AD A050905

# MAKING TEACHING CONCEPTS MORE EFFICIENT

Technical Report #1

1 Jan 1978

ARPA Order No. 3395

Name of Contractor:
Learning Research & Development Ctr.

Effective Date of Contract:
1 June 1977

Contract Expiration Date:
30 September 1978

Contract No. MDA903-77-C-0208
Principal Investigator:
Karen K. Block
(412) 624-4919

Short Title of Work:
Making Teaching Concepts
More Efficient

AD No.
DDC FILE COPY

DDC
RECEIVED
MAR 16 1978
F

Sponsored by the
Defense Advanced Research Projects Agency
Under ARPA Order No.: 3395
Contract No.: MDA 903-77-C-0208
Monitored by Leon H. Nawrocki: (ARI)

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Projects Agency of the U.S. Government.

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

# Efficiency in Teaching Vocabulary Concepts: Attainable or Not?[1]

Karen K. Block and Ellen S. McCaslin[2]

Learning Research and Development Center

University of Pittsburgh

[1]This research is sponsored by the Defense Advanced Research Projects Agency under ARPA Order No. 3395; Contract No. MDA903-77-C-0208; and Monitored by Leon H. Nawrocki, at The Army Research Institute. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Projects Agency of the U.S. Government.

[2]The authors would like to express their appreciation to their colleagues at the Learning Research and Development Center for helpful comments during the preparation of this report. The report directly reflects consultative comments from Charles Perfetti and Alan Lesgold, and indirectly reflects helpful suggestions from Isabel Beck, James Voss, James Greeno, Lauren Resnick, and Robert Glaser. Thanks are also due Ms. Debbie Lipesky for her help and great skill in typing and to Ms. Sandy Law for her assistance in final typing.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

# REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report #1

2. GOVT ACCESSION NO.:

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): Efficiency in Teaching Vocabulary Concepts: Attainable or Not?

5. TYPE OF REPORT & PERIOD COVERED: Technical Report No. 1, 1 Jun [illegible] - 1 Dec 77

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): Karen K./Block, Ellen S./McCaslin

8. CONTRACT OR GRANT NUMBER(s): MDA903-77-C-0208, ARPA Order-3395

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Learning Research and Development Center
3939 O'Hara Street
Pittsburgh, Penna. 15260

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:

11. CONTROLLING OFFICE NAME AND ADDRESS:
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

12. REPORT DATE: 1 Jan [illegible]78

13. NUMBER OF PAGES: 101 (193p.)

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):
Army Research Institute
5001 Eisenhower Avenue
Alexandria, Virginia 22333

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

DDC RELEASED MAR 16 1978 F

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Concept instruction; instructional design; transfer in instruction; instructional decision making; efficiency in instruction.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Please see next page

388006

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

There are a number of performance outcomes ($O_i$) associated with mastering a concept; we have identified seven and are developing seven instructional modules ($M_i$), each targeted toward one $O_i$. A maximally effective instructional sequence would be comprised of all the $M_i$. Our research is aimed at selecting the set of $M_i$, a subset of the total set, that satisfy given performance criteria.

The design of the $M_i$ was influenced by previous studies of instructional methods, design theory prescriptions, and selected psychological theories of semantic memory. The first stages of research require calibrating the size of the direct and indirect effects of each $M_i$; later stages will involve studying how selected combinations of $M_i$ influence levels of performance on the criterion tests and confirm or disconfirm the decision theory framework. Several "experimental" tests will also be used to assess the nature of the information that is learned.

The research will be conducted in an applied setting - a College Learning Skills Center. To date, 30% of the courseware materials needed have been designed. Future efforts will be devoted to further development of the decision theory framework, completion of the instructional database, and design of the CAI system that instruments the research.

ACCESS
DDC B.U. Section
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
DIST. SPECIAL
A

## SUMMARY

This program of research has set out to address the issue of efficiency in instruction in teaching vocabulary concepts. There are a number of performance outcomes associated with "knowing" a word; we have identified seven such outcomes and are developing seven instructional modules, one targeted toward each outcome. Instruction of this sort should be effective, but not necessarily efficient, in the sense that a great deal of learning time would be spent going through each of the seven modules in order to learn each outcome. We hypothesize in this technical report that it will not be necessary for subjects to receive instruction via all seven modules in order to achieve acceptable levels of performance on the criterion tests measuring each outcome. We hypothesize through a formal statement of the instructional focus decision that each teaching module will have a direct effect shown by performance on its related criterion test, but that it will also have an indirect effect, smaller than its direct effect, on the criterion tests measuring the other learning outcomes. If this is true, some smaller combination of teaching modules than the full set should produce acceptable levels of performance on the outcome measures, at a lower cost. The purpose of our program of research is to evaluate the hypothesis proposed by the decision theory framework and to determine the optimal combinations of modules to teach vocabulary concepts should the framework be confirmed.

The seven instructional modules are described in the report, five on a general level and two specifically, in terms of their instructional routines; i.e., the sets of stimuli, responses, and feedback each contains. The rationale for each module is presented, citing whether the module evolved from an instructional design point of view (e.g., an application of Merrill and Tennyson's 1977 work) or whether it is an application of psychological theory (e.g., Frase's 1975 model of prose processing; Rumelhart, Lindsay, and Norman's 1972 semantic network model, etc.).

To achieve relatively tight control over student learning processes, a CAI (Computer-Assisted Instruction) mode will be used to instrument the research. We have accepted delivery of two VT-52 DECSCOPE terminals and two sets of Vadic Corp. modems. During any given week, college student subjects will be pretested on general vocabulary concepts; they will then receive instruction on the words that were missed. They will return after an absence of one day to take an interim posttest, with automatic reteaching following on the words missed. Following another day's absence, they will return for a final session to take the battery of criterion tests and several "experiemntal" tests that we are developing to learn about the structure of information acquired.

The first stage of research requires calibrating the size of the direct and indirect instructional effects; later stages require studying how the combined use of the methods confirms or disconfirms the decision theory equations. In summary, the main accomplishments

to date of this research endeavor include: 1) A formalization of the decision theory framework that guides this research; 2) Descriptions of the seven instructional modules at a general level and a rationale for each; 3) The introduction of a cognitive psychology perspective on the design research so as to aid understanding of learning; 4) The rationale and design of two of the instructional modules in detail; and 5) The rationale and design of the pretest.

