objectives and the criteria for their assessment are designed for the program, which is a product.

What are the educational objectives of a program and how are they determined? The difficulty most often is not with the writing of the objective(s), but with the achievement of agreement among teachers on a particular set of behaviors being indicative of its presence at a determined level of intellectual knowledge or skill (Romiszowski, 1981, p. 86). Part of this difficulty results from a real divergence of opinion as to the aims of education, but part has also been caused by communications problems, because of sloppy concepts and misunderstood terminology. Educational objectives may be classified in a variety of ways. Romiszowski (1981) cites three of the most common: immediacy (long-term/immediate); performance-

type (cognitive/affective/psychomotor); and source (life-skill/methodological/content). This study deals with the immediate objectives of program content, as measured by two performance-type objectives, namely cognition and attitude. The overall cognitive objectives of the present program are, using Bloom's categories (cited in Romiszowski, 1981) to provide knowledge and comprehension about the material presented and the ability to apply, analyze, synthesize and evaluate the basic concepts of the material presented.

Although it is necessary to establish a test standard in order to determine whether the instructional objectives have been met as effectively as planned or hoped for, such criteria do not absolutely provide proof of amount of learning. It is possible, for example, that a student chooses to respond incorrectly for any one of a variety of reasons which could include affective factors. For example, Anderson (1972) noted that an individual's ability to process meanings (and hence, by implication, to perform on a given criterion) can fail for reasons such as fatigue, boredom, time pressure and/or difficult material.

Within the total evaluation, the relative importance of attitude objectives is of concern because attitudes can influence an individual's cognitive responses. In addition, a positive or negative attitude toward the topic of the program may well influence an individual's rating of the program itself. It is difficult for any but the most highly trained professional evaluator to separate how one

reacts to a given event (or instruction, etc.) from how one feels about it. The direction of the influence depends logically on the specific topic. Therefore, it is important that the program have affective goals and criteria to measure them.

In an educational context, an attitude is defined by Gagne (cited in Romiszowski, 1981, p. 190) as an acquired internal state that influences the choice of personal action. Romiszowski indicates that Gagne is "critical of the use of the term 'affective domain' (as used by Bloom and Krathwohl in their taxonomies) in that it emphasizes the emotional component of attitudes at the expense of their cognitive or behavioral aspects" (p. 190). It may be argued that feelings are essentially unteachable. It is possible, however, to transfer or attach certain specific feelings to certain specific events by, for example, classical or operant conditioning techniques. Whatever method is used, the environment is manipulated in order to condition the learner to exhibit the desired attitude when faced with a given circumstance. Often short-term attitudes are controlled or transformed by long-term values or goals. As Romiszowski (1981, p. 226) writes:

> Whatever the nature of the goal, the problem for the person is the same, namely to control his own behaviour so that it leads to the long-term goals, despite the fact that in the short term this may not be the easiest or the most pleasant path to follow.

Those who do this well may quite justifiably be said to have developed a skill - a skill of self-control, or of 'self-actualization' (to use Carl Rogers' terminology), or of living according to their value system (to use a term from Krathwohl's and Bloom's taxonomy of the affective domain).

It is interesting to note that the types of behavior described above have the characteristics of skills, that is, they require perception, decision, action and evaluation. In this sense, the categories of the affective domain's taxonomy do not describe different kinds of attitudes, but rather outline the different stages or levels of affective development. As such, it seems most logical that the measures of attitude focus on the processes rather than on the outcome of attitude development. In practice, however, it is difficult to set criteria for processes which are, by their very definition, in progress. An example of an outcome question is: "Did the program promote the desired attitude?"; while a process question would be: "Did the program contain the materials and activities it was supposed to?" The shift from outcomes to processes requires that a portion of the evaluation be designed to determine whether the classroom instruction adequately reflects the theory upon which the program is based. The formative evaluation is then conducted to monitor whether the program "looks and operates as it is supposed to" (Henerson, Morris & Fitz-

Gibbon, 1978, p. 17). This information is used to keep the program on track. The (immediate) attitude objectives (and measures) for the program used in this thesis were selected to focus on outcome or product. Also note that this decision must be related to whether or not attitude objectives can be realistically achieved by the time of measurement. When they can, then measurement of outcome is appropriate, whereas evaluation of processes should be emphasized when the objectives are not expected to be achieved by the time of testing. For the present study's program, "Out of the Mouths of Babes", the main affective objective is to provoke a demonstrated interest (that is, an interest indicated by some overt behavior) in language acquisition in children. This objective can be categorized in the first and second categories of Krathwohl's taxonomy, which are receiving and responding, respectively. The other three categories, valuing, organization and characterization by a value or value complex, cannot realistically be attained in one lesson, but would require a series of lessons to determine. Hence they do not form objectives of the present study.

Individual Aptitude

The specification of objectives in and of itself does not guarantee that all learners will benefit. Bracht (1970, p. 627) notes:

> Given a common set of objectives, some students will be more successful with one instructional program and

other students will be more successful with an alternative instructional program. Consequently, a greater proportion of students will attain the instructional objectives when instruction is differentiated for different types of students.

The problem of discovering interactions between individual characteristics (aptitudes) and instructional treatments is not new, and is treated at length in what is known as Aptitude x Treatment Interaction (ATI) research (Bracht, 1970; Bracht & Glass, 1968; and Cronbach & Snow, 1977). Snow (1978) defined aptitudes within an educational context as "student characteristics that predict response to instruction under a given instructional treatment" (p. 227).

This recognition of audience receptivity based on group heterogeneity must be considered one of the cornerstones of formative evaluation. It is especially important when dealing with educational television because of the time and cost restrictions of this production process. In order to maximally improve ETV and benefit all learners to the utmost, it is essential to consider individual learner characteristics when evaluating a given educational television program. But how can this best be done?

According to modern theory and research in text processing, one's original intention while reading determines what information will be processed. Information

retained which is not central to the educational objectives is termed incidental learning. The intention one has in mind when processing the material can be influenced by advanced goal statements (Klauer, 1984, p. 324). Intentions are assumed to have a motivating effect, which would lead to a greater amount of material learned. However, these same intentions may selectively influence what material is actually retained. Material relevant to the intentions is learned, while material irrelevant to them is not.

It is helpful for a moment to look at past attempts to improve learning by steering, or directing, the learner. Rothkopf (1966) argued that questions asked during the course of a lesson shape the student's attention and processing. According to this mathemagenic hypothesis, asking questions before a person reads a passage should cause him or her to search for the answers to these questions and therefore would not be expected to facilitate performance on new questions. Mathemagenic activities are those student activities that are relevant to the achievement of specified instructional objectives in specified situations or places (Rothkopf, 1970, p. 327). The concept of mathemagenic activity implies that the learner's actions play an important role in determining what is learned. That is, what a student does with a particular instructional material influences to a large extent what he or she acquires from it (learning).

It is important to note, however, that until the early to mid 1970s, most mathemagenic research dealt with verbatim recall rather than with comprehension (Anderson, 1972). Watts & Anderson (1971) argued that inserted questions requiring comprehension would in fact induce meaningful processing of text and thus improve performance on new criterion test questions which also entailed comprehension. Groups which received inserted questions that required them to apply the principooooooo new examples performed significantly better on a posttest than all other groups, including groups which received inserted questions that repeated examples described in the text. The mathemagenic research cited above relates to the effect of imposing overt manipulation on given instructional programming. The assumption is that a manipulation such as inserted questions does indeed affect learner response to the program. The ATI research relates to the effect of such overt manipulation on individual learners.

Implications of the Research

In formative evaluation, it is essential to understand the effects of manipulated or artificial treatments, not only in terms of possible applications, but, more importantly, in light of their influence on basic research methodology. Use of a tool as extraneous to the educational program as the PEAC system is may be viewed as a tampering with, or manipulation of, the educational proc[illegible] itself. Does the use of the PEAC system function

in a manner similar to that, say, of Rothkopf's inserted questions? That is, does it enhance or hamper cognitive recall or affective response? Does its effect depend solely on the overall evaluation question asked? This study was designed to explore these questions.

CHAPTER THREE

Method

Design

A 2 Usage (use or nonuse of Peac) x 3 Content Familiarity (high, medium and low) fixed design was used, with both factors between subjects. The pretest used to determine content familiarity was used again as the posttest, leaving open the possibility of including a repeated measures comparison. Aptitudes in both spatial visualization and verbal fluency were assessed for use as control variables.[1] The same design was replicated on two populations, described below as Study 1 and Study 2.

Subjects

The subjects in Study 1 were 55 students from a private English-language college in Montreal. For Study 2, 69 students from a public English-language Montreal college were used. Since subjects' prior knowledge of content was assessed via a pretest rather than manipulated, random assignment of groups was preserved, thus avoiding the limitations associated with an *ex post facto* design.

Materials

The instructional stimulus consisted of a 27-minute ETV program entitled "Out of the Mouths of Babes", produced in 1975 by the Canadian Broadcasting Corporation (CBC) and previously aired on the weekly "The Nature of Things" television series. The program, recorded on a 3/4-in. video cassette, examined the theory that a child's ability

to talk is partially innate; that babies are born with the capacity to extract not just words, but rules of grammar, from what they hear spoken around them. This particular program was selected for three reasons. First, its content and level of production were appropriate for college-level students. Second, it was believed that the content would be familiar to some students and unfamiliar to others, allowing an adequate number of learners in each level of content familiarity. Third, the program was considered a good representation of a "typical" ETV program in that it was of relatively high production quality and it dealt with an academic, though popular subject matter.

The pre-experiment measures consisted of a two-part prior knowledge cognitive test (content familiarity pretest), an attitude questionnaire, a spatial visualization test and a test of verbal fluency. Prior knowledge was assessed by two composite measures, one based on 6 short answer items and the other on 10 multiple choice items. The items tested levels of knowledge, comprehension and application of program content (see Appendix A). The attitude measure consisted of 14 scaled Likert-type questions which were related to subjects' attitudes toward the importance and relevance of the topic of language acquisition in children (see Appendix B). Each attitude item dealt with a discrete topic, and was therefore treated individually.

The results from the spatial visualization and verbal

fluency tests had been designated as potentially useful covariates. These abilities were measured because different individual aptitudes in these areas were thought to affect cognitive performance and possibly attitude towards either the visual medium of ETV and/or the content of the program. Spatial visualization and verbal fluency scores were also examined in light of their ability to identify a possible interaction between treatments.

Spatial visualization aptitude was assessed by the Spatial Visualization Test of the Dailey Vocational Tests (Dailey, 1965) which measures the ability to visualize objects presented two-dimensionally in three dimensions. The test consists of pairs of drawings, one of which is a flat pattern and the other a solid object made by folding the flat pattern. Small arrows on each pattern and its object indicate exactly the same position on both. Subjects must decide which of certain marked edges on the solid object correspond to the same edges on the flat pattern (see Appendix C). The Wide Range Vocabulary Test (French, Ekstrom & Price, 1963) was used to determine verbal fluency. It tests knowledge of word meanings by presenting one word, followed by five other words. Respondents must decide which of the five alternatives approximates most closely the meaning of the stem word (see Appendix D).

The introduction, instructions for each of the pre-treatment measures, information about program viewing in

one week, and the request to complete a post-viewing questionnaire were printed and delivered verbally to all groups by the experimenter (see Appendix H). This pre-treatment oral presentation was identical for all groups. For the experimental sessions, which took place exactly one week later, printed instructions were again used by the experimenter for an oral presentation to introduce herself, request that students watch the program, and give instructions regarding each of the post-viewing questionnaires. In addition, those groups using the PEAC system were instructed verbally from printed instructions in the use of the PEAC devices (see Appendices I and J for instructions to non-usage and usage groups, respectively).

For Study 1, all subjects were told that the overall question to be used in evaluating the program was: "How effective do you think this program is in demonstrating language acquisition in children?" Subjects not using the PEAC system simply were asked to keep the question in mind as they viewed the program. Subjects using the PEAC system were asked to rate the program in terms of this overall question, using their hand-held devices. Possible responses were: "very effective"; "fairly effective"; "not very effective"; and "very ineffective". It was initially intended to utilize the same overall (real-time) question for both samples. However, during Study 1 it was found that the interaction between this question and the characteristics of the program failed to provide for

discriminating evaluation. Nearly all subjects using the PEAC system rated the film as "very effective" or "fairly effective" at all times and in fact used their hand-held PEAC devices very infrequently. The decision was made to change the overall question for Study 2. All subjects for Study 2 were told that the overall evaluation question was: "How good do you think this program is, based on your overall reaction?" Once again, subjects not using the PEAC system were asked to keep this in mind during viewing. Subjects using the PEAC system were asked to: "Rate the program as good, fairly good, not very good or poor. Use your own judgment to make the evaluation, based on anything that strikes your overall reaction at any given moment." It was hoped that the change of question would accomplish two goals. First, the more global nature of this question would encourage students to be more discriminating about different aspects of the program (J. P. Baggaley, personal communication, January 25, 1986). Second, the simpler, more understandable phrasing would make the task easier and thus encourage more frequent responses.

For both studies, in classes using the PEAC system, the four possible real-time ratings were written on the blackboard as a reminder to subjects throughout the program.

The post-presentation dependent measures consisted of five parts. The cognitive and attitude measures given in the pre-presentation session were readministered (see

Appendices A and B). An additional section was attached to the cognitive component of the posttest. This section consisted of 11 "fill-in-the-blank type" questions designed to tap incidental learning (see Appendix E). It was thought that this measure might provide a more precise indication of the relative obtrusiveness or unobtrusiveness of the PEAC system. Subjects' own opinion about the effectiveness of the program was sought through 20 program evaluation rating questions (see Appendix F). These were different from the attitude measures in that the program evaluation questions tapped opinions about the presentation of this specific program whereas attitude items probed opinions related to the general topic -- that is, language acquisition -- apart from the ETV program. Finally, 16 demographic questions ascertained information such as age and sex (see Appendix G).

Procedure

Study 1 involved three intact groups, two experimental (n = 16 and n = 13), and one control (n = 26). Study 2 also involved three intact groups, two experimental (n = 23 and n = 17) and one control (n = 29). The three groups in each of the studies consisted of different sections of the same course, and were taught by the same instructor. The students had been placed in a given section based on their college's scheduling system. Any potentially different learner characteristics were assumed to have been distributed to classes in a more or less random fashion,

and condition assignment was made to most closely equalize n-sizes in conditions (the two smaller classes were the experimental groups and the largest class was the control group).

The first session for all groups was held during the first class period of the winter term. The experimenter was introduced by the regular instructor, who informed the class that no one was obliged to participate, and that the experiment was not in any way related to their actual class. The experimenter then provided a brief introduction about the study, and explained that any questions related to the purpose and rationale of the study would be answered at the end of the experiment. She also told the class that results would be made available to the instructor in several weeks time.