The main tasks toward the completion of which work is currently progressing include: 1) Completion of the design of the computer based instructional research system; 2) Completion of the detailed design of the remaining five instructional modules; 3) Development of the instructional database with the vocabulary words that have been selected; 4) Design of the criterion tests; and 5) Design of the "experimental tests."

TABLE OF CONTENTS


| Section | Page |
|---|---|
| Addressing a Need for Efficiency in Instruction | 1 |
| Formulating the Design Problem | 3 |
| Formulating the Instructional Focus Decision | 14 |
| The Johnson and Stratton Study: Evidence for Indirect Instructional Effects Only? | 17 |
| Designing Instruction to Achieve Direct Effects Too | 25 |
| General Conditions on the Design of the $M_i$ | 30 |
| Gneral Features of the $M_i$ to Be Tested | 44 |
| Design of the Computer Assisted Experimentation System | 55 |
| Procedures for Word Selection | 58 |
| Materials Design: Pretest Construction | 63 |
| Components of the $M_i$ | 69 |
| The CLASSIFY Modules: Rationale and Design | 71 |
| CLASSIFY Routine Composition and Sequence | 74 |
| The DEFINE Module: Rationale and Design | 79 |
| DEFINE Routine Composition and Sequence | 79 |
| Test Development | 84 |
| Summary of the Research Endeavor to Date | 86 |
| Reference Notes | 88 |
| References | 89 |
| Appendix A: Numerical Examples of the Decision Theory Framework | 94 |

## Addressing a Need for Efficiency in Instruction

In these days of retrenchment, it seems important to re-examine our current procedures for designing instruction to determine whether they might be revised to make instruction more economical and efficient. Current instructional technologies contain, for the most part, guidelines and procedures that have been shown to be effective for accomplishing training objectives. However, for those designers who take the matter of training efficiency fairly seriously, effectiveness is not the sole criterion for selecting training procedures. Designers who want procedures that are both effective and economical will recognize a continuing need to analyze current practices to identify those practices that might be revised to achieve greater economy.

The need to achieve economy along with effectiveness in instructional procedure is not a new objective for instructional technology. The problem with past efforts, however, is twofold: First, the question of efficiency has not often been attacked directly, and, second, findings have not been systematically related to design theory. There seems to be little in the way of past research that presents alternatives in instructional procedure that vary in efficiency and then examines results on post-test measures of achievement, or other more pragmatically-derived measures.

Research in the area of tailored testing (see Ferguson & Hsu, 1971) is a pertinent example of research targeted toward

achieving greater efficiency in test procedure. That research grew out of, and was, in turn, related to test design theory. To date, there appears to be no good analogue of this type of research in the field of instructional design.

In researching efficiency of design, one would begin by defining instructional procedure consonant with current practice; one's goal would be to identify a substantially-reduced procedure that is functionally equivalent to the initial, more elaborate procedure. The contracted procedure would reduce instructional time and development costs. Some critics might view research with these objectives as crass materialism without redeeming features. However, we believe that this research objective can be defended on humanistic as well as materialistic grounds, and, most importantly as we hope to show, poses some interesting scientific questions.

Elaborating the steps of the above procedure, it is first necessary to specify the maximally-effective (ME) instructional procedures for the particular instructional objectives being considered. Each ME procedure would be consistent with current design theory and, if possible, would be confirmed by observations of actual practice. By definition, the ME procedures would be designed to do the best job possible, with minimal regard for instructional cost (where cost might be defined in various ways, such as student time spent in learning or development expense). Next, the ME procedures should be analyzed to identify points

where economies might be achieved while permitting outcomes to be maintained within an acceptable range. Of crucial importance to the identification of such points is information about functional relations between parameters of instruction, considered alone and in combination, and levels of outcome performance that can be expected. Central also is information about generalized effects of instruction as well as its more specific effects, since this information would provide a basis for retaining some tasks and removing others.

In summary, research with the objective of making instruction more efficient requires undertaking several different steps: First, it is necessary to define what current design theories suggest as the maximally-effective (ME) procedures for the strategy questions faced. Then, points in the effective procedures are identified where the instruction can be made shorter or simpler. Then, research problems associated with hypotheses of functional equivalence are identified and pursued. The research to be described in this paper represents an attempt to apply this approach in order to identify more efficient methods of instructional design for teaching vocabulary concepts.

## Formulating the Design Problem

The importance of vocabulary concept acquisition must be underscored, for in every subject area and in every training setting, new vocabulary concepts must be taught. These concepts can be technical vocabulary associated with a particular field

of application or they can be general vocabulary taught to improve students' general language facility, enabling them to read more difficult text materials and to communicate with more variety and precision. Because concepts are prerequisite to reaching numerous educational and training objectives and because significant amounts of instructional time are spent in teaching vocabulary concepts, it is important that the technology of teaching concepts become more highly developed. By highly developed, we mean, of course, that procedures used must not only be effective, but must also be demonstrably efficient.

When designers refer to building student "understanding" of new words taught, they imply that such understanding is reflected in a variety of student performances. Terminal objectives for a program of instruction in vocabulary might typically include any of the following: recalling and explaining the meaning of a word, differentiating between examples and nonexamples of a concept, recognizing a definition or a synonym, constructing sentences that use a word correctly, constructing new examples of a concept, or using a word to understand sentences or to communicate ideas in writing or through speech.

In the terminology of Bloom's (1956) taxonomy, one would thus want to build capability ranging from the knowledge level (where given a target word, the student recalls its meaning or recognizes its meaning when presented among a set of distractors) to higher levels, such as the application level (where given ideas

whose content matches the conditions of a target word's use, the student uses the target word in written or oral expression of these ideas). Each of the other objectives stated previously can also be found at some level of Bloom's taxonomy.

In considering the question of student performances that reflect "knowing" the meaning of a word, Cronbach (1942, 1943) suggested that word knowledge is multi-faceted, and that to measure it, one would need several types of test items comprising structurally different formats to assess competence on each aspect. Different types of word knowledge requiring different types of tests suggest that target objectives for instruction in vocabulary concepts involve qualitatively different behavioral outcome statements. The crucial question for the specification of instructional procedure is to determine whether different conditions of instruction are truly required for the various outcome statements.

Current design theories do indeed suggest that different instructional conditions are needed for different types of learning outcomes. In fact, the whole point to creating taxonomies or typologies of learning outcomes for instructional purposes is that the resulting types specified entail different conditions of instruction. This point can be illustrated in reference to two of the previously-identified vocabulary objectives, namely, stating a definition and classifying instances. In Gagné's

(1977) typology, stating a definition would be considered as verbal information learning, while classifying would be considered as a specific intellectual skill, i.e., learning a concept.[1] The two behaviors would thus be built through different training sequences. In a similar vein, Klausmeier (Klausmeier, Ghatala, & Frayer, 1974) has also consistently emphasized in his own model for the development of concept learning the need to distinguish classifying from defining. Klausmeier stresses that the two skills do not automatically develop together. One can also point to empirical research (Johnson & O'Reilly, 1964) demonstrating that conditions of learning that produce competence in classifying will only produce competence in defining when defining practice is given during concept acquisition. In further support of this point, Eustace (1969), who applied Gagné's hierarchy theory to design instruction for the concept noun, argued that separate behavioral levels must be considered in laying out instructional objectives, and that associated with the different levels are different instructional conditions.

Based on these considerations, the formulation that clearly emerges from a design theory perspective is that there are i learning outcomes that need to be separately considered and that

---

[1] Gagné's system also encompasses other vocabulary objectives: recognizing a synonym could be considered a verbal chain, constructing sentences could reflect rule application, constructing new examples could be a cognitive strategy, and using a word in communication could reflect applying a higher order rule.

there are i sets of qualitatively different instructional conditions, each optimal for one of the i outcomes.[2] A one-to-one matching rule is in effect, i.e., for each of the i terminal objectives there is a matched instructional procedure.

Two questions resulting from the design theory perspective now come into play:

> Is there an optimal sequence of learning tasks within the instructional procedure for an objective?
>
> Is there some way to merge the instructional procedures across objectives to enable fewer than the full set to be developed? (i.e., Through suitable selection of some outcomes for direct instructional focus and suitable design of instructional task features, is it possible to achieve j outcomes directly and i - j outcomes incidentally?)

The answer to both of these questions is critically dependent upon our knowledge of transfer.

For the first question, the sequencing question, the answer comes from empirical studies where the supposition is that ability to perform one task facilitates the learning of a second. The existence of facilitation is the logical basis for placing

---

[2] We do not want to specify the value of i at this point. It is possible for i to range from a minimum of two (representing the simple extremes of "knowing" and "being able to use") to some unknown higher maximum value that could be obtained by carving up the response repertoire into finer and finer pieces. For our formulation to "work", all we need to be able to claim is that there is some number of outcomes greater than one for which different instructional conditions exist. We believe that case is made in the text.

one task before another in an instructional sequence. Resnick (Resnick with Ford, 1976) reports on a set of studies that assessed the outcomes of instruction on children who did arithmetic tasks in different orders. The studies showed that teaching in hierarchical sequence is the best way to assure that all children in a group learn the objectives. Resnick also reported, however, that a minority of children were able to learn the more complex objectives without intervening instruction. This finding could provide the basis for more economies in instructional design in the context of sequencing questions where real or potential hierarchical relations exist. Resnick herself noted that skipping of prerequisites was in fact a faster way to learn.

The answer to the second question, the instructional focus question, must come from studies that assess both direct and indirect outcomes of instruction.[3] Data from these kinds of studies provide a basis for deciding which objectives must be directly taught and which are natural consequences of direct teaching to the others. The central matter of concern to make instruction more efficient, as we are here defining efficiency, is the extent

---

[3]We are distinguishing between sequence and focus questions in two ways, logical and empirical. Sequence questions are concerned with optimal ordering of a fixed number of instructional tasks. Focus questions are concerned with reducing the number of instructional tasks required to meet a fixed number of objectives. The questions are also distinguished in terms of the nature of the evidence that designers need to be able to make design decisions that permit more economical procedures (see text).

to which designers are currently willing and/or able to make instructional focus decisions. Willingness to consider transfer in design decisions is a function of designer attitudes toward transfer; ability to make design decisions based on transfer is a function of our research knowledge base.

Designers exhibit various attitudes toward transfer which have been built through experience with various program development efforts and with various populations of students. That designers are not totally in accord in their attitudes toward direct teaching and transfer can be illustrated by this example: We asked two experienced designers with credentials in the field to respond to an instructional focus question applied to our vocabulary objectives.[4] The question was posed:

> If you were teaching new vocabulary words and you wanted to produce a variety of capabilities ranging from the knowledge level, like knowing a synonym, to higher levels, like classifying and applying, what kinds of experiences would you provide in the instructional sequence? Specifically, would you provide practice on each type of task that you wanted the students to master?

One designer stated that indeed she would. The second designer said that she might not, but that she did not know at this point what types of tasks would be included in the instructional sequence. The second designer went on to speculate about the

---

[4]Both designers questioned are Research Associates at the Learning Research and Development Center, University of Pittsburgh.

existence of transfer relations among the criterion tasks but she stated that she could not specify what those relations would be without first laying out the tasks. Putting aside questions of reliability of a one-item test, along with questions of validity (e.g., do designers with different attitudes actually produce different instructional products?), the above responses do suggest variation in developers' attitudes toward transfer.

However, before we abandon as the prevalent position our one-to-one matching rule derived from design theory in favor of a more liberalized opinion taking possible transfer into account, let's reflect for a moment upon developments in the instructional design field at large. With the advent of structured curriculum models, analytic tools such as task and hierarchy analysis have come into frequent use. When applying these techniques, attention is focused on specifying component behaviors for the purpose of insuring that none is overlooked in instruction. The trend has been toward more careful analysis of terminal behaviors in order to insure that there are no gaps in the instructional sequence. It seems reasonable to suggest that careful skill descriptions might give rise to a related tendency to teach directly each skill identified.

Designer attitudes toward transfer are also no doubt influenced by the learner characteristics of their target populations, as well as by the task characteristics. Many studies

that have investigated the role of intelligence have found that more intelligent students show greater transfer (Craig, 1953; Werner, 1930). Ellis (1965) suggested that a reasonable interpretation of this finding is that brighter students tend to seek out relationships and are, therefore, more likely than the less bright students to have a set for transfer. In these days of universal education, however, designers are concerned that all students learn. In the curricular area of beginning reading, where reasonably well conducted curriculum evaluations exist, the emerging conclusion (for example, Beck & McCaslin, in press) is that direct instructional models using a "specific skills" approach are among the more effective curricula for populations of slower learners. The theory of teaching tasks advocated in one educational psychology text (Becker, Englemann, & Thomas, 1971) supports this position also; its authors maintain that children learn only what they are taught, a point given further emphasis by Gage and Berliner (1975) who state in reference to Thorndike's transfer of substance notion that "these studies of transfer make us aware that we might often do better to teach children exactly what we want them to know" (p. 172).