After answering any permissible questions, large envelopes containing the three pre-treatment questionnaires were distributed. Students were asked to write the last four digits of their telephone number in a box provided on the envelope, as a means of coding responses while ensuring anonymity. Students were asked to remove all three questionnaires from their envelope and set aside the second and third. The experimenter then asked students to begin by reading to themselves the instructions printed on Questionnaire 1, and then write their answers to the questions in the space provided. Students were reminded to complete all sections of the first questionnaire, but not

to go on to Questionnaire 2 until told to do so. They were told they would have 15 minutes to answer Questionnaire 1 (which included the cognitive and attitude items). The experimenter reminded the class when only two minutes remained. When the 15 minutes had elapsed, students were asked to put aside the first questionnaire and look at the second one. The experimenter read aloud the instructions for Questionnaire 2, which was the Spatial Visualization Test, and informed students that their time limit was 10 minutes and that once again they would be reminded of the time when 2 minutes remained. The experimenter reminded the class when only two minutes remained. After 10 minutes had elapsed, students were asked to set aside Questionnaire 2 and begin the last questionnaire, which was the Wide Range Vocabulary Test. Students were asked to read the instructions to themselves and begin. They were told of the 10 minute time limit and were reminded when only 2 minutes remained. For all questionnaires, students were told to put up their hand if they had any difficulty and the experimenter would help them individually. When all three questionnaires had been completed, students put them back in the envelope and envelopes were collected. Students were told that the second session would occur in one week.

The experimenter returned in one week. All groups were given identical instructions introducing the presentation and informing subjects that they would be

asked to answer a number of questions after viewing. The experimental group also received a standard five-minute session on the use of the PEAC system (including an explanation of the overall real-time question as mentioned above). This procedure was one which is most commonly used in formative evaluation situations employing the PEAC system (Baggaley, 1982a, p. 4). When any permissible questions had been answered, the 27-minute program was shown. Immediately following the program, envelopes were distributed to the students. Students were asked to write the last four digits of their phone number on the envelope, to take out all three questionnaires and to set aside the second and third. The experimenter then asked students to begin by reading to themselves the instructions printed on Questionnaire 1, and then answer the questions. Students were reminded to complete all sections of the first questionnaire, but not to go on to Questionnaire 2 until told to do so. They were told they would have 15 minutes to answer Questionnaire 1 (which again included the cognitive and attitude items). The experimenter reminded the class when only two minutes remained. When the 15 minutes had elapsed, students were asked to put aside the first questionnaire and look at the last two questionnaires. They were told that since there would be no time limit for the last two questionnaires, they could begin the second one (which was the program evaluation ratings) and go on to Questionnaire 3 (demographic

variables) when they had finished Questionnaire 2.

The experimenter explained that although each student would be allowed to leave the classroom for a break when he or she had completed Questionnaires 2 and 3, she herself would be available to answer any questions once all students were finished. Students were asked to replace all three questionnaires in their envelope and place the envelope on a table near the door on their way out of the classroom. When all students were finished, the experimenter thanked them for their participation and answered any questions the students asked.

CHAPTER FOUR

Results

Scoring Procedures for Dependent Variables/Aptitude Tests

The spatial visualization test and the verbal fluency test were scored for total correct responses, as prescribed in their respective manuals. Each of the 33 items on the spatial visualization test had a value of one point for a correct response. Each of the 48 items on the verbal fluency test had a value of one point for a correct response.

The two cognitive tests, the short answer format and the multiple choice format, were used as measures of content familiarity in the pretest. The six items of the short answer test were assigned 2 points each and scored as correct responses, with 12 points possible. Partial credit (1 point) was given to responses containing less than the total concepts required. Strict application of model answers was used. The multiple choice test was also scored for the number of correct responses. Each of the 10 items was assigned 1 point, for a total of 10 possible points.

Due to the potentially subjective nature of scoring some of the short answer items, an inter-rater reliability was calculated. To obtain inter-rater reliability results, a selection of short answer tests was scored by two individuals, neither of whom had any knowledge as to the tests' group membership. An inter-rater reliability coefficient of .91 was obtained.

The attitude section, designed to measure attitude about various aspects of the topic of language acquisition in children, consisted of 14 Likert-type questions. Each question was coded individually, giving a value from one to five to each of the responses.

As mentioned in Chapter Three, an attempt was made to gather real-time data via the PEAC system during the actual viewing of the program. This was done in order to determine whether any differences in frequency or nature of real-time responses were related to responses made on the criteria (dependent variables). However, malfunction of the data collection units or their programming prevented real-time data frequency counts for Study 2. Note that failure of the units in no way influenced the independent variable of PEAC.

The first two cognitive posttests consisted of the same two that were used in the pre-treatment session, and were scored the same way. The third test, designed to further explore the relation between use of the PEAC system and different types of learning, consisted of 11 incidental questions of a "fill in the blank" type. One question (#1) was worth 5 points, another (#11) was worth 3 points, with the remaining nine questions worth 2 points each, for a total of 26 points possible. Partial credit (1, 2, 3 or 4 points for #1; 1 or 2 points for #11; and 1 point for the remaining 9 questions) was given to responses containing fewer than the total concepts or words required. Strict

application of model answers was used.

The attitude section of the post-treatment test was identical to that of the pre-treatment session and was scored in the same manner.

The items in the program evaluation rating section, designed to measure opinion about the actual program, varied in nature, with some (Section I, #1 - 3) having a short answer format, some (Section II, #1 - 4) a multiple choice format, and others (Section III, #1 - 7 and Section IV, #1a and #1b) a Likert-type scale format. Some questions (Section II, #4, Section III, #2, #4 - 7 and Section IV, #1a and #1b) allowed for a mid-scale response (5-point scale), others (Section II, #1 - 3, and Section III, #1a - 1e and #3) were of a forced-choice nature (4-point scale). Responses to the three short answer questions were not analyzed statistically due to the variety of possible answers, none of which could be deemed correct or incorrect. Responses for all other questions were coded and given a value of from 1 to 4 or 5, depending on the number of possible choices.

Various demographic information was also collected and tabulated.

Tests of Assumption

Prior to the analyses of effects, several tests of assumptions were necessary. Due to the need to employ intact groups within each study, equivalence of groups was first assessed. Analyses of variance (ANOVAs) were

completed on the three groups of each study on all five pre-session dependent measuring instruments (the spatial visualization, verbal fluency, cognitive short answer, cognitive multiple choice and attitude tests). No differences emerged in any comparisons, thus verifying group equivalence.

It was also necessary to test for homogeneity of regression for analyses involving covariates. The test of interaction showed no effect for either covariate in Study 1. However, because the verbal fluency interaction was found to be significant ($F(1,65)=4.88$, $p=.031$) in Study 2, it was not used in any analyses for that study as a covariate, but was treated as a blocking variable.

It was initially thought that the content of the program would provide a broad, discriminating range of prior knowledge. However, results of the cognitive pretests in both studies yielded a narrow and somewhat low level of achievement. Use of this variable would therefore be statistically dubious given standard measurement error, so content familiarity was dropped as a useful independent variable in both studies.

Study 1

Study 1 involved data from the sample drawn at the private college only. As will be seen below, differences in the samples of Study 1 and Study 2 beyond design and surface (private vs. public schools) considerations emerged in all pretest performances, further amplifying the need

for separation of the two groups initially thought to possibly be equivalent.

Cognitive Tests

Means and standard deviations for the two covariates -- spatial visualization and verbal fluency -- and the three cognitive variables -- short answer, multiple choice and incidental learning scores -- are listed in Table 1. Results of multivariate and univariate analyses of variance or covariance yielded no significant results.

Attitude Tests

Since each of the attitude items dealt with different components of the program and content, they were not clustered to form composite scores. Rather, they were examined individually using non-parametric analyses. Neither Chi-square tests nor Mann-Whitney U tests yielded any significant differences between groups for either the attitudes expressed on the pretest or those expressed on the posttest. Medians, means and standard deviations are shown in Table 2.

Despite the ordinal nature of the data, a repeated measures analysis of variance was conducted in order to determine whether or not any changes occurred from the pretest to the posttest between the treatment and the control groups. Due to the nature of the repeated measures analysis, pairs with at least one missing case were deleted from the analysis. A total of 7 cases were deleted from the experimental group, reducing the n size to 22 from 29,

Table 1

Cognitive and Aptitude Measures For Study 1

Dependent variable	With PEAC			Without PEAC		
	$\bar{X}$	SD	n	$\bar{X}$	SD	n
Spatial visualization	23.14	6.05	29	22.42	6.37	26
Verbal fluency	26.17	8.45	29	27.08	6.20	26
Pretest						
Short answer	3.83	1.42	29	3.50	1.45	26
Multiple choice	3.10	1.57	29	2.81	1.50	26
Posttest						
Short answer	4.38	2.24	29	3.77	1.56	26
Multiple choice	7.90	1.52	29	8.00	1.50	26
Incidental	15.59	2.77	29	16.12	3.30	26

Table 2

Medians, Means and Standard Deviations for All Attitude Items For Study 1*

Question	With PEAC		Without PEAC	
	Pre-test	Post-test	Pre-test	Post-test
1) It is important to understand how an individual learns a language.	1.71 1.66 (.55)	1.91 1.86 (.52)	1.83 1.92 (.94)	1.94 2.04 (.87)
2) Knowing how children learn to talk makes one more socially attractive.	3.94 3.86 (1.01)	3.87 3.79 (.86)	3.70 3.73 (.83)	3.75 3.77 (.71)
3) Children are usually an interesting topic.	1.69 1.64 (.56)	1.83 1.82 (.67)	1.75 1.85 (.88)	1.88 1.89 (.71)
4) Knowing about how children learn language will help a person to teach their own children to speak more quickly and effectively.	2.00 2.11 (.96)	2.25 2.52 (1.09)	1.83 1.89 (.82)	2.38 2.46 (1.10)
5) At least one course in child development would be useful for everyone -				
a) for those who plan to have children.	1.67 1.63 (.57)	1.77 1.72 (.59)	1.31 1.46 (.71)	1.87 1.85 (.61)
b) for those who don't plan to have children.	2.50 2.58 (.86)	2.55 2.59 (.84)	1.97 2.12 (.99)	2.19 2.42 (.95)
6) It is unfair to expect people who are not parents to be interested in children.	3.89 3.70 (.99)	3.73 3.52 (.95)	3.79 3.46 (1.14)	3.89 3.73 (.96)

Table 2 - continued

7) Documentary prgrams on various aspects of child development should be shown regularly on televsion.	2.04 2.28 (1.10)	3.68 3.62 (.78)	2.00 2.17 (1.01)	3.23 3.15 (.83)
8) The importance of how an infant learns his or her mother tongue is exaggerated.	2.27 2.42 (.83)	3.59 3.48 (.63)	2.83 2.83 (.87)	3.57 3.39 (.80)
9) Knowing whether speech is culturally or biologically determined is irrelevant.	2.35 2.38 (.58)	2.14 2.21 (.68)	2.71 2.71 (.69)	1.89 1.96 (.92)
10) Parents should use correct grammar when speaking to their children.	2.14 2.25 (.79)	4.28 4.21 (.90)	2.00 2.08 (.83)	4.29 4.31 (.62)
11) Knowing about how children learn language will help a person to be a better parent.	3.75 3.52 (1.08)	2.71 2.79 (1.18)	3.00 3.21 (1.25)	3.00 3.04 (.96)
12) The problem with television is that all the information goes by so quickly that you can't remember very much.	2.44 2.70 (1.19)	3.82 3.55 (.99)	3.00 3.00 (1.14)	3.83 3.69 (.84)
13) Television is a better way to learn about topics like language acquisition because it makes everything concrete and interesting.	3.19 3.13 (.97)	2.67 2.66 (1.01)	3.20 3.21 (.83)	2.14 2.23 (.82)

Top digit is the median, second digit is the mean and () is the the standard deviation.

while 2 cases were deleted from the control group, reducing the n size to 24 from 26. The adjusted means and standard deviations used in the repeated measures analysis are presented in Table 3. None of the response distributions was statistically significantly skewed, thus enforcing the interpretability of the results. Results showed significant main effects for repeated measures for Questions #4 ($F(1,44)=11.19$, $p<.002$), #5a ($F(1,44)=17.60$, $p<.001$), #7 ($F(1,44)=33.66$, $p<.001$), #8 ($F(1,44)=24.04$, $p<.001$), #9 ($F(1,44)=12.70$, $p<.001$), #10 ($F(1,44)=250.16$, $p<.001$), #12 ($F(1,44)=16.64$, $p<.001$), and #13 ($F(1,44)=29.10$, $p<.001$). Three of the eight items (#4, 5a, and 8) show that both groups expressed a more negative attitude about the importance or relevance of language acquisition in children after viewing the program, while results for Questions 9 and 10 indicated the opposite. Both groups were more likely to think that television was a better medium (#12) and a better way to learn about topics like language acquisition (#13), but the groups expressed more negative attitude about documentary programs (#7) after viewing.

Program Evaluation Rating Tests

As mentioned in the Scoring Procedures for Dependent Variables section above, the three program evaluation rating short answer items of Section I were not statistically analyzed. A tally of the responses, however, indicated that a majority of 40 out of 55

Table 3

Adjusted Means and Standard Deviations For All Attitude Items for Repeated Measures Analysis For Study 1*

Question	With PEAC		Without PEAC	
	Pre-test	Post-test	Pre-test	Post-test
1) It is important to understand how an individual learns a language.	1.64 (.58)	1.77 (.53)	1.96 (.95)	2.08 (.88)
2) Knowing how children learn to talk makes one more socially attractive.	3.91 (.97)	3.86 (.94)	3.79 (.83)	3.83 (.70)
3) Children are usually an interesting topic.	1.64 (.58)	1.77 (.53)	1.92 (.88)	1.92 (.72)
4) Knowing about how children learn language will help a person to teach their own children to speak more quickly and effectively.	2.14 (.99)	2.64 (1.14)	1.88 (.80)	2.50 (1.10)
5) At least one course in child development would be useful for everyone -				
a) for those who plan to have children	1.64 (.58)	1.82 (.59)	1.50 (.72)	1.88 (.61)
b) for those who don't plan to have children	2.55 (.74)	2.59 (.80)	2.17 (1.01)	2.46 (.98)
6) It is unfair to expect people who are not parents to be interested in children.	3.91 (.87)	3.50 (1.06)	3.63 (1.01)	3.79 (.93)

Table 3 - continued

7) Documentary programs on various aspects of child development should be shown regularly on television.	2.18 (1.10)	3.59 (.80)	2.17 (1.01)	3.08 (.83)
8) The importance of how an infant learns his or her mother tongue is exaggerated.	2.45 (.86)	3.50 (.67)	2.83 (.87)	3.38 (.82)
9) Knowing whether speech is culturally or biologically determined is irrelevant.	2.41 (.59)	2.18 (.73)	2.71 (.69)	2.00 (.93)
10) Parents should use correct grammar when speaking to their children.	2.18 (.73)	4.27 (.70)	2.08 (.83)	4.38 (.58)
11) Knowing about how children learn language will help a person to be a better parent.	3.50 (1.10)	2.77 (1.19)	3.20 (1.25)	3.08 (.97)
12) The problem with television is that all the information goes by so quickly that you can't remember very much.	2.73 (1.20)	3.68 (.95)	3.00 (1.14)	3.71 (.86)
13) Television is a better way to learn about topics like language acquisition because it makes everything concrete and interesting.	3.18 (.96)	2.64 (1.00)	3.20 (.83)	2.20 (.83)

* Top digit is the mean, () is the standard deviation.

students thought that the program's greatest strength (#1) was the use of children as examples, while 20 out of 55 students found the greatest weakness (#2) was the use of the narrators themselves. Other responses to and tallies for Question #1 included: the program was informative and clear, 10; and the topic itself was interesting, 4. Other responses to and tallies for Question #2 included: the program had no weaknesses, 5; incomprehensible verbal responses from children, 3; too clinical, 2; and boring, 2. Twelve students did not answer this question. Responses to and tallies for Question #3 -- the best way to improve the program -- varied, with 15 students not answering; 5 saying it was fine the way it was; 5 suggesting the addition of more examples; and the remainder offering a variety of unique comments ranging from "add graphics" to "direct the program to parents" to "outline the important points at the beginning and end of program" to "edit out the narrators as 'talking heads'".

Each of the other 17 program evaluation rating items dealt with different aspects of the actual program. Therefore, each item was examined individually. Response frequencies are presented in Table 4. A Chi-Square test was performed for each variable to determine whether differences existed between the treatment groups. No significant differences were found for any of the questions.

Table 4

Frequencies of Program Evaluation Ratings For Study 1

Section	Question	Responses	Frequencies	
			With PEAC	Without PEAC
II	1) How effective do you think this program was as an introduction to the topic of language acquisition in children?	a) very effective	14	7
		b) fairly effective	15	18
		c) not very effective	0	1
		d) very ineffective	0	0
	2) For yourself, would you judge the content of the program as being	a) very effective	11	5
		b) fairly effective	17	17
		c) not very effective	1	4
		d) very ineffective	0	0
	3) Do you feel that too much or too little information was presented?	a) way too much	0	0
		b) too much	4	0
		c) just right	15	16
		d) too little	10	9
		e) way too little	0	1
	4) Would you recommend to others that they watch this program?	a) strongly recommend	3	1
		b) recommend	16	11
		c) neutral	10	14
		d) discourage	0	0
		e) strongly discourage	0	0
III	1a) narration	a) excellent	5	3
		b) good	15	13
		c) fair	7	9
		d) poor	2	1
	1b) content	a) excellent	12	4
		b) good	15	17
		c) fair	2	4
		d) poor	0	1
	1c) continuity	a) excellent	5	3
		b) good	20	15
		c) fair	4	5
		d) poor	0	3

Table 4 - continued

1d) pacing	a) excellent	4	2
	b) good	14	10
	c) fair	11	9
	d) poor	0	4
1e) children used as examples	a) excellent	21	16
	b) good	7	7
	c) fair	1	2
	d) poor	0	1
2) The photography in the program was aesthetically	a) excellent	2	2
	b) good	14	11
	c) average	11	8
	d) fair	2	4
	e) poor	0	1
3) "Talking heads" (narrators talking directly to the camera was a technique used in the program. Do you think is was used	a) far too often	1	1
	b) too often	8	6
	c) just the right amount	19	18
	d) too rarely	1	1
	e) far too rarely	0	0
4) The content was logically presented and well-organized.	a) strongly agree	4	2
	b) agree	18	19
	c) neutral	7	4
	d) disagree	0	0
	e) strongly disagree	0	1
5) In general, the content was easy to remember.	a) strongly agree	3	1
	b) agree	23	20
	c) neutral	3	5
	d) disagree	0	0
	e) strongly disagree	0	0
6) On a continuum from dry to fascinating, how would you rate this program?	a) very dry	0	1
	b) dry	1	2
	c) average	7	10
	d) slightly fascinating	16	13
	e) fascinating	5	0
7) If this were a weekly series, and you "accidentally" watched this program, would you	a) definitely tune in	2	1
	b) maybe tune in	19	10
	c) neutral	3	5
	d) probably not tune in	5	6
	e) definitely not tune in	0	4

Table 4 - continued

IV	1a)	I learn more from an educational television program if I have specific questions to answer beforehand because I only concentrate on the questions.	a) strongly agree	3	2
			b) agree	12	12
			c) don't know	4	5
			d) disagree	9	6
			e) strongly disagree	1	0
	1b)	I learn less from an educational television program if I have specific questions to answer beforehand because I only concentrate on the questions.	a) strongly agree	0	2
			b) agree	5	6
			c) don't know	6	7
			d) disagree	17	11
			e) strongly disagree	1	0

Real-Time Data from PEAC

Subjects used their hand-held PEAC devices an average of 1.24 times, after the initial setting, per student for the entire program. Response frequencies for real-time data are presented in Table 5.

Demographic Variables

Demographic data are shown in Table 6.

Study 2

The results of Study 2 were based on data from students at the public college. Data were tested in three sets of analyses (one for each type of criterion variable category) in the same manner as Study 1.

The first of the three sets of analyses examined the short answer, multiple choice and incidental recall scores. Next, subjects' separate attitudinal responses were analyzed for overall group differences and for changes from pre-treatment scores to post-treatment scores. Third, subjects' program evaluation ratings were examined for group differences, determined by students' responses to questions probing opinions about various production aspects of the program. The specific tests and scoring procedures are identical in all respects to those of Study 1.

Cognitive Tests

Means and standard deviations for spatial visualization and verbal fluency scores, as well as those for the three cognitive variables -- short answer, multiple choice and incidental recall -- are listed in Table 7. A

Table 5

Real-Time Response Frequencies (PEAC Data) For Study 1

Response	Total frequency
A) Very effective	29
B) Fairly effective	51
C) Not very effective	7
D) Very ineffective	2

Table 6

Frequencies of Demographic Variables for Study 1

Question	Response	Frequencies With PEAC	Frequencies Without PEAC
1) Sex	Male	3	10
	Female	26	16
2) Age	Under 18 yrs	15	3
	18-20 yrs	14	23
3) Parental status	Is a parent	0	1
	Not a parent	29	25
4) School Major	Humanities	7	3
	Sciences	5	10
	Business	1	10
	Fine Arts	14	3
	Other	2	0
5) Previous course in psychology	Yes	18	18
	No	11	8
6) Hours per week of ETV viewing	Less than 1 hr	14	11
	1-2 hrs	[illegible]	12
	3-4 hrs	3	3
7) Hours per week of commercial TV viewing	Less than 1 hr	7	3
	1-2 hrs	4	4
	3-4 hrs	10	9
	5-6 hrs	5	5
	More than 6 hrs	3	5
8) Owns VCR	Yes	21	20
	No	8	6

Table 6 - continued

9) Hours per week of movie viewing on VCR (if applicable)	Less than 1 hr	5	4
	1-2 hrs	10	8
	3-4 hrs	6	7
	5-6 hrs	1	1
	More than 6 hrs	7	0
10) Most important reason for TV viewing	Entertainment	17	13
	Relaxation	5	8
	Learning	2	2
	Relief from boredom	1	2
	Keeping up socially	2	1
11) Considers TV a good instructional medium	Yes	28	21
	No	1	2

Table 7

Cognitive and Aptitude Measures For Study 2

Dependent Variable	With PEAC			Without PEAC		
	$\bar{X}$	SD	n	$\bar{X}$	SD	n
Spatial visualization	15.70	7.48	40	16.89	7.26	29
Verbal fluency	18.63	6.87	40	18.38	5.38	29
Pretest						
Short answer	2.65	1.48	40	2.35	1.52	29
Multiple choice	1.90	1.15	40	2.36	1.25	29
Posttest						
Short answer	2.05	1.65	40	3.41	1.99	29
Multiple choice	5.60	1.78	40	6.18	1.49	29
Incidental	13.23	4.46	40	13.66	5.02	29

multivariate analysis of covariance (MANCOVA) (Spatial Visualization only), using the short answer and the multiple choice scores of the pretest and the posttest as dependent variables, resulted in a significant Hotellings test ($F(2,65)=6.12$, $p=.004$) for the main effect of PEAC. No other overall effects were significant. Within the PEAC effect, univariate analyses showed the short answer scores to be significant ($F(1,66)=10.26$, $p=.002$), while the multiple choice scores only approached significance ($F(1,66)=3.5$, $p=.066$), both on the posttest only. Table 8 represents the multivariate and univariate comparisons for these dependent measures.

The analyses showed that, based on the short answer posttest scores, non-users of PEAC (the control group) outperformed users of PEAC (the experimental group). Subsequent analyses using various alternative combinations of variables yielded the same main effect for PEAC. These analyses included: the above multivariate analysis (MANOVA) and univariate analyses which dropped the covariate; analyses which used content familiarity in spite of its narrow range (the pretest short answer scores) as a control factor (three-level blocking); and analyses which used either spatial visualization or verbal fluency as a control factor (three-level blocking).

The above MANOVAs provided essential information regarding the central hypothesis of the study, namely, that use of the PEAC system influences cognitive recall.

Table 8

Multivariate and Univariate Tests on Short Answer Pretest (SPR), Multiple Choice Pretest (MPR), Short Answer Posttest (SPO), and Multiple Choice Posttest (MPO) Scores For Study 2

Effect	Multivariate test (Hotelling's)			Univariate tests Error Mean Square: SPR - 2.27 MPR - 1.39 SPO - 3.24 MPO - 2.84			
	F	df	p		F	df	p
PEAC	3.73	(1,63)	.009	SPR -	.63	(1,66)	.429
				MPR -	1.89	(1,66)	.174
				SPO -	10.26	(1,66)	.002
				MPO -	3.50	(1,66)	.066

However, a true test of the repeated measure became useful with the loss of prior knowledge as an independent variable. That is, rather than using pretest performance as a blocking variable, overall pretest performance was used as a baseline and compared to posttest performance. The data were thus submitted to a BMDP2V repeated measures analysis of covariance with the short answer pretest and posttest scores as the repeated measures and PEAC as the treatment factor. The results produced a significant interaction ($F(1,66)=11.6$, $p=.001$). While both experimental and control groups performed at the same level on the pretest, the control group, which was not distracted by the PEAC evaluations, improved significantly (Tukey, $q(67)=4.06$, $p<.01$) and the experimental group actually performed slightly worse on the posttest! On the posttest comparison, the control group also performed significantly better than the experimental group (Tukey, $q(67)=5.56$, $p<.01$). A graphic representation of this interaction appears in Figure 1.

No significant effects were found for incidental learning.

Attitude Tests

As in Study 1, attitude items were examined individually using non-parametric analyses. Mann-Whitney U tests on each item revealed significant differences between groups on the pretest for Items #2 ($U=306.0$, $p<.009$), #4 ($U=286.5$, $p<.004$), #9 ($U=300.0$, $p<.007$), and #11 ($U=324.0$,

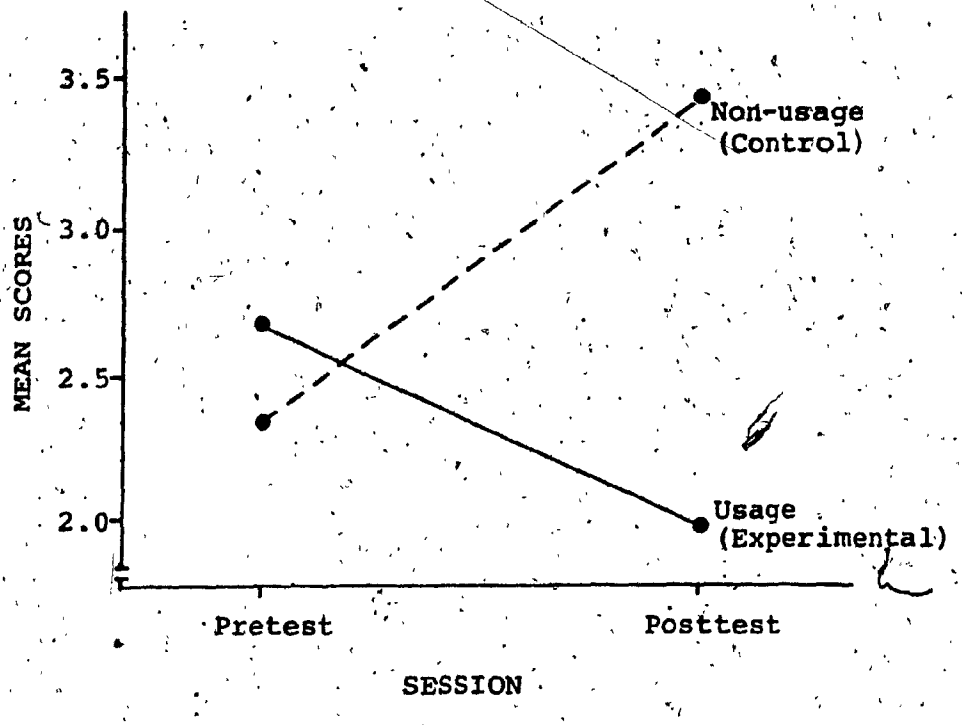


Figure 1. Mean scores on short answer pretest and posttest for usage and non-usage of PEAC for Study 2.

p=.04). Medians, means and standard deviations for all attitude items are shown in Table 9. Results showed that on the pretest the experimental group attached more value to the importance of language acquisition in children than did the control group. No significant differences between groups were found for attitudes indicated on the posttest. (A more conservative Chi-square analysis produced the same results.)

In order to assess changes in attitudes from pretest to posttest between groups, a repeated measures analysis of covariance was performed. Due to missing data, 12 cases were deleted from the experimental group, reducing the n size to 28 from 40, while 6 cases were deleted from the control group, reducing the n size to 23 from 29. The adjusted means and standard deviations used in the repeated measures analysis are presented in Table 10. None of the response distributions was statistically significantly skewed, thus enhancing the interpretability of the results. Results showed significant main effects for the repeated measures for items #1 ($F(1,49)=4.45$, p=.04), #7 ($F(1,49)=49.55$, $p<.001$), #8 ($F(1,49)=50.97$, $p<.001$), #10 ($F(1,49)=192.11$, $p<.001$), and #12 ($F(1,49)=18.22$, $p<.001$). Results from items 8 and 10 indicate that after viewing the program, students in both the experimental and the control groups expressed more negative attitude toward the importance and/or relevance of language acquisition in children. For item #1, however, students in both groups

Table 9.

Medians, Means and Standard Deviations for All Attitude Items For Study 2*

Question	With PEAC		Without PEAC	
	Pre-test	Post-test	Pre-test	Post-test
1) It is important to understand how an individual learns a language.	1.82 2.00 (1.03)	1.77 1.77 (.71)	1.88 2.08 (1.12)	1.88 1.90 (.72)
2) Knowing how children learn to talk makes one more socially attractive.	2.79 2.69 (1.08)	3.07 2.90 (1.18)	3.61 3.40 (.76)	3.60 3.24 (1.02)
3) Children are usually an interesting topic.	1.39 1.67 (.93)	1.65 1.77 (.87)	1.91 1.96 (.79)	1.84 1.90 (.82)
4) Knowing about how children learn language will help a person to teach their own children to speak more quickly and effectively.	1.41 1.55 (.72)	1.82 1.97 (1.04)	2.04 2.12 (.83)	1.85 2.03 (1.05)
5) At least one course in child development would be useful for everyone -				
a) for those who plan to have children	1.53 1.54 (.56)	1.68 1.79 (.88)	1.77 1.88 (.93)	1.80 1.83 (.76)
b) for those who don't plan to have children	2.36 2.53 (1.05)	2.41 2.51 (.85)	2.50 2.67 (.87)	2.46 2.63 (1.01)
6) It is unfair to people who are not parents to be interested in children.	3.96 3.76 (.91)	3.80 3.39 (1.41)	3.73 3.37 (1.08)	3.91 3.69 (1.17)

Table 9 - continued

7) Documentary programs on various aspects of child development should be shown regularly on television.	1.89 2.21 (1.32)	3.96 3.90 (.72)	2.12 2.28 (1.02)	3.72 3.55 (.83)
8) The importance of how an infant learns his or her mother tongue is exaggerated.	2.00 2.03 (.72)	3.76 3.69 (.73)	2.46 2.60 (.82)	3.65 3.38 (.86)
9) Knowing whether speech is culturally or biologically determined is irrelevant.	2.50 2.50 (.83)	2.25 2.36 (.74)	2.88 2.80 (1.00)	2.56 2.69 (.93)
10) Parents should use correct grammar when speaking to their children.	2.63 2.58 (.86)	4.47 4.44 (.64)	2.10 2.17 (.70)	4.26 4.35 (.48)
11) Knowing about how children learn language will help a person to be a better parent.	2.83 3.03 (1.33)	2.89 2.85 (1.04)	2.35 2.76 (1.07)	2.88 2.90 (1.15)
12) The problem with television is that all the information goes by so quickly that you don't remember very much.	2.41 2.60 (.99)	3.61 3.23 (.99)	2.14 2.33 (.92)	3.54 3.21 (1.15)
13) Television is a better way to learn about topics like language acquisition because it makes everything concrete and interesting.	2.95 2.92 (.95)	2.38 2.54 (.79)	2.50 2.79 (.93)	2.39 2.66 (1.01)

* Top digit is the median, second digit is the mean, and () is the standard deviation.