Thus, despite the second designer's comments concerning the likelihood of transfer (suggesting that fewer than $i$ tasks would be needed to build $i$ learning outcomes), it seems fair to suggest that the attitude toward transfer apparent in the field is more

conservative than liberal. We will conclude that the majority of designers follow the design theory prescription and adhere to a one-to-one matching rule.

Frase (1978), in discussing the newer analytic approaches for characterizing test performance (and hence instructional objectives), aptly notes that the current tendency toward greater specificity in skill definition has led to failures to consider more general processes, i.e., abilities that could support performance across a wide variety of tests. If this point were pressed, it might force the development of design strategies that explicitly consider instructional focus questions. Currently, such strategies seem to be the exception rather than the rule.

To summarize, we have thus far attempted to make the case that a significant portion of designers would likely follow a one-for-one matching rule when developing instruction for i qualitatively different learning outcomes. If transfer really exists, however, the instruction they will design will be inefficient in the sense that such instruction teaches more than is necessary. In applying the most conservative attitude toward transfer, each instructional task would be targeted toward a given criterion task; in order for students to master each objective specified, they would be provided with experience with each instructional task. Obviously, such a minimal transfer

assumption leads to an instructional sequence of maximal cost in terms of instructional time and development costs.

By contrast, if it were possible to assume that a subset j of the i instructional tasks could produce acceptable levels on the j matched criterion tasks and also on the i - j unmatched criterion tasks, then significant savings in instructional time and development costs could be achieved. The key element for achieving greater efficiency in concept instruction is the existence, to a degree, of indirect as well as direct instructional effects.

In our proposal (Block, Note 1), we suggested that, in the area of vocabulary concept instruction, there is some basis for the belief that various instructional methods produce substantial indirect instructional effects. We shall treat the adequacy of this evidence in a forthcoming section. At this point, however, we feel it necessary to achieve a more precise formulation of the instructional focus question.

Formulating the Instructional Focus Decision

In order to make concept instruction more efficient, it is necessary to make instructional focus decisions. The choice situation, stated more formally, is this:

1. There are n learning outcomes, $O_i$, and the same number of instructional methods, $M_i$, and criterion tests, $CT_i$.

2. For each $O_i$, there exists a unique instructional condition, $M_i$, that produces a higher level of performance on a given criterion test, $CT_i$, than for any other single instructional condition. This instructional condition, $M_i$, also tends to elevate the levels of performance on the other $CT_j$ for $j \neq i$, but not, in general, to the extent that it raises the level of performance for $CT_i$.

3. Associated with each $M_i$ is an instructional cost, $c_i$.

4. Associated with each criterion test is a lower bound of acceptable performance, $lb_i$.

5. Performance on any $CT_i$ is a function of the number and type of $M_i$ encountered.

6. Performance on any $CT_i$ can be decomposed into direct and indirect instructional effects. Let:

$\Delta_i$ = direct effect of $M_i$ on $CT_i$ for all i.

$\sigma_{ij}$ = indirect effect of $M_j$ on $CT_i$ for all i, j.

(i) Assume $\Delta_i > \sigma_{ij}$ for all i, j.
(ii) Assume $\Delta_i = \Delta_j$ for all i, j.
(it is not assumed that $\sigma_{ij_1} = \sigma_{ij_2}$ for $j_1 \neq j_2$).

and so, therefore,

$$CT_i = \Delta_i + \sum_{j \neq i} \sigma_{ij}, \text{ if all } M_i \text{ are used. (the ME sequence)}$$

If we let $S = \{j | M_j \text{ is used}\}$ and $\sigma_{ii} = \Delta_i$, then

we can state that

$CT_i = \sum_{j \in S} \sigma_{ij}$, if only some of the $M_j$ are used. (the reduced sequence)

7. Total instructional cost, $C_i$, is the sum of costs of individual $M_i$, that is,

$$C_i = \sum_{j \in S} c_j.$$

(It is not assumed that costs are reduced when more than one method is used).

8. The instructional focus question entails selection of $k<n$ methods so that total instructional cost is minimal and a stated performance criterion is optimized by suitable choice of the $M_i$. The goal might be to:

   a. Select the $M_i$ to insure that performance on all $CT_i$'s does not fall below $lb_i$. That is, to max (min ($CT_1$, $CT_2$, ... $CT_n$) subject to $C_i \leq W$.

   b. Select the $M_i$ to maximize the average of levels of performance on the $CT_i$. That is, to max

$$\left( \frac{\sum_{i=1}^{n} CT_i}{n} \right)$$

   subject to $C_i \leq W$.

   c. Select the $M_i$ to maximize the maximum level of performance on a given $CT_i$. That is, to max (max ($CT_1$, $CT_2$, ... $CT_n$) subject to $C_i \leq W$.

We have explored several hypothetical (see Appendix A) numerical examples and have found that, for the same cost limitation, one can obtain a different $M_i$ composition in the solution set depending upon the performance criterion to be met. Further, whether or not different solution sets are entailed when using different performance criteria is a function of the size of $\Delta_i$ and $\sigma_{ij}$, the direct and indirect effects. We shall delay serious treatment of scaling issues at this point however.

In a general way, the above formulation seems to reflect a fair picture of how things might work in a situation where instructional focus decisions would be made. It cannot be regarded as a thoroughly substantiated theory because the assumptions made require more careful and extensive testing against the data from studies of transfer. However, the reason we wanted to state the decision problem more formally is that we wanted to be as explicit as possible in regard to the main elements of the decision as we see them and we wanted a theoretical framework to help guide and sharpen our research on instructional focus decisions.