Table 10

Adjusted Means and Standard Deviations for All Attitude Items for Repeated Measures Analysis For Study 2*

Question	With PEAC		Without PEAC	
	Pre-test	Post-test	Pre-test	Post-test
1) It is important to understand how an individual learns a language.	2.00 (1.05)	1.64 (.49)	1.96 (.98)	1.83 (.65)
2) Knowing how children learn to talk makes one more socially attractive.	2.82 (1.06)	2.89 (1.17)	3.39 (.78)	3.21 (1.00)
3) Children are usually an interesting topic.	1.71 (.98)	1.82 (.94)	1.87 (.69)	1.96 (.71)
4) Knowing about how children learn language will help a person to teach their own children to speak more quickly and effectively.	1.54 (.74)	1.96 (1.00)	2.09 (.85)	2.09 (1.04)
5) At least one course in child development would be useful for everyone -				
a) for those who plan to have children	1.57 (.57)	1.71 (.76)	1.87 (.97)	1.83 (.72)
b) for those who don't plan to have children	2.61 (1.10)	2.46 (.88)	2.70 (.88)	2.74 (.96)
6) It is unfair to expect people who are not parents to be interested in children.	3.71 (.98)	3.46 (1.40)	3.39 (1.08)	3.65 (1.19)

Table 10 - continued

7) Documentary programs on various aspoects of child development should be shown regularly on television.	2.07 (1.30)	3.86 (.65)	2.26 (1.05)	3.48 (.79)
8) The importance of how an infant learns his or her mother tongue is exaggerated.	2.04 (.74)	3.79 (.69)	2.57 (.84)	3.39 (.94)
9) Knowing whether speech is culturally or biologically determined is irrelevant.	2.61 (.79)	2.36 (.83)	2.74 (1.01)	2.91 (.85)
10) Parents should use correct grammar when speaking to their children.	2.54 (.92)	4.36 (.68)	2.17 (.72)	4.35 (.49)
11) Knowing about how children learn language will help a person to be a better parent.	3.21 (1.40)	2.79 (.99)	2.78 (1.09)	3.09 (1.16)
12) The problem with television is that all the information goes by so quickly that you can't remember very much.	2.43 (.96)	3.07 (1.02)	2.35 (.93)	3.09 (1.16)
13) Television is a better way to learn about topics like language acquisition because it makes everything concrete and interesting.	2.92 (.94)	2.50 (.84)	2.78 (.95)	2.65 (.98)

* Top digit is the mean, () is the standard deviation.

showed that they attached more value to the importance of the topic of language acquisition after viewing. Both groups showed a more favorable attitude toward the medium of television after viewing (#12), and a more negative attitude toward documentary programs (#7).

Program Evaluation Rating Tests

As for Study 1, the three program evaluation rating short answer items of Section I were not statistically analyzed. A tally of the responses showed that a majority of 50 out of 69 viewers thought that the program's greatest strength (Question 1) was the number of examples of children used, while 16 out of 69 students felt the program's greatest weakness (Question 2) was the nature of the narrators themselves. Other responses to and frequencies for to Question 1 included: clarity of presentation, 5; and the topic of language acquisition itself, 8. Other responses to and frequencies for to Question 2 included: no weaknesses, 9; too much repetition, 5; and boring, 3. Responses to the best way to improve the program (Question 3) varied, with 23 students not answering this question, 8 saying it was fine as is, 7 suggesting better narration and 7 suggesting more examples of children.

Chi-square tests were performed on all of the other variables to determine whether differences existed between the treatment groups. The response frequencies are presented in Table 11. Results showed a

Table 11

Frequencies of Program Evaluation Ratings For Study 2

Section	Question	Responses	Frequencies With PEAC	Frequencies Without PEAC
II	1) How effective do you think this program was as an introduction to the topic of language acquisition in children?	a) very effective	19	10
		b) fairly effective	18	18
		c) not very effective	3	0
		d) very ineffective	0	0
	2) For yourself, would you judge the content of the program as being	a) very effective	18	12
		b) fairly effective	22	14
		c) not very effective	0	1
		d) very ineffective	0	1
	3) Do you feel that too much or too little information was presented?	a) way too much	1	0
		b) too much	4	1
		c) just right	25	24
		d) too little	10	3
		e) way too little	0	1
	4) Would you recommend to others that they watch this program?	a) strongly recommend	6	3
		b) recommend	22	12
		c) neutral	10	12
		d) discourage	2	1
		e) strongly discourage	0	0
III	1a) narration	a) excellent	6	4
		b) good	16	11
		c) fair	14	10
		d) poor	4	2
	1b) content	a) excellent	8	6
		b) good	27	19
		c) fair	5	2
		d) poor	0	0
	1c) continuity	a) excellent	4	3
		b) good	18	13
		c) fair	16	8
		d) poor	2	2

Table 11 - continued

1d) pacing	a) excellent	4	4
	b) good	16	10
	c) fair	16	8
	d) poor	4	4
1e) children used as examples	a) excellent	36	19
	b) good	4	5
	c) fair	0	2
	d) poor	0	1
2) The photography in the program was aesthetically	a) excellent	0	0
	b) good	21	12
	c) average	14	11
	d) fair	4	1
	e) poor	1	3
3) "Talking heads" (narrators talking directly to the camera) was a technique used in the program. Do you think it was used	a) far too often	3	0
	b) too often	14	3
	c) just the right amount	22	21
	d) too rarely	1	3
	e) far too rarely	0	0
4) The content was logically presented and well-organized.	a) strongly agree	4	1
	b) agree	27	18
	c) neutral	4	6
	d) disagree	5	2
	e) strongly disagree	0	0
5) In general, the content was easy to remember.	a) strongly agree	0	2
	b) agree	27	16
	c) neutral	10	8
	d) disagree	3	1
	e) strongly disagree	0	0
6) On a continuum from dry to fascinating, how would you rate this program?	a) very dry	0	1
	b) dry	5	3
	c) average	17	16
	d) slightly fascinating	15	5
	e) fascinating	3	2
7) If this were a weekly series, and your "accidentally" watched this program, would you	a) definitely tune in	9	1
	b) maybe tune in	16	12
	c) neutral	8	4
	d) probably not tune in	6	9
	e) definitely not tune in	1	1

Table 11 - continued

IV	1a) I learn <u>more</u> from an educational television program if I have specific questions to answer beforehand because I only concentrate on the questions.	a) strongly agree		4	2
		b) agree		15	16
		c) don't know		3	4
		d) disagree		14	5
		e) strongly disagree		4	0
	1b) I learn <u>less</u> from an educational television program if I have specific questions to answer beforehand because I only concentrate on the questions.	a) strongly agree		2	0
		b) agree		11	5
		c) don't know		7	5
		d) disagree		16	16
		e) strongly disagree		4	1

significant difference ($\chi^2(3, N=67) = 8.96$, $p=.03$) for item #3 of Section III. The results show that the PEAC users were more dissatisfied with the "talking heads" narration than non-users.

Demographic Variables

Demographic data are presented in Table 12.

Table 12

Frequencies of Demographic Variables for Study 2

Question	Response	Frequencies With PEAC	Frequencies Without PEAC
1) Sex	Male	14	14
	Female	26	15
2) Age	Under 18 yrs	1	6
	18-20 yrs	26	23
	21-23 yrs	6	0
	24-25 yrs	2	0
	Over 25 yrs	5	0
3) Parental status	Is a parent	1	1
	Not a parent	39	28
4) School Major	Humanities	1	4
	Sciences	37	5
	Business	1	17
	Fine Arts	1	1
	Other	0	2
5) Previous course in psychology	Yes	9	29
	No	31	9
6) Hours per week of ETV viewing	Less than 1	12	11
	1-2 hrs	18	11
	3-4 hrs	7	4
	5-6 hrs	3	1
	More than 6 hrs	0	2
7) Hours per week of commercial TV viewing	Less than 1	7	3
	1-2 hrs	8	6
	3-4 hrs	11	3
	5-6 hrs	7	4
	More than 6 hrs	7	13
8) Owns VCR	Yes	26	21
	No	14	8

Table 12 - continued

9) Hours per week of movie viewing on VCR (if applicable)	Less than 1	3	3
	1-2 hrs	10	6
	3-4 hrs	9	9
	5-6 hrs	1	1
	More than 6 hrs	1	3
10) Most important reason for TV viewing	Entertainment	18	20
	Relaxation	5	7
	Learning	3	2
	Relief from boredom	4	0
	Keeping up socially	2	0
	Other	2	0
11) Considers TV a good instructional medium	Yes	29	29
	No	5	0

CHAPTER FIVE

Discussion

While the central hypothesis of this study was supported under certain conditions, the results presented a number of surprises and methodological complications. The main hypothesis, which stated that use of the PEAC system would influence cognitive recall, was supported for critical content when a very general real-time evaluation question was used. The more technical, or specific question of Study 1 was not influential because it was simply ignored. Regarding the incidental material measure, contrary to the initial hypothesis based on time honored mathemagenic effects so frequently found in prose research, no effects were found under any circumstances. However, this outcome would be expected in light of the learners' virtual non-use of the PEAC hand-units. The major methodological setback of the study was the failure to obtain discriminating levels of prior knowledge on the pretest. Thus, the predicted interaction between the Familiarity and Usage factors could not be tested. Finally, the transferability of results from prose to video-based media was dealt an additional blow when the individual aptitudes of spatial visualization and verbal fluency did not prove to be effective predictors with any of the dependent measures. The implications arising out of the above results are discussed below.

PEAC Usage

When considering the effect of PEAC usage, it is essential that one first look at the frequency of real-time responses made with the hand-units. In previous evaluations with the PEAC methodology (Baggaley, 1982a, 1982b, and 1982c; Baggaley & Smith, 1982; CPB, 1981), viewers have responded relatively frequently while watching a program. This is in sharp contrast to both the average of 1.24 responses (after the initial setting) for the entire 27-minute program in Study 1 and the observed low but unrecorded frequency of responses for Study 2. Two questions must therefore be addressed: why did this low level of response frequency occur for both studies; and why did the apparent rate of average responses for Study 2 increase over those for Study 1? Several factors may have contributed to these events.

First, in any evaluation, there must be at least one question on which to base the evaluative judgment. The opportunity for a viewer to overtly evaluate is an event separate from that of simply viewing. An individual viewer chooses to participate or not participate (CPB, 1980, p. C-16). The directing influence of the evaluation question(s) may be critical not only to the nature of the responses, but also to their frequency. As noted in Chapter Two, most researchers employ a global question for a variety of reasons, including generalizability of data and technical aspects of the PEAC system.

For Study 1, in an attempt to more finely explore the link between the task demand of evaluation and the particular effects of PEAC usage, a somewhat more specific evaluation question was employed, namely, "How effective do you think this program is in demonstrating language acquisition in children?" Although it was initially intended to utilize this question for both samples, results from Study 1 showed that the interaction between this question and the characteristics of the program not only failed to provide for discriminating evaluation, but did not elicit the frequency of responses essential to the evaluation process of the PEAC system. Nearly all the subjects using the PEAC system rated the film as "very effective" or "fairly effective" at all times and in fact used their hand-held PEAC devices very infrequently. The specific and rather lengthy phrasing of the evaluation question may have been ignored or forgotten by viewers. Since the question appeared, in effect, too complicated for viewers to keep in mind while simultaneously watching the program, they seemed to choose simply to watch the program. In other words, they chose not to participate in the evaluation. If, on the other hand, they were actively evaluating, and simply never changed their opinion, the same conclusion must still be drawn, that the question was inappropriate. The procedure provided no useful information. Since in Study 1 the PEAC users hardly ever used their hand units during the program, the main

difference between the treatment and control groups, that is, the use or non-use of PEAC, in practical terms virtually disappeared. This would certainly account for the lack of differences between the groups on the dependent measures.

The question employed for Study 2 was global and in keeping with previous research: "How good do you think this program is, based on your overall reaction?" It was hoped that the nature of this question would encourage students to be more discriminating about different aspects of the program and that the simpler, more understandable phrasing would make the task easier and thus encourage more frequent responses. In light of the fact that the only manipulated design difference between Study 1 and Study 2 was the question, the observed increase in the frequency of responses for Study 2 suggests that the question itself plays a crucial role in whether people respond. Response rate, in turn, appears to influence the obtrusiveness of the PEAC methodology, as seen by the effect produced on cognitive recall. The act of holding the hand units and being asked to evaluate seems not to be what differentiates the PEAC methodology from traditional techniques nor what defines the task demand of evaluative viewing. Instead, the task demand seems to be defined by the overt request of directed attention in the form of the overall evaluation question. The resulting activity then determines how various types of learning are influenced.

From the standpoint of this series of experiments, it was the change in the real-time question which produced both an increase in response rate and the overall effect for PEAC as initially hypothesized. However, it is obvious that a second major factor, held constant across both studies, was the likely cause of general low response rate, that being the nature of the program itself. As mentioned in Chapter Three, one of the reasons for the selection of the particular program used was that it was considered a good representation of a typical ETV program. Viewing of ETV is not, in general, a highly emotive activity. A typical program would probably not elicit the individual attitude peaks and valleys of a program dealing with a controversial subject matter or a program designed largely to entertain. In other words, ETV does not usually bring out strong opinions. The PEAC system, however, is useful in formative evaluation precisely because it records measurable moment-by-moment changes in reactions to specific program segments which are then compared with one another as well as against a baseline. It is this type of coordinate data which allows the producer and/or designer to make revisions and improve the program. If the program does not stimulate this type of continuous discriminating reaction, the primary strength of the PEAC system appears to be lost, unless of course one wished simply to ensure the presence of an attitudinal plateau. Without access to frequent responses in real-time data, the researcher is unable to tie cognitive

learning to viewers' reactions. Programs and/or questions which do produce frequent responses may in fact be useful. For cases in which frequent responses are not elicited, post-presentation evaluation techniques seem the best alternative. While the use of the PEAC system does not preclude the use of post-presentation techniques (in fact, to the author's knowledge, the system is always used in conjunction with post-viewing questionnaires), the cost and effort seems worthwhile only if useful additional information is obtained. Furthermore, the level of obtrusiveness, especially at increasing levels of responding, may obscure the evaluation of the principle aim of the program, that is, learning. If the PEAC system is used because the ETV program is felt to have sufficient attitudinal variability, the measure of cognitive effectiveness should probably be conducted separately.