While main elements of the above decision problem seem to be in accord with design theory and selected empirical facts, it is necessary to discuss certain aspects of the formulation more fully. Specifically, the major assumption on which the hypothesis of efficiency is based is the existence of both direct and indirect instructional effects which are the outcomes of various methods for teaching vocabulary concepts. The study most

directly pertinent to this assumption is the study performed by Johnson and Stratton in 1966 which examined five methods of teaching vocabulary concepts. The Johnson and Stratton design was such that both direct and indirect transfer effects were assessed. Since this study provides an important context and point of departure for our work, it will be examined here in detail.

## The Johnson and Stratton Study: Evidence for Indirect Instructional Effects Only?

The Johnson and Stratton study is a classic one that is mentioned in numerous educational psychology texts. In that study, general psychology students learned four new vocabulary concepts: opulent, alacrity, altercation, and chide (Thorndike-Lorge frequency counts between 1 and 6) by one of five teaching methods. There were four single methods (definitions, sentences, classification, and synonyms) and a mixed method. The single methods were designed to be as distinct from one another as possible (i.e., both the stimuli presented and types of responses required were different among the four methods).

The first method, the definitions method, consisted of presenting students with definitions. The definitions were written with the help of several dictionaries to characterize each word in a specific way and to place it in a higher-order class (i.e., a genus et differentia format was used). For example,

the definition used for altercation was:

> When two or more people express different opinions, get excited, and contradict each other, the event is called an altercation. Thus an altercation is a social interaction characterized by heated exchange of opposing arguments. Now write a definition of altercation in your own words.

Definitions for all four words were printed on one side of a sheet of paper, with spaces for responses after each word; hence the definitions were in view as the students paraphrased.

The second method, the sentence method (known also as a words-in-context method) consisted of a short story of 174 words in which each of the four word concepts appeared twice. The story had "easy" context in that only the four words should have been unfamiliar. Subjects were instructed to read the story and learn the four new words. At the bottom of the page were four incomplete sentences which students were instructed to complete by using each of the four new words once.

The third method, the classification method, was presented as a booklet. It consisted of short descriptions of objects and events, arranged in blocks of five, one classifiable under each of the four concepts and one irrelevant. At the top of each block, the four new words and "none" were printed. Subjects were instructed to classify each of the items by writing the correct word on a blank line next to the item. The first block of five items was presented on the first page of the booklet; the answers to the first block and the second block of five items were continued on the

second page, etc. There were six blocks, of five events each, to be classified, and correct answers were given for each item. The booklet was arranged to make it easy for subjects to turn back to check answers.

The fourth method, the synonyms method, consisted of presenting subjects with a booklet for synonym training. The first page of this booklet informed subjects that their task was to learn the meanings of four concepts. Four short statements appeared next: "Alacrity means eagerness. Altercation means squabble. Chide means to criticize. Opulent means luxuriant." Then each target concept was listed below the statements with a blank next to it; four other synonyms were also listed there (e.g., reproach, quarrel, lavish, and promptness). The subjects were required to match each of the four concepts to the appropriate synonym by writing one synonym in each blank. Answers were shown on the next page, along with another block of four different synonyms to be matched to the same four concepts. Four blocks of four synonyms were given in this manner. The arrangement made it easy for subjects to check answers.

The fifth method, the mixed method, was a combination of the four preceding methods, with some editing of individual methods. The definitions were abridged. Each new word appeared only once in context of a sentence. Then two synonyms and one example were given. Johnson and Stratton provide this illustration of the mixed method:

> To chide someone is to talk to him to get him
> to correct his mistakes. Chide means to criticize

> or reproach. Thus a mother might chide her children for fighting with each other. An example might be a group of fellows poking fun at a boy with dirty clothes. Now write in your own words what chide means.

Following the paraphrase requirement, students did one block of matching target concepts to their synonyms (with correct answers provided) and one block of classification (again with correct answers provided).[5]

The various methods were designed to require roughly the same amount of time to complete. Johnson and Stratton report that each method required about twelve minutes. The teaching materials were used in a classroom setting, and nine days later, students took various tests in the same classroom setting.

There were four differently-constructed test items for each of the concepts learned. Each item was designed to measure the specific instructional effects of one of the four single methods; hence three of the items measured the indirect instructional effects of any given single method. While Johnson and Stratton

---

[5]Johnson and Stratton edited the individual methods in combining them into a mixed method in order to make time spent in learning equivalent for each of the five methods tested. Their research question posed for the mixed method was therefore: given that instructional time is fixed, what are some reasonable ways to fill it using a combination of the individual methods? We point this out here because our definition of a mixed method will differ from Johnson and Stratton's since the research question we pose is different.

did not report the exact conditions of their testing procedure, we think it reasonable to assume them equivalent to the test procedure reported in the norming procedure for the same study.

On the test matched to the definitions method (the definitions test), the four words were printed with spaces for short responses. The directions told students to explain briefly what each word meant, even if they had to guess. On the test matched to the sentences method, there were eight incomplete sentences (two for each concept) in which the target word was present. Students were directed to add a few words to give each sentence a meaningful ending, even if they had to guess. On the classification test, 20 instances (four for each target concept and four unrelated) appeared in irregular order and students were directed to classify the items as one of the four concepts or "none of these." On the synonyms test, the target word appeared with ten possible choices; the ten choices included one synonym for each target word. Students were directed to select the synonym for the target word. There were four trials for each target word. The tests were given in the order described, an order hypothesized to minimize cumulative learning effects. The first two tests were collected as each was completed; the second two were apparently completed together, and then collected.

Definitions were each rated 0-4 by two judges using a dictionary as a criterion; 32 was the maximum score (highest sum of the judges' ratings was 8 x 4 words). Both sentences for each target word from

the sentence test were scored together and the combination rated 0-4; yielding an 8 (sum of two judges' ratings) x 4 words = 32 maximum. To place the other two test scores on the same scale, since each word had in effect four trials each on the classification and synonyms tests, the maximum score of 4 was doubled to 8, yielding a 32 maximum score for the classification and synonyms tests. Reliabilities of the judges' ratings (assuming the norming procedure ratings are comparable) yielded decent, but not exceptionally high, reliabilities in the .60-.97 range (median .78) for the definitions and sentence completion tests. For the classification and synonyms tests K-R 20 reliabilities were .65-.89 (median .79).

From these data on the stability of the scores and from the examples of test items reported by Johnson and Stratton, we might want to adapt our treatment of *direct and indirect* effects in our decision framework to allow for variations across replications. That is, we could adapt our treatment so that achieving a particular size of indirect effect would be probabilistic (or there would be a likelihood function relating the chances that any obtained value would be within a particular range). We shall not do this here, however, because we believe that other considerations, such as whether indirect effects were in fact attained, are currently more important. The reliabilities reported were sufficient for us to believe that their findings deserve serious consideration.