The strength of the above argument relies in part on the assumption that the learners of Study 1 and Study 2 performed in a relatively equivalent manner regarding the treatment. It was initially expected that learners from Study 1 would outperform those of Study 2, but it was not expected that the difference would interact with the presence or absence of the PEAC system. To examine potential differences, tests were performed between sample performance only to detect any differential effects. As was noted, there was no PEAC effect on multiple choice items. Thus, the interaction between samples and multiple

choice scores which did emerge obviously had no direct effect on the independent variable. The short answer test which yielded the PEAC effect did not produce an interaction. The PEAC effect therefore remained uncontaminated by any differences between the groups.

The last potentially significant fly in the ointment is the apparent floor effect on the short answer test. This result would be a genuine dilemma if it were the outcome of a single source, that is, non-learning. However, three separate features of the cognitive measures suggest a different picture. First, purely from a statistical standpoint, a floor effect tends to reduce the likelihood of finding differences. Thus, it is indeed surprising that any effect could have emerged at all, and presumably did so because it was "real". Second, while short answer responses remained somewhat depressed following the presentation, multiple choice performance improved between 200 and 300 percent, thus indicating that a reasonable level of learning of one sort (recognition) occurred. The viewers were in fact paying some kind of attention, and did learn something. Third, the lack of an effect in Study 1 would be expected in light of the non-participation of learners. However, in Study 2, short answer performance for the PEAC group actually decreased marginally even though their multiple choice answers improved 300 percent, while the non-PEAC group improved to a significant degree on both measures. Thus, all of the

learner performance on all measures of both studies was reasonable, and would be expected under the circumstances. It would therefore be unreasonable to assume that the short answer test results were anomolous. The large n-size further supports the validity and reliability of the data. That the college students learned little from the program is an issue addressed below, but the substantive significance of the PEAC effect is seen to be on solid, albeit undramatic ground.

Selective Effects

The effect of the PEAC system methodology raises several important issues about the nature of measuring cognitive recall in a formative evaluation context. The issue of cognitive recall deals with the degree to which information presented in the program is assimilated by the student in terms of factual, comprehensive or applied knowledge (Anderson, 1972; Romiszowski, 1981). In both studies, three measures were used to assess cognitive recall: short answer, multiple choice and "fill-in-the-blank" types of tests. In distinguishing the three formats, none is universally appropriate as a measure of recall. Rather, the appropriateness of their use depends upon the educational objectives of the ETV program. Previous research, however (Anderson, 1972), has indicated that short answer tests, which require the student's own words, are more discriminating of knowledge which includes comprehension and application. This is because these types of tests are

more likely to require cognitive processing than multiple choice or "fill-in-the-blank" type responses, which are more likely to be verbatim phrases or paraphrases. Assuming then that short answer responses entail the reformulation of assimilated knowledge (into written answers), this measure of recall provides an indication of the effect of PEAC usage on the processing of information. The contention that short answer responses are in fact more discriminating of comprehension and application was supported when results revealed the short answer posttest scores to be the sole contributing measure to the overall main effect of PEAC. This result suggests that use of the PEAC system affects an individual's "deeper" processing of information as opposed to simple information acquisition. While the issue of constructive versus reconstructive memory extends beyond the scope of this thesis (Zangwill, 1972), it is worth noting that the perennial concern over the passive nature of learning from television (which would encourage surface learning and discourage deep learning) may have manifested itself in the recognition versus recall differences.

Extrapolating further from Anderson (1972) and Rothkopf (1970), the task of continuous real-time evaluation may further encourage surface processing and recall of verbatim material, at the expense of material involving comprehension. Given the low response rate, however, any increase in incidental recall could not be

expected. The use of the PEAC system does not seem to influence recognition memory, at least at the usage level found here, as was demonstrated by the multiple choice scores. Thus, the kind of incidental material measured in the study (all basically surface, verbatim-type questions) did not require conscious cognitive processing in the same manner as short answer material involving comprehension. Reflecting on the results obtained for the multiple choice items, recall of this incidental material would not have been expected to be reduced by the PEAC system in either study.

Attitude measures. There were no differences between PEAC users and non-users for any of the attitude items. The only differences which did emerge were from the pre- to the posttest. In Study 1, all three items which dealt with television as a medium for information dissemination (#7, 12 and 13) were significant, and showed the greatest substantive change. It is interesting to note that two (#7 and 12) of these three items also produced effects in the second study. Following the presentation, viewers were more positive about the use of TV as an instructional tool (#12 and 13). However, they were significantly less positive about using a documentary to accomplish the instructional task (#7). These results suggest that while the experience inclined students toward a positive view of using television as a learning medium, they seemed to dislike the documentary format. These conclusions are

nevertheless strange because students did like this documentary (see Program Evaluation Ratings, section II, items #2 and 4). Therefore, one can only conclude that they don't want educational documentaries "shown regularly on television", as the question specified. However, viewers appear to both approve of and recommend the documentary in the context in which they found themselves, that is, watching ETV in the classroom. Unfortunately, this favorable disposition failed to result in a favorable level of learning, as will be discussed below.

Changes were also revealed for some items which dealt with opinions about the value of language acquisition (#4, 5a, 8 and 9 in Study 1, and #1 and 8 in Study 2). These differences, though statistically significant, were small, with responses remaining positive on both the pretest and posttest for Questions #1, 4, 5a and 9. These results suggest that students respect the importance of language acquisition as a topic and thus were favorably inclined to watch a program on that topic.

Curiously, the only "attitude" item which directly referred to information content in the program (#10), elicited changes in both studies. This item was, upon review, determined to be as much cognitive as affective, and in fact produced some change in both learning and attitude as a result of specific information dissemination.

Program evaluation rating measures. The program evaluation ratings indicated that, overall, students liked

the program. Those items which required short answers revealed that most viewers liked the program very much, especially the segments involving examples of children. It is interesting that quite a few viewers could find nothing to exemplify "the greatest weakness" nor even could they suggest an improvement. However, dissatisfaction was expressed about the narrative technique of "talking heads", which is a style often used in ETV and documentary programs.

No effect of PEAC usage was demonstrated for any of the program evaluation rating items in Study 1, and only one item (#3) showed differences between groups in Study 2. This one item indicated that users were more dissatisfied with the narration technique of "talking heads" than were non-users.

Overall learning. One final issue worth noting is the level of overall learning. While surface performance on the posttest was fair, as shown by the multiple choice scores, comprehension was low, as evidenced by the short answer scores. This seems to suggest two guidelines for the improvement of ETV. First, collaboration between the instructional designer and the ETV producer needs to be encouraged. Specifically, the two professionals (or teams) must work together to develop solid, effective material which really meets the pre-determined objectives. As has been mentioned, the program used in this study was selected because it was deemed representative of ETV. The actual

educational objectives of its original producers were not known, so for the purposes of this study, the program objectives were defined by the experimenter (see Chapter Three). Thus, the described objectives may not be those of the original producers at all. This liberty of defining the program's educational objectives was taken to illustrate a point. Namely, most ETV is produced either by instructional designers or ETV producers, seldom by both together. The situation of an educator imposing objectives on a program is not a rarity. Teachers, for example, often select a film or videotape for classroom viewing from a catalogue. They then show the program in the context of their course, presumably on the assumption that students will obtain some measure of cognitive as well as other types of learning from it. ETV producers, on the other hand, may have created a program with the primary objective of enjoyment or entertainment. Learning and enjoyment are not mutually exclusive, but to ensure an overlap, producers and educational designers must work together.

The second guideline suggested by the learning performance is the need for the specification of objectives for ETV. While much has been written about objectives for traditional educational media, such as print, little is available as a guideline for the development and implemention of ETV objectives. If producers do not have precisely defined objectives, they can hardly be accused of failing to meet them.

Implications

The PEAC system methodology may be useful for measuring affective variables, but only for programs whose objectives include the attempt to change attitude. The results obtained from evaluating attitude for an educational program do not seem to be transferable to an assessment of a program's cognitive learning effectiveness. It may not be useful to assess attitude toward programs, such as the one used, which are geared toward cognitive change (see discussion of program objectives above). The PEAC system methodology has been advocated as not artificially changing the act of viewing (Radio-Quebec, 1984). This study has indicated this to be true for attitude, the system's area of strength, and false for cognitive aspects, an area for which the system was probably not designed.

Two key ideas emerge from these results. First, it would be inappropriate to condemn the PEAC system for not functioning well under circumstances for which it was not designed. In fact, this study provides tentative support for its use in the context of its proper environment. Second, instructional evaluators must take great care to use the system for its intended purpose, and not generalize to variables beyond its scope. While this is obvious, the main issue extends just beyond. The greatest danger seems not to lie in the misuse of the PEAC system in a cognitive domain, because it quite simply provides no information. Rather, the danger lies in evaluators' reliance solely on

the PEAC system for the formative evaluation of ETV productions. The sophistication and "hard data" produced by the PEAC system may lull evaluators into feeling that their job is not only done, but done thoroughly and well. If the objectives of a program are partially cognitive, however, then those objectives must be dealt with independently of the PEAC-involved evaluation. Cognitive evaluation techniques are admittedly more difficult and in some respects less precise because they cannot pinpoint performance to specific time frames of the program. Thus, they likely fall very low in a producer's priorities, but this is exactly the point. The PEAC system may convince producers and instructional designers that they have identified all of the program's weak spots, and consequently, they are fully prepared to make any necessary revisions. However, this happy task also serves to distract them from what should be (from an educator's point of view), other essential objectives. The affective information provided by the PEAC system is inherently concrete, tangible, and, from an "artistic standpoint", remediable. Cognitive aspects are far more elusive and difficult to address, yet if the medium of television is to serve a multiple role in formal and informal educational environments, it must attend to the facts as well as the opinions.

In conclusion, both instructional designers and ETV producers need to examine their goals and the relation of

these goals to the media used. What does it mean to watch educational television and how is this viewing affected by evaluation techniques? Any evaluation research must take place within the context of a multitude of factors, including objectives, individual characteristics, motivation, environmental conditions and so on. Suggestions for future research include closer looks at various types of real-time evaluation questions and ETV programs. Concentration is needed both on the effect of the PEAC methodology on attitudes in strongly emotive programs, and on its utility in assessing cognitive change when frequent real-time responding does occur. Since different audiences may respond in different manners, formative evaluation techniques need to be explored with a cross-section of target audiences. Individual aptitudes almost certainly influence evaluation just as they do learning; how and under what circumstances they do so needs to be explored.

To improve the instructional diversity and utility of ETV, what is clearly needed, is: first, an understanding by all involved parties of the implications of what they are doing; second, a willingness to agree upon common objectives and so what is necessary to accomplish them; and finally, a framework for application (preferably based on theory) which at once addresses the constraints of the interaction while guiding the decision-making toward the best possible quality. It is toward the development of a more comprehensive and valid system of formative evaluation that this study has hopefully contributed.

References

Anderson, R. C. (1972). How to construct achievement tests to assess comprehension. Review of Educational Research, 42, 145-170.

Anderson, R. C., & Kulhavy, R. W. (1972). Learning concepts from definitions. American Educational Research Journal, 9, 385-390.

Baggaley, J. P. (1980). Psychology of the TV image. Farnborough: Gower.

Baggaley, J. P. (1982a). Audience analysis: 'You don't smoke, eh?'. (Report to the Canadian Cancer Society). St. John's, Newfoundland: Memorial University.

Baggaley, J. P. (1982b). Cancer prevention and literacy. (Report to the Canadian Cancer Society). St. John's, Newfoundland: Memorial University.

Baggaley, J. P. (1982c). Electronic analysis of communication. Media in Education and Development, 15, 70-73.

Baggaley, J. P., & Smith, K. (1982). Formative evaluation in rural education. Media in Education and Development, 15, 173-176.

Baggaley, J. P. (1985a). Design of a TV character with visual appeal for preschool children. Journal of Educational TV, 11(1), in press.

Baggaley, J. P. (1985b). Formative evaluation of educational television. Edited version of addresses , November, 1984, to the Human Sciences Research Council, Pretoria.

Baker, E. L. & Alkin, M. C. (1973). Formative evaluation of instructional development. AV Communication Review, 21, 389-418.

Bracht, G. H. (1970). Experimental factors related to aptitude-treatment interactions. Review of Educational Research, 40, 627-645.

Bracht, G. H., & Glass, G. V. (1968). External validity for experiments. American Educational Research Journal, 5, 437-474.

Cambre, M. A. (1981). Historical overview of formative evaluation of instructional media products. Educational Communications & Technology Journal, 29(1), 3-25.

Cambre, M. A. (1982). Formative evaluation and the new technologies. In J. P. Baggaley (Ed.) Experimental research in TV instruction, Vol. 5, pp. . Montreal: Concordia University.

Canadian Broadcasting Corporation (Producer). (1975). Out of the mouths of babes [videotape]. Montreal: Author.

Caron, A., Giroux, L., & Mayrand, C. (1982). Evaluation preliminaire du systeme PEAC: Proposition d'un devis experimental. Unpublished manuscript, Universite de Montreal, Departement de communication, Montreal.

Chu, G. C., & Schramm, W. (1979). Learning from television: What the research says (4th ed.). Washington, D.C.: National Association of Educational Broadcasters.

Coldevin, G. O. (1980). Formative research in television presentation strategies: Guidelines for effective production. In J. P. Baggaley (Ed.). Experimental research in TV instruction, Vol. 3, (pp. 62-80). St. John's, Newfoundland: Memorial University of Newfoundland.

Corporation for Public Broadcasting. (1981). A comparison of three research methodologies for pilot testing new television programs. Washington, D.C.: Author.

Corporation for Public Broadcasting. (1980). Proceedings of the 1980 technical conference: Qualitative television ratings. Washington, D.C.: Author.

Cronbach, L. J., & Snow, R. E. (1977). Aptitudes and instructional methods. New York: Irvington.

Cronbach, L. J., & Webb, N. (1975). Between-class and within-class effects in a reported aptitude x treatment interaction: Reanalysis of a study by G. L. Anderson. Journal of Educational Psychology, 67, 717-724.

Dailey, J. T. (1965). Spatial Visualization Test of the Dailey Vocational Tests. Boston: Houghton Mifflin Company.

Dick, W. (1980). Formative evaluation in instructional development. Journal of Instructional Development, 3(3), 3-6.

Festinger, L. (1957). A theory of cognitive dissonance. Evanston, IL: Row, Peterson.

Festinger, L. (1962). Cognitive dissonance. Scientific American, 207(4), 93-107.

Festinger, L., & Maccoby, N. (1964). On resistance to persuasive communication. Journal of Abnormal & Social Psychology, 68, 359-366.

French, J. W. (1951). The description of aptitude and achievement tests in terms of rotated factors. Psychometric Monographs (No. 5). Chicago: University of Chicago Press.

French, J. W., Ekstrom, R. B., & Price, L. A. (1963). Manual for Kit of Reference Tests for Cognitive Factors. Princeton, NJ: Educational Testing Services.

Geis, G. L., Burt, C. W., & Weston, C. B. (1984). Instructional development: Developmental testing. Paper presented at the annual convention of the American Educational Research Association, New Orleans.

Goldman, R. J., & LaRose, R. (1981). Assessment of audience feedback systems for research and programming. Washington, D.C.: Corporation for Public Broadcasting.

Gooler, D. D. (1980). Formative evaluation strategies for major instructional development projects. Journal of Instructional Development, 3(3), 7-11.

Green, B. (1982). Market analysis and strategic planning at TVOntario. In J. P. Baggaley (Ed.), Experimental research in TV instruction, Vol. 5, (pp.). Montreal, ConcordiaUniversity.

Gropper, G.L. (1967). Does "programmed" instruction need active responding? AV Communications Review, 15, 5-22.

Gropper, G. L., Lumsdaine, A. A., & Shipman, V. (1961). Improvements of televised instruction based on student responses to achievement tests (Studies in televised instruction, Report No. 1). Pittsburgh: Metropolitan Pittsburgh Educational Television Stations WQED-WQEX & American Institute for Research.

Guilford, J. P., & Zimmerman, W. S. (1956). Guilford-Zimmerman Aptitude Survey Part VI Spatial Visualization. Beverly Hills, CA: Sheridan Supply Company.

Henerson, M. E., Morris, L. L., & Fitz-Gibbon, C. T. (1978). How to measure attitudes. Beverly Hills, CA: Sage.

Ho, A. Y-K. C. (1981). Conjugate reinforcement as a method of studying television viewing behaviour. Unpublished master's thesis, Concordia University, Montreal.

Hoban, C. F. (1956). Research and reality. AV Communication Review, 4, 3-20.

Jurgemeyer, F. H. (1982). Programmed instruction: Lessons it can teach us. Educational Technology, 22(May), 20-22.

Klauer, K. J. (1984). Intentional and incidental learning with instructional texts: A meta-analysis for 1970-1980. American Educational Research Journal, 21, 323-339.

Lashley, K. S., & Watson, J. B. (1921). A psychological study of motion pictures in relation to venereal disease campaigns. Social Hygiene, 7, 181-219.

Lindquist, E. F. (1951). Preliminary considerations in objective test construction. In E. F. Lindquist (Ed.), Educational measurement, (pp. 119-158). Washington, D.C.: American Council on Education.

Lumsdaine, A. A. (1947). Experimental research and the improvement of teaching film. Educational Screen, 26, 254-255.

Madaus, G. F., Stufflebeam, D. L., & Scriven, M. S. (1983). Program evaluation: A historical overview. In G. F. Madaus, M. S. Scriven, & D. L. Stufflembeam (Eds.). Evaluation models: Viewpoints on educational and human services evaluation. Hingham, MA: Kluwer Academic Publishers.

Markle, S. M. (1967). Empirical testing of programs. In P. C. Lange (Ed.), Programed instruction (Sixty-Sixth Yearbook of the National Society for the Study of Education, Part II). Chicago: University of Chicago Press.

Mitchell, P. D. (1979). A micro-analytic procedure for improving instructional materials. In T. Page & A. Whitlock (Eds.), Aspects of educational technology. London: Kogan Page.

Nickerson, R. B. (1979). The formative evaluation of instructional television programming using the program evaluation analysis computer (PEAC). In J. P. Baggaley (Ed.), Experimental research in TV instruction, Vol. 2, (pp. 121-125). St. John's, Newfoundland: Memorial University.

Orr, D. B. (1966). The evaluation of televised instruction. AV Communication Review, 14, 363-370.

Paluzzi, J. V. (1979). Interest and presentational formats in TV instruction. In J. P. Baggaley (Ed.), Experimental research in TV instruction, Vol. 3, (pp. 23-25). St. John's, Newfoundland: Memorial University.

Radio-Quebec. (1984). Variabilite de la reponse PEAC. Montreal: Author, Service de la recherche.

Radio-Quebec. (1985). Recherche formative pour la serie "Arrimage". Montreal: Author, Service de la recherche.

Romiszowski, A. J. (1981). Designing instructional systems. London: Kogan Page.

Rothkopf, E. Z. (1966). Learning from written materials: An exploration of the control of inspection behavior by test-like events. American Educational Research Journal, 3, 241-249.

Rothkopf, E. Z. (1970). The concept of mathemagenic activities. Review of Educational Research, 40, 325-336.

Scriven, M. S. (1967). The methodology of evaluation. In R. W. Tyler, et al. Perspectives of communication evaluation (AERA monograph series on curriculum evaluation, no. 1), (pp. 39-83). Chicago: Rand McNally.

Skinner, B. F. (1957). Science and human behavior. New York: Macmillan.

Snow, R. E. (1978). Theory and method for research on aptitude processes. Intelligence, 2, 225-278.

Twyford, L. C. (1954). Profile techniques for program analysis. AV Communication Review, 2, 243-261.

Tyler, R. W. (1942). General statement on evaluation. Journal of Educational Research, 35, 492-501.

Watts, G. H., & Anderson, R. C. (1971). Effects of three types of inserted questions on learning from prose. Journal of Educational Psychology, 62, 387-394.

Weston, C. B. (1986). Formative evaluation of instructional materials: An Overview of approaches. Canadian Journal of Educational Communication. 15, 5-17.

Winer, L. R. (1981). An application of Brunerian theory to instructional simulation: Spatial visualization factorial research designs & wooden blocks. Unpublished master's thesis, Concordia University, Montreal.

Zangwill, O. L. (1972). Remembering revisited. Quarterly Journal of Experimental Psychology, 24, 123-138.

Zimbardo, P. G., Cohen, A. R., Weisenberg, M., Dworkin, L., & Firestone, I. (1966). Control of pain motivation by cognitive dissonance. Science, 151, 217-219.

APPENDICES

Appendix A

Prior Knowledge Pretest and Cognitive Recall Posttest

Questionnaire 1

Section I - Short Answer Questions

Answer each question in section I in your own words. Be as brief and concise as possible in the space provided.

1) Do parents generally correct their children's speech errors, and if so, how?

2) Children have a tendency to look for general rules of language. This is called "over-regularization". Give an example of over-regularization.

3) What would you consider the single most important thing a parent does to facilitate a child's language acquisition?

4) a) What part of a child's language acquisition is biologically determined (innate), if any?

4) b) What part of language acquisition is determined by a child's environment, if any?

5) Does language develop in qualitatively different stages, or as a quantitative accumulation of the same basic principles?

Section II - Multiple Choice Questions

Answer each question in section II by circling the letter of your response. Circle only one letter for each question.

1) The fact that babies perceive speech at such an early age implies that this ability is

   a) passed on from parents to children all over the world through practice
   b) biologically determined
   c) dependent on cultural norms and customs
   d) none of the above

2) A child's grammar is close to that of an adult by the time the child has reached the age of

   a) 2 years
   b) 3 years
   c) 4 years
   d) 5 years

3) The ability to combine words into two and three word sentences is called

   a) grammatical phrasing
   b) telegraphic speech
   c) horophrasing
   d) mastery

4) A three year old is given a large brown cardboard box, with one side colored yellow. When asked to point to the front of the box, he is most likely to point to

   a) the yellow side
   b) the side facing him
   c) the top of the box
   d) any side at all

5) Infants are highly responsive to the human voice when they are as young as _____

   a) 1 day
   b) 1 week
   c) 2 weeks
   d) 1 month

6) Grammatical features of English are acquired

a) in the same order, although at different rates for different children
b) at the same rate, although in different order for different children
c) in the same order and at the same rate by all children whose native tongue is English
d) at different rates and in different order for different children

7) In which of the following situations would a three year old be most likely to indicate the correct side of an object as its front?

a) when the object is a person
b) when the object has one side of a different color
c) when the object is familiar to them
d) when the object is moveable

8) When a child is given an incorrectly constructed sentence, she will

a) not know what the errors are and repeat the errors in her own speech
b) know what the errors are but not repeat them in her own speech
c) not know what the errors are and not repeat them in her own speech
d) all of the above, depending on the circumstances

9) Horophrase is the correct term for

a) the one word sentence
b) gutteral utterances of newborns
c) the two or three word sentence
d) none of the above

10) The overapplication of grammatical rules is best exemplified by a child saying

a) "open the door" instead of "door open"
b) "he went" instead of "he goed"
c) "he goed" instead of "he went"
d) "himself" instead of "herself"

Appendix B

Attitude Pretest and Attitude Posttest

Questionnaire 1 Section III

Answer each question in section III by placing a check ( ) over the term that best indicates your response.

1) It is important to understand how an individual learns a language.

Strongly agree	Agree	Don't know	Disagree	Strongly disagree

2) Knowing how children learn to talk makes one more socially attractive.

Strongly agree	Agree	Don't know	Disagree	Strongly disagree

3) Children are usually an interesting topic.

Strongly agree	Agree	Don't know	Disagree	Strongly disagree

4) Knowing about how children learn language will help a person to teach their own children to speak more quickly and effectively.

Strongly agree	Agree	Don't know	Disagree	Strongly disagree

5) At least one course in child development would be useful for everyone -

a: for those who have or who plan to have children

Strongly agree	Agree	Don't know	Disagree	Strongly disagree

b: for those who don't plan to have children

Strongly agree	Agree	Don't know	Disagree	Strongly disagree

6) The problem with television is that all the information goes by so quickly that you can't remember very much.

Strongly agree	Agree	Don't know	Disagree	Strongly disagree

7) It is unfair to expect people who are not parents to be interested in children.

Strongly agree	Agree	Don't know	Disagree	Strongly disagree

8) Documentary programs on various aspects of child development should be shown regularly on television.

Strongly agree	Agree	Don't know	Disagree	Strongly disagree

9) The importance of how an infant learns his or her mother tongue is exaggerated.

Strongly agree	Agree	Don't know	Disagree	Strongly disagree

10) Knowing whether speech is culturally or biologically determined is irrelevant.

Strongly agree	Agree	Don't know	Disagree	Strongly disagree

11) Parents should use different levels of language when speaking to children of different ages.

Strongly agree	Agree	Don't know	Disagree	Strongly disagree

12) Knowing about how children learn language will help a person to be a better parent.

Strongly agree	Agree	Don't know	Disagree	Strongly disagree

13) Television is a better way to learn about topics like language acquisition because it makes everything concrete and interesting.

Strongly agree	Agree	Don't know	Disagree	Strongly disagree

Appendix C

Spatial Visualization Test

Questionnaire 2

This section is a test to see how well you can understand how the pattern of an object can be folded or rolled into a solid figure. In the sample on the next page (after the answer sheet) are two drawings. The drawing at the left of the page is a flat pattern. The drawing at the right is the solid object that is made by folding the pattern.

Notice the small arrows on both the flat pattern and the object ( ). These arrows appear in exactly the same position on each pattern and its object. For example, the arrows on the solid object in sample show exactly the same corner as the one marked with arrows on the flat pattern. Note, too, the dotted lines on the flat pattern. These dotted lines show where the folds would be in the object.

Each number on the solid object indicates an edge that corresponds to an edge marked with a letter on the flat pattern. Look at the solid object. Notice the numbers 1, 2, and 3. Each number is a separate question, so there are three questions to answer in this drawing. You are to decide which letter (A, B, C, D, E, F, G, or H) on the flat pattern corresponds to each of the numbers 1, 2, and 3 on the solid object.

In order to answer question 1, find the number 1 on the solid object. When the object is unfolded, that edge falls into the position marked A in the flat pattern. Therefore, the answer to question 1 is A. On your answer sheet, write the letter A on the space beside number 1.

Now find the number 2 on the object. What letter on the flat pattern is in the same position as the number 2 on the object? For question 2, write in the appropriate letter. Go ahead with question 3. Write your answer on the answer sheet.

The right answer for question 2 is E; for question 3 it is F. If you did not mark these answers, look at the drawings again to see what the right answers should be. Cross out any wrong answer you have and mark the right no.

Mark only ONE answer for each question. If you change your mind about an answer, cross out your first answer and write the other one in the space. PLEASE DO NOT MARK ON THE PAGES WITH THE DRAWINGS. WRITE ONLY ON YOUR ANSWER SHEET. If you finish before time is up, you should go back and check your work.

ANSWER SHEET

Page 1

1 ____
2 ____
3 ____

Page 2

4 ____
5 ____
6 ____

Page 3

7 ____
8 ____
9 ____
10 ____
11 ____
12 ____
13 ____

Page 4

14 ____
15 ____
16 ____
17 ____

Page 5 (continued)

18 ____
19 ____
20 ____
21 ____

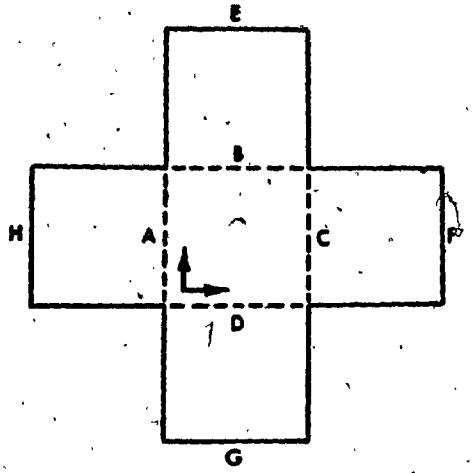
Page 6

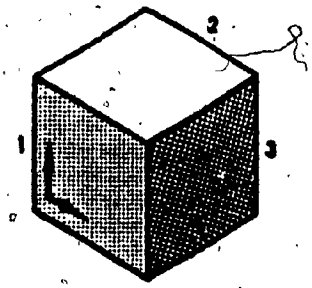
22 ____
23 ____
24 ____
25 ____
26 ____
27 ____

Page 7

28 ____
29 ____
30 ____
31 ____
32 ____
33 ____

SAMPLE

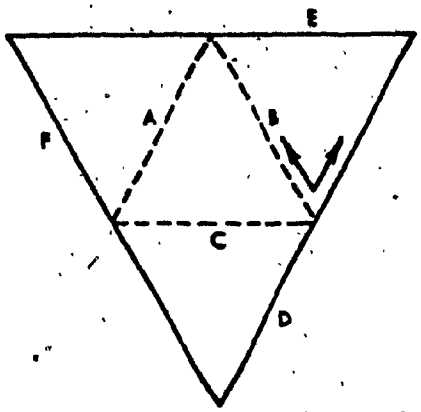


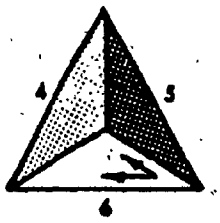


Do not turn the page until told to do so.

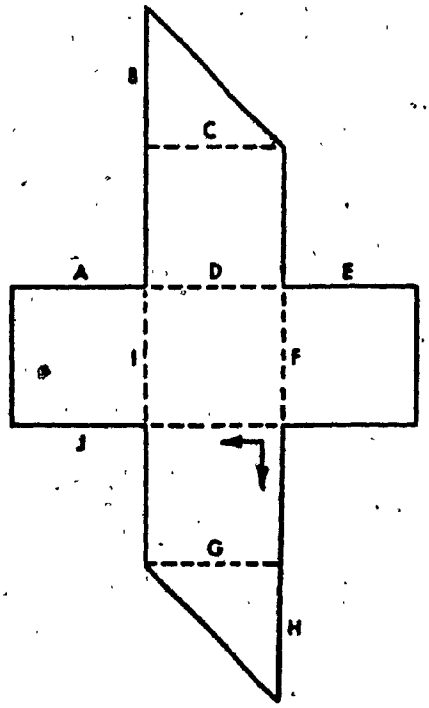
SPATIAL VISUALIZATION TEST

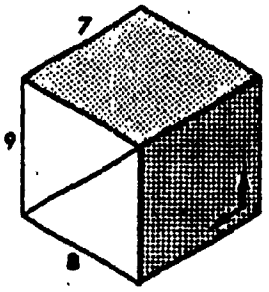
To the examinee: Begin with item 4 and continue through the booklet. Make no marks in this booklet.