The Johnson and Stratton findings can be summarized briefly. In general, all methods groups (the four single method groups and

the mixed method group) performed significantly better than did a control group that had worked on an irrelevant set of materials during training. Also, the group that received mixed method instruction achieved higher scores than did any single method group. However, there are no significant differences within the four single method groups.

One statement made by Johnson and Stratton is encouraging with respect to a hypothesis that tasks not directly matched to properties of the test can produce practically significant amounts of learning. In summarizing pertinent aspects of their findings, they noted, "Transfer from training by one method to tests corresponding to the other methods was 100%" (p. 48). This is indeed an encouraging remark and accurately reflects certain aspects of their data. Presumably, the statement was based on the fact that no single method excelled in its influence on total (across all four) test scores. Each single method -- definitions, (sentences, classification, and synonyms) -- also produced similar levels of performance of each of the four tests considered separately. These data support the statement that training by $M_i$ (where $i \neq j$) supports performance on $CT_j$ to the same degree as training by $M_j$.

However, the data from the Johnson and Stratton study do not provide adequate evidence for the existence of direct instructional effects. Evidence of direct instructional effects would require a demonstration that $M_i$ produces significantly better performance on

$CT_i$, than does $M_j$ (or some pooled combination of $j \neq i$). Johnson and Stratton found no such evidence of specific instructional effects. On the contrary, groups receiving each of two single methods (definitions and synonyms) did not perform best on their corresponding test, in comparisons relating each single method mean to the pooled means of the other three groups. Further, as reported previously, no single method was superior to any other method on each of the four tests considered separately. As specified by our theoretical framework for instructional focus decisions, if there are no direct instructional effects, by definition there can be no indirect instructional effects. Hence, the Johnson and Stratton study was not an adequate test of our theoretical framework.

In the face of the Johnson and Stratton data, we still believe our theory to be correct and we are even encouraged by their data to pursue additional tests. They show that performance on $CT_i$ can in fact be elevated by $M_j$, which in turn leads us to believe that real indirect effects do exist. We think that the absence in the Johnson and Stratton study of direct effects poses for us the design problem of lesser difficulty; indirect instructional effects are generally much more problematic to try to achieve.

Some readers may assert at this point that it is not possible to simultaneously meet both requirements that we have established for an adequate empirical test of our theoretical framework. That is, they would charge that in order to achieve a demonstration of

direct effects, the indirect effects would necessarily become small and insignificant (therefore showing that more economical instructional procedures do not exist). However, we believe that such a supposition has yet to be demonstrated in our particular situation, and further, useful information will come out of work that quantifies the relative sizes of the effects obtained under certain conditions. Therefore, one goal of our research will be to design the $M_1$ so that direct instructional effects are obtained.

Designing Instruction to Achieve Direct Effects Too

In order to obtain direct instructional effects, it is necessary to gain tighter control over student learning processes. Commenting upon their large transfer effect, Johnson and Stratton state:

> One interpretation for the large transfer effect is that the subject sets concept acquisition as his goal, varying his methods on his own initiative and testing himself so that the intended differences between training methods vanish. Thus each method takes over some of the advantages of the mixed method (p. 53).

Their interpretation seems to suggest that if better control over the student learning process could be achieved, the large transfer effects might be reduced and hence a specific methods effect might indeed apear. A computer will be used to instrument our research and will permit a greater degree of control than Johnson and Stratton attained. Since we are developing the methods in a CAI mode, we will be able to manipulate more elementary instructional events than did Johnson and Stratton. For example, to control

learner processes in a definitions method, one might first require the learner to paraphrase each key aspect of the definition separately, then require the learner to put them together. In contrast, the Johnson and Stratton definitions method required a complete paraphrase "all at once." The unit of instructional control in our CAI example is smaller (and hence more elementary) because the events given instructional attention would presumably be components of the larger terminal performance.

Johnson and Stratton suggested another explanation for the large transfer effects. As it turns out, what they suggested in 1966 would be quite well received today:

> Another possibility [for the large transfer effects] is that the transfer occurs on the testing day when the subject treats the four tests as problems and uses whatever he can recall *to solve them.* One with training on sentences may recall the story, for example, and use this information to formulate a definition or to choose a synonym (p. 53).

In effect, Johnson and Stratton are suggesting here that transfer occurs at the time of the test when information stored in memory is combined with prior knowledge of certain verbal procedures (e.g., formulating a definition) and the two, working together, can support performance on a criterion test. By today's standards, of course, such a theory of test performance is both gross and incomplete. Nevertheless, information stored in memory does play an important role in test performance. That the structure and content of such information ought to be explicitly

considered in understanding transfer has been emphasized by Voss (Note 2).

One might wonder, in terms of the purposes of our research, how explicitly considering the representation of remembered information or the mental processes involved in criterion test performance will help us solve our particular applied problem. Using the Johnson and Stratton suggestion that $CT_i$ = information recalled + prior procedural knowledge, it is easy to see that performance on criterion tests will differ to the extent that information recalled is different from method to method (or to the extent that prior procedural knowledge is different, a variable which at this level of detail in a formulation is difficult to surmise from method to method). It is possible that distinctly different methods leave the same memory traces; if this is true, it might not be possible to obtain a specific methods effect, and it might not make sense to retain the theoretical framework previously stated for this area of application. The more that we can learn about the content and structure of information retained from exposure to a given method, and about the processes that use the information, the more we become able to discern whether our applied goals can realistically be met. In our research, we intend to use several "experimental" tests, in addition to criterion tests, in order to measure aspects of the knowledge that is learned from a given method.

Ten years have passed since the Johnson and Stratton work and much research has been performed to show that an adequate model of

performance on a criterion test takes into account more than just the nature of information learned during the instructional episode. A current model might suggest that performance on any criterion test following any $M_i$ will be a function of:

(1) The information about word meaning that can be retrieved at the time of the test. This information will be a function of:

   (a) the knowledge structure that was built during instruction (which in turn is a function of the type and structure of the information presented, how prior learning interacts with the information presented, and other properties of the instructional task, such as type of response required and content of informative feedback, all of which influence phases of cognitive learning -- encoding, storage, etc.);

   (b) the compatibility of any retrieval cues given by the test with the knowledge structure that was built;

(2) The extent to which the information recalled plus prior learning can support performance of the form required by the criterion test; and

(3) "Test taking" skills, i.e., skills specific to a particular type of test that can be learned through exposure to instructional tasks whose properties are similar to the test. (Such aspects as familiarity with responses and familiarity with correct structural form on tests requiring constructed responses such as definitions fall into this category.)