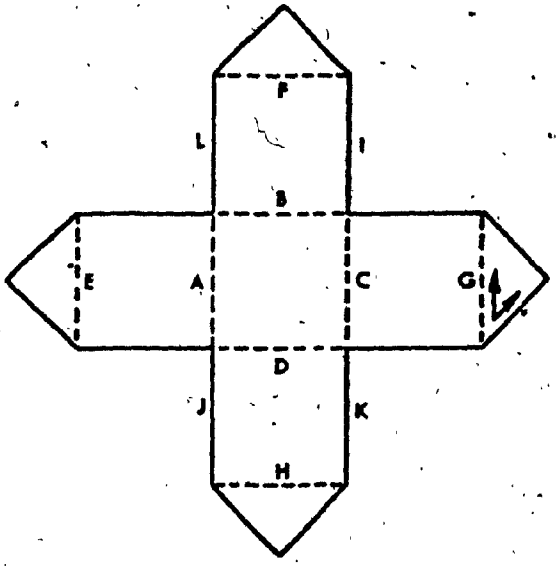


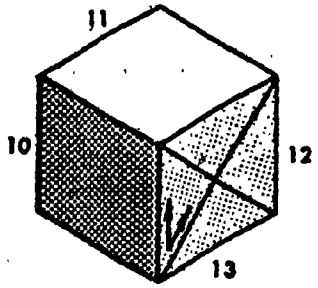


Go on to the next page.



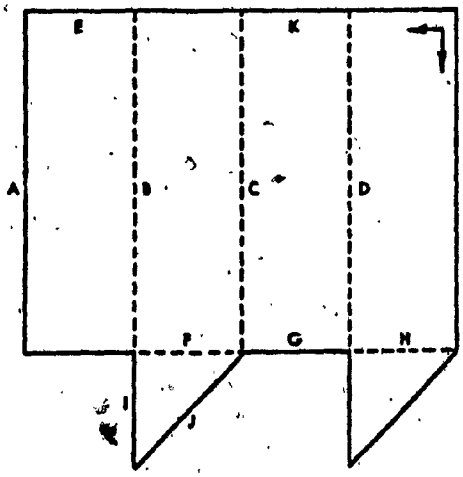


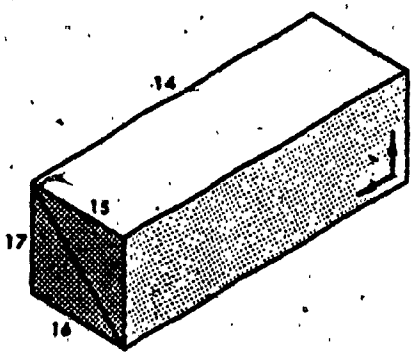


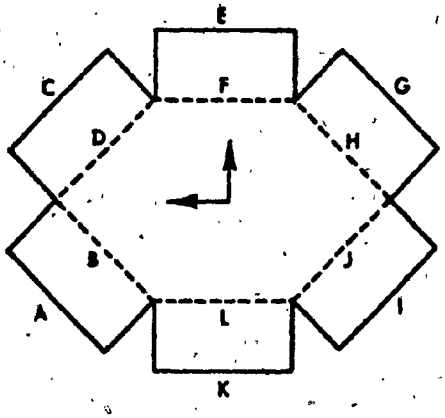


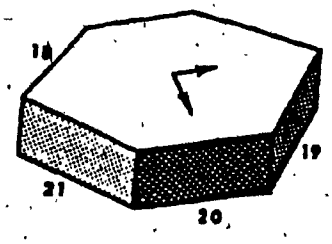
Go on to the next page.

— MAKE NO MARKS IN THIS BOOKLET. —

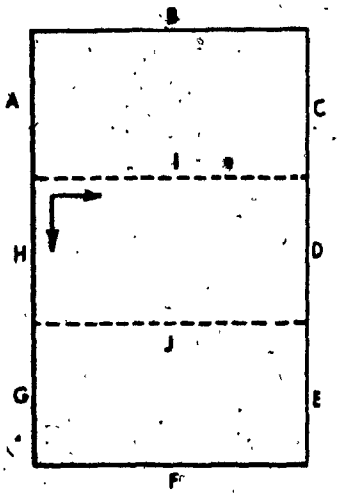


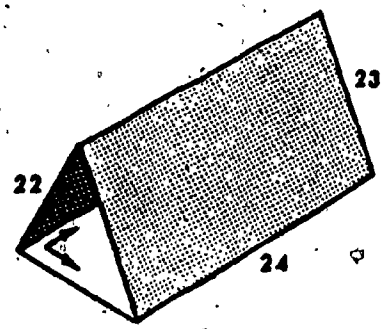


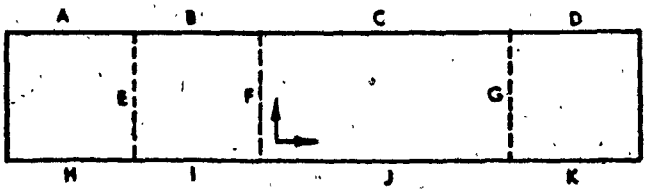


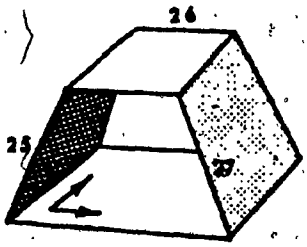


Go on to the next page.



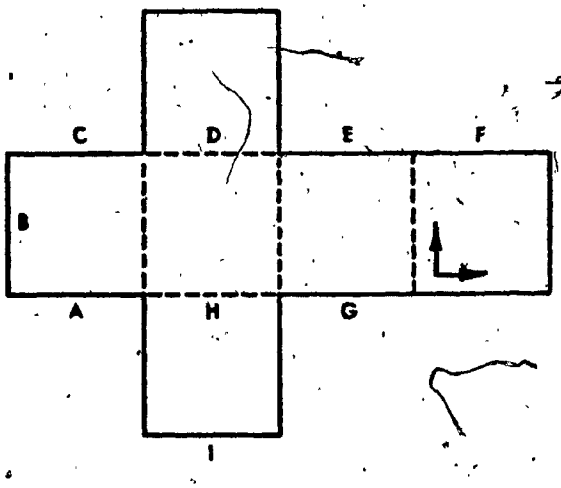


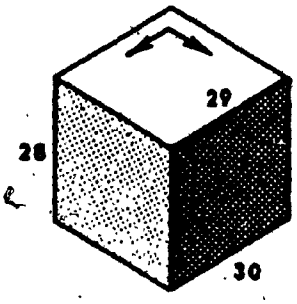


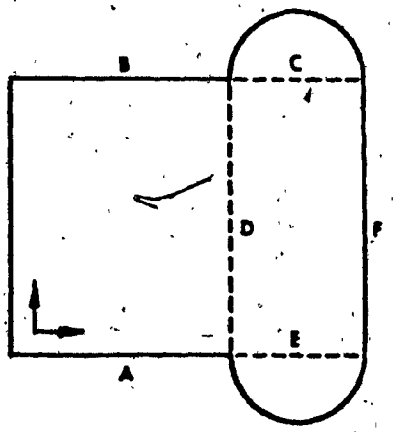


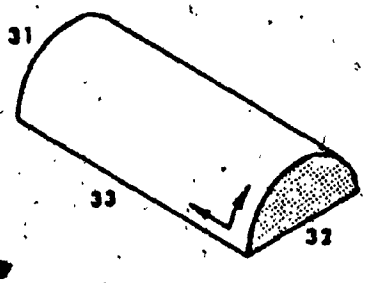
Go on to the next page.

— MAKE NO MARKS IN THIS BOOKLET. —









END OF TEST

Appendix D

Verbal Fluency Test

Questionnaire 3

This is a test of your knowledge of word meanings. Look at the sample below. One of the five lettered words has the same meaning or nearly the same meaning as the word above.

jovial

A. refreshing
B. scarce
C. thickset
D. wise
E. jolly

The answer to the sample item is E.

For each word, circle the letter beside the word that is closest in meaning.

DO NOT BEGIN UNTIL TOLD.

1. cottontail
   A-squirrel
   B-popular
   C-boa
   D-marshy plant
   E-rabbit

2. marketable
   A-partisan
   B-jocular
   C-marriageable
   D-salable
   E-essential

3. boggy
   A-afraid
   B-false
   C-marshy
   D-dense
   E-black

4. gruesomeness
   A-blackness
   B-falseness
   C-vindictiveness
   D-drunkenness
   E-ghastliness

5. loathing
   A-diffidence
   B-laziness
   C-abhorrence
   D-cleverness
   E-comfort

6. bantam
   A-fowl
   B-ridicule
   C-cripple
   D-vegetable
   E-ensign

7. evoke
   A-wake up
   B-surrender
   C-reconnoiter
   D-transcend
   E-call forth

8. unobtrusive
   A-unintelligent
   B-epileptic
   C-illogical
   D-lineal
   E-modest

9. terrain
   A-ice-cream
   B-final test
   C-tractor
   D-area of ground
   E-weight

10. capriciousness
   A-stubbornness
   B-courage
   C-whimsicality
   D-amazement
   E-greediness

11. maelstrom
   A-slander
   B-whirlpool
   C-enmity
   D-armor
   E-majolica

12. tentative
   A-critical
   B-conclusive
   C-authentic
   D-provisional
   E-apprehensive

13. placate
   A-rehabilitate
   B-plagiarize
   C-depredate
   D-apprise
   E-conciliate

14. surcease
   A-enlightenment
   B-cessation
   C-inattention
   D-censor
   E-substitution

15. apathetic
   A-wandering
   B-impassive
   C-hateful
   D-prophetic
   E-overflowing

16. paternoster
   A-paternalism
   B-patricide
   C-malediction
   D-benediction
   E-prayer

17. opalescence
   A-opulence
   B-senescence
   C-bankruptcy
   D-iridescence
   E-assiduity

18. lush
   A-stupid
   B-luxurious
   C-hazy
   D-putrid
   E-languishing

19. curtailment
   A-expenditure
   B-abandonment
   C-abridgment
   D-improvement
   E-forgery

20. perversity
   A-adversity
   B-perviousness
   C-travesty
   D-waywardness
   E-gentility

21. calumnious
   A-complimentary
   B-analogous
   C-slanderous
   D-tempestuous
   E-magnanimous

22. illiberality
   A-bigotry
   B-imbecility
   C-illegibility
   D-cautery
   E-immaturity

23. clabber
   A-rejoice
   B-gossip
   C-curdle
   D-crow
   E-hobble

24. sedulousness
   A-diligence
   B-credulousness
   C-seduction
   D-perilousness
   E-frankness

25. shortcake
A-condiment
B-pastry
C-fruit
D-sweetmeat
E-vegetable

26. hardtack
A-nail
B-textile
C-weapon
D-wood
E-biscuit

27. commendable
A-pleasurable
B-charitable
C-lucrative
D-proscriptive
E-laudable

28. nonchalant
A-sarcastic
B-discourteous
C-noble
D-unconcerned
E-unsophisticated

29. coloration
A-pigmentation
B-alternation
C-configuration
D-prevention
E-taint

30. aridity
A-bitterness
B-surface
C-sonority
D-dryness
E-torridity

31. demoniacal
A-aloof
B-mythical
C-thoughtful
D-fiendish
E-eccentric

32. highroad
A-mountain road
B-right of-way
C-main road
D-roadbed
E-concrete road

33. befog
A-dampen
B-forget
C-whip
D-mystify
E-belittle

34. platoon
A-tableland
B-bridge of boats
C-body of soldiers
D-commonplace remark
E-frigate

35. dullard
A-peon
B-duck
C-braggart
D-thief
E-dunce

36. momentously
A-frivolously
B-moderately
C-weightily
D-momentarily
E-modishly

37. corroboratory
A-plausible
B-anticipatory
C-confirmatory
D-explanatory
E-esoteric

38. figurine
A-metaphor
B-mine
C-poem
D-organ
E-statuette

39. rancorous
A-malignant
B-jubilant
C-abashed
D-inglorious
E-careless

40. inveteracy
A-habitualness
B-migration
C-bravery
D-covering
E-hatefulness

41. choler
A-anger
B-chorister
C-guard
D-saliva
E-refrigerator

42. vacillation
A-purification
B-wavering
C-expulsion
D-tempting
E-foolishness

43. aggrandizement
A-theft
B-impeachment
C-derision
D-amazement
E-enlargement

44. effulgence
A-prominence
B-outline
C-change
D-radiance
E-energy

45. aphasia
A-loss of speech
B-drunkenness
C-anemia
D-loss of memory
E-rash

46. panoplied
A-philosophical
B-dressed in armor
C-panting
D-frenzied
E-atavistic

47. sacrosanct
A-sacrificial
B-dormant
C-inviolable
D-superficial
E-gullible

48. prurience
A-modesty
B-sapience
C-provender
D-lust
E-security

Appendix E

Incidental Recall Test

Questionnaire 1

Section IV*

Answer each question in section IV in the space provided.

1) The early sequence of the program showed examples of children's speech at different ages. List the ages used.

2) What example was used to illustrate the relational terms "big" and "little"?

3) In the scene showing a mother and her three children in the kitchen, what kind of food or dish was being prepared?

4) Did the opening scene, that is, the first image in the program, take place indoors or outside?

5) What was the name of the researcher who invented the "wod" test?

6) a) Was there any music during some or all of the program?

Yes No (Circle one)

b) If yes, how would you rate it?

Distracting Neutral Enhancing

7) How many examples were given to illustrate the "wod" test?

* Section IV followed Section III of the Attitude posttest (shown in Appendix B).

8) How old was Katie in the program?

9) Was all of the narration (that is, the things said by the researchers who spoke during the program) in a Canadian accent? If not, can you name what country or countries their accents appear to be from?

10) Was the ending of the program summarized by a male or a female?

11) How were the terms "warm" and "cold" illustrated in the program?

Appendix F

Program Evaluation Ratings Questionnaire

Questionnaire 2

Section I

Answer each question in section I using one or two sentences.
Use the space provided.

1) What do you think is the program's greatest strength?

2) What do you think is the program's greatest weakness?

3) How could this program be improved to make it more effective for instructional purposes?

Section II

Answer each question in section II by circling the letter which best indicates your response.

1) How effective do you think this program was as an introduction to the topic of language acquisition in children?

a) very effective
b) fairly effective
c) not very effective
d) very ineffective

2) For yourself, would you judge the content of the program as being

a) very relevant
b) fairly relevant
c) not very relevant
d) very irrelevant

3) Do you feel that too much or too little information was presented?

a) way too much
b) too much
c) just right
d) too little
e) way too little

4) Would you recommend to others that they watch this program?

a) strongly recommend
b) recommend
c) neutral
d) discourage
e) strongly discourage

Section III

Consider the following aspects of the program and indicate how you would rate their quality by placing the appropriate letter in the space provided.

1) a) excellent
b) good
c) fair
d) poor

____narration

____content

____continuity

____pacing

____children used as examples

2) The photography in the program was aesthetically

Excellent Good Average Fair Poor

3) "Talking heads" (narrators talking directly to the camera) was a technique used in the program. Do you think it was used

Far too often	Too often	Just the right amount	Too rarely	Far too rarely

4) The content was logically presented and well-organized.