A model such as this provides a basis for understanding why $CT_i$ is high, or low, following an $M_j$, in terms of particular component factors. As such, it also provides a basis for selective emphasis of particular experiences, information, or procedures in the design of the $M_i$. To the extent that we can manipulate levels of

competence on the above factors through appropriately designed interventions, we can control levels of performance on the $CT_1$. The above model is simply another way of decomposing performance on $CT_1$; rather than considering it here to be a sum of direct and indirect effects, we are recognizing that performance can be analyzed another way. In working through the design problem that is the focus of this paper, it is very useful to refer to concepts contained within the above decompositon.

In summary, an analysis of the Johnson and Stratton research has been instructive with respect to identifying ingredients needed for an adequate empirical test of our decision framework. We noted that a main requirement is the achievement of direct instructional effects. To obtain them, we will use instrumentation allowing better control. We also noted that since criterion test performance is a function of knowledge built during instruction, we will attempt to measure that knowledge since it might provide a key for understanding the size of effects that can be expected.

## General Conditions on the Design of the $M_i$

As originally stated, the $O_i$ for a program of instruction in vocabulary represent qualitatively different performances with a word and lie along a continuum from simple unitary performances to more complex performances involving linguistic reasoning and application. There are alternative ways to divide up the response continuum; there are many verbs and their objects that could be used as learning outcome statements. For our purposes, the $O_i$ ought to represent a range from simple "knowing" to "being able to use." Convention suggests students ought to be able to tell you the meaning of a word, use it in a sentence, and select a synonym for it. To these conventional objectives, Johnson and Stratton added to the $O_i$ list the ability to classify instances, which at first appears to be simply an interesting twist. With further thought, however, a case can be made from an instructional design perspective for considering the meaning of a word to be a classificatory concept (see Block, in press).[6] It seems sensible, therefore, to retain the four Johnson and Stratton learning outcomes for our $O_i$ list. These learning outcome statements are:

1. Stating the meaning of the target word;
2. Selecting a synonym from a set of distractors;

---

[6]Our colleagues at SWRL have come to a similar conclusion (see Humes, 1976).

3. Differentiating between examples and non-examples of the concept; and

4. Completing sentences meaningfully using the target word.

Three additional outcomes come to mind as useful performances with a word. A fifth outcome might involve constructing new examples of the concept. This would be a natural extension of $O_3$, requiring generation of positive instances of the concept class. Another outcome showing ultimate mastery of a word would be its incorporation into productive vocabulary which in turn would support its spontaneous use in composition and speech. Finally, there might be an outcome reflecting a literacy about words, an understanding of lexical structures that tie words together that can be reflected in diagrams and continua. Therefore, three additional outcomes would be:

5. Constructing new examples of the concept;

6. Using the word spontaneously to communicate ideas in writing and thorough speech; and

7. Creating lexical structure such as Venn diagrams or continua within which the target word may be located.

We believe that these seven $O_1$ adequately reflect a broad range of different performances and that they are a creditable list. We have argued previously from a design theory point of view that $O_1$ and $O_3$ require different instructional conditions

($M_i$), and we could carry the argument through using Gagné's typology for the other $O_i$.

Each $M_i$ addresses itself toward a particular $O_i$. As previously stated, in order to meet the structural requirements of the design problem, it is necessary that each $M_i$ produce direct instructional effects on its $CT_i$ that are larger than its indirect effects on $CT_j$; but, at the same time, the indirect effects must not be small. That is, the $M_i$ must do a good job of specific teaching without being too good (in the sense of being too limiting). The rest of this section and the section that follows detail our rationale for the selection of the $M_i$ that will hopefully fulfill the requirements of our design framework. We first reexamined the Johnson and Stratton teaching methods in our quest for a reasonable $M_i$ composition.

A major empirical principle of training-to-test transfer is that transfer of training is greatest when the training conditions are highly similar to those of the ultimate testing conditions (Ellis, 1965). This means that the more similar the properties of $M_i$ are to those of $CT_i$, presumably will be the greater the direct instructional effects. This principle was at least partially applied by Johnson and Stratton in their instructional design. It should be obvious from the section of this paper that elaborated the Johnson and Stratton work that the response requirements of their criterion tests were very similar (and in some cases identical) to those of

the related training methods. This similarity of response properties clearly shaped their $M_1$ design. A remaining question, however, is how they decided which particular stimuli would be paired with which particular type of response within each $M_1$.

To review, the stimuli Johnson and Stratton used were: 1) a genus et differentia definition, 2) a brief synonym statement, 3) a story with the words appearing in context twice, and 4) concept instances to be classified. It seems reasonable to expect that learning would occur with S-R pairs other than the particular pairs used in their study. That is, given a definition as an S, students could probably complete a sentence using the target word as an R, as long as they had prior knowledge of the procedure of sentence construction. The response properties between training and test would still be identical (if this hypothetical teaching sequence were paired with the sentence construction test), but learning would happen with a different S-R pairing than that used by Johnson and Stratton.

We therefore need to understand what the basis might be for the particular pairings achieved. Johnson and Stratton did not provide a rationale for their pairings except that they wanted each method to be a good representative of that method and they wanted no overlap among methods. The former point implies that they thought there must be some consistent definition among teachers and other educators for each of the four methods of interest. However, there appears to us to be no such clear and

consistent definition of vocabulary instructional methods among practitioners and educators (see Block, in press). The methods definitions that are used mainly refer to the syntactic structure of the stimulus content (see Ellson, 1976) and hence are not well specified since they detail neither response content nor feedback. The reading instructional literature suggests four structural variations for a stimulus presentation but says nothing about other aspects of the instructional presentation.