Strongly agree	Agree	Neutral	Disagree	Strongly disagree

5) In general, the content was easy to remember.

Strongly agree	Agree	Neutral	Disagree	Strongly disagree

6) On a continuum from dry to fascinating, how would you rate this program?

Very dry	Dry	Average	Slightly fascin-ating	Fascin-ating

7) If this were a weekly series, and you "accidentally" watched this program, would you purposely tune in next week?

Definitely tune in	Maybe tune in	Neutral	Probably not tune in	Defini-ately not tune in

Section IV

Answer these questions by placing a check ( ) over the term that best indicates your response.

1) a: I learn more from an educational television program if I have specific questions to answer beforehand because I am forced to concentrate.

Strongly agree	Agree	Don't know	Disagree	Strongly disagree

b: I learn less from an educational television program if I have specific questions to answer beforehand because I only concentrate on the questions.

Strongly agree	Agree	Don't know	Disagree	Strongly disagree

Appendix G

Demographic Questionnaire

Questionnaire 3

Place your response to each question as indicated.

1) Sex: _____ Male _____ Female

2) Age _____ Under 18 _____ 18-20 _____ 21-23 _____ 24-25 _____ Over 25

3) Do you have any children of your own?

_____ Yes _____ No

4) What is your major area of study?

Humanities _____
Sciences _____
Business _____
Fine Arts _____
Computer Science _____
Other _____

5) Have you ever taken a general psychology or a child psychology course?

_____ Yes _____ No

6) How many hours of educational (or public) television to you watch a week?

_____ Under 1 _____ 1 - 2 _____ 3 - 4 _____ 5 - 6 _____ Over 6

7) How many hours of commercial television do you watch a week?

_____ Under 1 _____ 1 - 2 _____ 3 - 4 _____ 5 - 6 _____ Over 6

8) Do you have a VCR?

_____ Yes _____ No

9) If you answered yes to #8, how many hours do you watch a week of movies, etc. (not including replayed commercial or ETV programs)?

_____

10) Rank the following reasons that you would answer the question "Overall, I watch TV...", starting with "1" as the most important reason and "5" as the least important.
Use a rank only once.

_____ for entertainment
_____ for relaxation
_____ for learning
_____ for relief from boredom
_____ to keep up socially
_____ other -- please specify

11) Do you feel that TV is a good instructional medium, and why?

Appendix H

Pre-experimental Session Oral Instructions For All Groups

Pre-experimental session oral instructions

For All Groups

First of all, I'd like to thank you all very much for allowing us to take up some of your time today.

My name is and I am a graduate student in the Educational Technology program at Concordia University. Currently, we are conducting research on the evaluation of certain films and video programs. Today we would like to collect some preliminary information, and next week we will ask you to evaluate a TV program in the same way that is done at CBC, TVOntario and so on. I think you'll find it fascinating.

For anyone who is interested in the results that we obtain from this particular study, we will be supplying a copy of the final research report to your teacher when it is completed.

As I just said, what I'd like you to do today is to complete three short questionnaires. The first questionnaire today will give us an idea of how much you know about the topic of next week's video, while the other two will give us some other general information which I will explain when we get to them. I'll pass out the questionnaires now.

[Wait until all learners have the forms.]

Take all the materials out of the envelope, and please keep the form labelled "QUESTIONNAIRE 1" on your desks, and set the other two aside for now. Please do not put your

name on any of the forms.

Now what I would like you to do is put the last four digits of your phone number in the box on the front of the envelope. This will allow us to match the forms you fill out today with the evaluation forms that you will be given next week. However, note that all of your responses will remain entirely anonymous. The number is for matching purposes only. What I would like you to do now is also put the last four digits of your telephone number in the upper right hand corner of the questionnaire, once again to ensure that we can match them up later.

The first questionnaire, as I said, asks you some questions about the video's topic -- langauge acquisition in children. Please simply do the best you can on these questions. You will have about 10 minutes to complete this first questionnaire, and I'll let you know when there are two (2) minutes remaining. Now please carefully read the instructions and begin. Are there any questions at this point? [Pause] If you have any questions while completing Questionnaire 1, please raise your hand and I'll come and help you individually.

[Remember to note the starting time.] [After eight minutes...]

There are two minutes remaining. [Wait two minutes.]

Please stop, and set this questionnaire aside. The next two questionnaires are designed to give us a general idea about your group's visual and verbal tendencies.

Television is, obviously, a manipulation of images and words. In order to effectively evaluate how these images and words deliver the message intended by the producer, it is extremely useful to measure group tendencies in terms of interpreting visual and verbal stimuli. Remember, all this information is anonymous, so don't worry about your performance. Just do the best you can. Also, after the evaluation next week, I'll explain all this in more detail and answer all your questions.

Now, pick up the form labelled QUESTIONNAIRE 2. What I'd like to do is read through the instructions out loud while you read along silently. If there are any questions regarding how to complete the questionnaire, I'll answer them before you begin. I'll begin reading the instructions now.

[Read instructions aloud.]

Are there any questions? You have about 10 minutes to complete this second questionnaire, and I'll let you know when there are two minutes remaining. If you have any questions, while completing this questionnaire, please raise your hand and I'll come help you individually. [Note the starting time.]

[After eight minutes...]

There are two minutes remaining. [Wait two minutes.]

Please stop, and set this questionnaire aside. Pick up the form labelled QUESTIONNAIRE 3. Read through the instructions. You have 10 minutes to complete this last

questionnaire, and I'll let you know when there are two minutes remaining. If you have any questions, while completing this questionnaire, please raise your hand and I'll come help you individually. [Note the starting time.]

[After 8 minutes...]

There are two minutes remaining. [Wait two minutes.]

Please stop, and place all three questionnaires back in the envelope. Make sure that the last four digits of your phone number are on the envelope.

[Collect the questionnaires and check to make sure that digits of phone number have been indicated.]

Thank you very much for your time. I will be back next week to ask you to evaluate the video program.

Appendix I

Experimental Session Oral Instructions

For PEAC Non-usage Groups

Experimental session oral instructions

for PEAC non-usage groups

Once again I'd like to thank you all very much for allowing us to take up some of your time today. Is there anyone who was not here last week?

[If yes, "You are welcome to participate in today's evaluation of a TV program. I'll answer any questions you might have later. Okay?"]

As we mentioned last week, the research we are conducting at Concordia involves the evaluation of certain films and videos. Today we would like your help in evaluating the effectiveness of an instructional program. The program was specifically designed to show how children acquire language and is intended for a cegep course in psychology or in child psychology. In other words, after watching the program, students should be able to understand the material presented. They should be able to analyze, synthesize and evaluate the basic concepts presented. While watching today's program, we would like you to assume the role of a student and learn as much as you can, as we will be asking you a number of questions afterwards about the program.

FOR STUDY 1: The overall question we are interested in is: How effective do you think this program is in demonstrating language acquisition in children?

FOR STUDY 2: The overall question we are interested in is: How good do you think this program is, based on your overall reaction?

The purpose of this evaluation is to help the producers to determine specific points when the program is very interesting as well as those points which are not interesting. It is by seeing the program's strengths and weaknesses through your eyes that the program may be improved.

Are there any questions before we begin?

[Show program.]

[When the program is over, hand out envelopes containing post-viewing questionnaires.]

Now that you've seen the program, I'd like you to answer the questionnaires being handed out. Take all the materials out of the envelope. Please do not put your name on any of the forms. What I would like you to do is put the last four digits of your phone number in the box on the front of the envelope. This will allow us to match the forms you fill out today with those you completed last week., Once again, note that all of your responses will remain entirely anonymous. The number is for matching purposes only. What I would like you to do is also put the last four digits of your telephone number in the upper right hand corner of each of the questionnaires.

All the questionnaires are related in various ways to

the evaluation of the program you just saw. Now please keep the form labelled "QUESTIONNAIRE 1" on your desks, and set the other two aside for now. The first questionnaire today will give us an idea of how much you remember about the content of today's video, as well as your opinions about its topic. The other two questionnaires will give us some other information which I will explain when we get to them. You will have about 15 minutes to complete the first questionnaire, and I'll let you know when there are two (2) minutes remaining. Now please carefully read the instructions and begin. Are there any questions at this point? [Pause] If you have any questions while completing Questionnaire 1, please raise your hand and I'll come and help you individually.

[Remember to note starting time. Also, play the amount of time by ear, depending on how many students finish within 15 minutes.]

[After 13 minutes]

There are two minutes remaining. [Wait two minutes]

Please stop, and set this questionnaire aside. The next questionnaire will give us an idea about your evaluation of specific features of this video. Remember, all this information is anonymous, so don't worry about giving your honest opinion.

You will have about 10 minutes to complete the second questionnaire, and I'll let you know when there are two (2) minutes remaining. Now please carefully read the

instructions and begin. Are there any questions at this point? [Pause] If you have any questions while completing Questionnaire 2, please raise your hand and I'll come and help you individually.

[Remember to note starting time]

[After 8 minutes]

There are two minutes remaining. [Wait two minutes]

Please stop, and set this questionnaire aside. The next questionnaire simply asks for some demographic facts about each of you. When you have completed Questionnaire 3, take all three questionnaires and put them in the envelope on your desk.

[Wait until everyone looks as though they've finished and confirm this by asking:]

Is there anyone who has not finished yet?

[Collect all questionnaires, once again making sure that the four phone digits have been indicated.]

Once again, thank you for your time. The results will be given to your teacher in the near future.

Appendix J

Experimental Session Oral Instructions

For PEAC Usage Groups

Experimental session oral instructions

for PEAC usage groups

Once again I'd like to thank you all very much for allowing us to take up some of your time today. Is there anyone who was not here last week?

[If yes, add "You are welcome to participate in today's evaluation of a TV program. I'll answer any questions you might have later. Okay?"]

As we mentioned last week, the research we are conducting at Concordia involves the evaluation of certain films and videos. Today we would like your help in evaluating the effectiveness of an instructional program. The program was specifically designed to show how children acquire language and is intended for a cegep course in psychology or in child psychology. In other words, after watching the program, students should be able to understand the material presented. They should be able to analyze, synthesize and evaluate the basic concepts presented. While watching today's program, we would like you to assume the role of a student, and learn as much as you can, as we will be asking you a number of questions afterwards about the program.

FOR STUDY 1: The overall question we are interested in is: How effective do you think this program is in demonstrating language acquisition in children?

FOR STUDY 2: The overall question we are interested in is: How good do you think this program is, based on your overall reaction?

The program is 25 minutes long.

The purpose of this evaluation is to help the producers to determine specific points when the program is very effective as well as those points which are not effective. It is by seeing the program's strengths and weaknesses through your eyes that the program may be improved.

In order to evaluate the program <u>during</u> viewing, we'll be using the response units which are being handed out now. They're very simple to use and you will use them to record your own reaction to the film while watching it. We'll go over exactly how they work once everyone has received one.

[Wait until everyone has a hand unit.]

Now, look at the window display on your hand unit and raise your hand as soon as you see a little red dot in the right corner. That means that the hand unit is ready to use.

[If all hand units are not synchronized, collect them and begin again.]

Each hand unit has been preprogrammed to record your own individual responses throughout the entire film. You'll notice the four buttons marked A, B, C, and D. These buttons will be used to record your opinion of

different parts of the program.

FOR STUDY 1: If, while watching the program, your overall reaction at a given moment is that the program is very effective, you press "A". If you feel that it is fairly effective, you press "B", while if it's not very effective, you press "C" and if, in your own opinion, the program is very ineffective, you press "D".

FOR STUDY 2: If, while watching the program, your overall reaction at a given moment is that the program is good, you press "A". If you feel that it is fairly good, you press "B", while if it's not very good, you press "C" and if, in your opinion, the program is poor, you press "D".

We've placed this chart up here beside the video monitor to remind you of what the letters represent. Use your own judgment to decide whether or not a particular moment of the program is good or not. For example, you may find certain things like the narration, or the scenery, or the content itself, or the characters, good or fairly good or not very good or poor. Once you have made a response by pressing a particular key, that letter remains shown in the display window. Try pressing the different keys now to see how they each appear in the display. [Pause] As you can

see, it is not necessary to keep your finger on a key to register a response. You may, of course, press the keys as frequently or as seldom as you like. It is only necessary, however, to press a key when your reaction changes. Does everyone see how they work?

Now, just before we start viewing, let's give the film the benefit of the doubt by pressing the "B" key for

FOR STUDY 1: "fairly effective".

FOR STUDY 2: "fairly good".

This allows us to synchronize all of your responses when we analyze the data later. Once you have pressed the "B" key, please make sure you don't press another one until the program begins.

Again, are there any questions about how to use the hand units?

[Show program.]

[When the program is over, hand out envelopes containing post-viewing questionnaires.]

Now that you've seen the program, I'd like you to answer the questionnaires being handed out. Hold on to your hand units for a moment. Take all the materials out of the envelope. Please do not put your name on any of the forms. What I would like you to do is put the last four digits of your phone number in the box on the front of

the envelope. This will allow us to match the forms you fill out today with those you completed last week., Once again, note that all of your responses will remain entirely anonymous. The number is for matching purposes only. What I would like you to do is also put the last four digits of your telephone number in the upper right hand corner of each of the questionnaires.

Also, I'd like you to notice the number on the top of your hand unit. [Hold master unit up in the air to show where the number is to be found.] Please write this number just below the four digits of your phone number, on the envelope. This is to make sure that your responses on the hand unit and your responses on the questionnaires can be related to each other.

All the questionnaires are related in various ways to the evaluation of the program you just saw. Now please keep the form labelled "QUESTIONNAIRE 1" on your desks, and set the other two aside for now. The first questionnaire today will give us an idea of how much you remember about the content of today's video, as well as your opinions about its topic. The other two questionnaires will give us some other information which I will explain when we get to them. You will have about 15 minutes to complete the first questionnaire, and I'll let you know when there are two (2) minutes remaining. Now please carefully read the instructions and begin. Are there any questions at this point? [Pause] If you have any questions while completing

Questionnaire 1, please raise your hand and I'll come and help you individually.

[Remember to note starting time. While everyone is filling out Questionnaire 1, collect hand units, making sure that the identification numbers are indicated on envelope. Also remember that students may need more than 15 minutes. Play be ear.]

[After 13 minutes]

There are two minutes remaining. [Wait two minutes]

Please stop, and set this questionnaire aside. The next questionaire will give us an idea about your evaluation of specific features of this video. Remember, all this information is anonymous, so don't worry about giving your honest opinion.

You will have about 10 minutes to complete the second questionnaire, and I'll let you know when there are two (2) minutes remaining. Now please carefully read the instructions and begin. Are there any questions at this point? [Pause] If you have any questions while completing Questionnaire 2, please raise your hand and I'll come and help you individually.

[Remember to note starting time]

[After 8 minutes]

There are two minutes remaining. [Wait two minutes]

Please stop, and set this questionnaire aside. The next questionnaire simply asks for some demographic facts about each of you. When you have completed Questionnaire

3, take all three questionnaires and put them in the envelope on your desk.

[Wait until everyone looks as though they've finished and confirm this by asking:]

Is there anyone who has not finished yet?

[Collect all questionnaires, once again making sure that the four phone digits have been indicated.]

Once again, thank you for your time. The results will be given to your teacher in the near future.