The question before us concerns the extraction of a rationale for achieving four particular S-R pairings out of a domain of 4·4 = 16 possible pairings that could have been used. The rationale could have been that Johnson and Stratton wanted the content, or object, of the response (e.g., stated <u>what</u>? - stated a <u>definition</u>) to be reflected in the stimulus portion of the method. Hence their presentation of a definition as the stimulus paired with the definition paraphrase response. In this way, students did not have to recode the information presented into another linguistic form. By contrast, if the words had been presented in context and students were required to write a definition, they would have had to recode the stimulus information to do so.

On the other hand, the rationale could be partially reflective of laboratory learning paradigms. The classification method is a direct analogue of discovery learning (reception paradigm) procedures used with classificatory concepts (see Bourne, 1966). The synonym method is similar to repeated study-test trial

procedures for paired associates learning, although the content was not, strictly speaking, paired associates since the R terms varied, i.e., numerous synonyms were used as R terms for the target word (the S).[7] It is more similar to a paired associates concept learning procedure (see Bourne, 1966). In 1966, there were no laboratory paradigms for the definitions or the sentences methods (see Melton, 1964 for the paradigms of that day) but there are today. The sentences method could be related to paradigms for studying story comprehension and the definitions method to prose learning studies involving concepts, attributes, and connectives (see Frase & Silbiger, 1970). The central point about paradigms is not a point about differencies in procedure; however, what is of interest are the different information structures (an association, a classification rule, etc.) associated with a given type of learning paradigm. Varying the structure of the information taught might be a good rationale for the type of variation in the stimulus content to be represented in the design of the $M_1$.

Thus far we have noted that Johnson and Stratton designed their $M_1$ so that properties of the training response matched the properties of the criterion test response. This design procedures leaves open the question of the structure of the stimulus information, and it is necessary to introduce other considerations such as recoding or a

---

[7] Interestingly, this method of teaching meaning treats a word more like a relational, rather than a classificatory concept, a content distinction important in design theory (see Glaser, 1968).

reference to basic learning paradigms in order to rationalize the particular S-R pairings that were used. It is possible, however, that current versions of empirical transfer principles more extensively specify the properties needed during training for optimal performance on the test. Work by Merrill and Wood (1975) speaks to this question.

Currently, Merrill and his colleagues have devoted much attention to the problem of formulating instructional prescriptions that have empirical support (see Merrill, Olsen, & Coldeway, 1976). In current instructional design literature, Merrill and Wood have formulated the most specific statements concerning the relationship of instructional tasks to the task properties of criterion tests. Merrill and Wood have proposed a consistency hypothesis. The consistency hypothesis states that, "Student performance in relationship to a specified instructional task is facilitated to the degree to which the type of content operations and the response conditions involved are consistent among objectives, presentation displays, practice displays, and test displays" (1975, p. 2·8).

Display is a term used to refer to a piece of information given a student (e.g., a word definition, an example, task directions). In proposing the consistency hypothesis, Merrill and Wood are stating that not only must response conditions be similar between training and test, but also content operations must be similar. To understand the consistency hypothesis in terms of its implications for $M_1$ design, it is first necessary to understand the term content operation.

In Merrill and Wood's taxonomy of instructional design, instructional objectives are classified according to the content operation (identity, descriptive, productive) and response condition (rule remembering, rule using, rule finding) combination intrinsic to the objective. The term content operation refers to types of instructional content that might be thought to vary in complexity; an identity refers to equivalence or substitutability of two content elements; descriptive content is a logical combination (e.g., union, intersection, if and only if) of content elements; and productive operations are ways of combining content elements so as to produce qualitatively new elements. The clearest examples of contrasting content complexity come from mathematics: an identity is A = B; descriptive: definition of an isosoceles triangle; productive: finding the area of a triangle that is equal to 1/2 times the base times the altitude. The reason for the phrase content operation, rather than simply content type, or content structure, is that the basic elements of subject matter content are thought to be best described in terms of relationships between domain and range concepts (that is, as rules of various complexity). Various sorts of responses can be supported by these various forms of content knowledge and the response conditions, too, vary in complexity from simple "remembering" what was taught, to using it, to "discovering" it (rule finding).

Applying the above taxonomic system to our vocabulary objectives, it seems that an identity would be the semantic relation of synonymy,

and a descriptive operation would be a word definition stated in terms of concepts (and the relationship between these concepts) that comprise the definition. Just what a productive operation (which Merrill and Wood note is actually a proposition) would be is a bit less clear cut. It seems to reflect a generative aspect of content, i.e., producing new things from given ones. Merrill and Wood define a productive content operation thusly:

> The referents of the range concept are produced by composition, decomposition, or some other combination operation in which the referents of the domain concepts are qualitatively changed as they are combined to produce referents of the range concepts (p. 2·3).

In math, this means:

| | BASE | HEIGHT | | AREA OF TRIANGLE |
|---|---|---|---|---|
| referents of domain: | 30" | 15" | referents of range: | 225 sq. " |

In vocabulary, whether or not a productive operation is involved in a particular objective depends on one's interpretation of what constitutes a qualitative change. There are various $O_i$ that require the production of new referents of the range concept from quite different referents of the domain concept. Suppose, for example, using sycophant as the target word:

| referents of domain: | Lunchroom context: | referents of range: | |
|---|---|---|---|
| | The sycophant had to eat lunch alone because nobody would sit with him. | | The sycophant did not receive any get well cards when he was in the hospital. |
| | Classroom context: | | |
| | The sycophant was always last when the students chose up sides for the team. | | |

The productive rule might be: A sycophant is not well regarded by his peers. To produce the referent of the range concept it was

necessary to decompose (in the sense of develop specific interpretations for particular contexts) the phrase, "is not well regarded by his peers." It seems reasonable, therefore, that (at least) our $O_i$ dealing with sentence completion might reflect a productive operation (that supports performance at the using level). The $O_i$ dealing with stating a definition and classifying reflect descriptive operations (that support performance at the recall and at the using - classifying - levels) while the $O_i$ dealing with synonyms entails identity. (We shall not treat the other three $O_i$ for reasons that will be apparent).

The consistency hypothesis applies only to descriptive operations at the remembering and using levels (using a definition entails classification, according to Merrill and Wood, p. ii) and to productive operations at the remembering and using levels. What is critical for our purposes is the recommendations they make for instructional strategy design. The consistency hypothesis would apply to three of our first four $O_i$ (stating a definition, classifying, and completing sentences). Selecting synonyms would be eliminated because the consistency hypothesis does not apply to identities.

To review, the consistency hypothesis requires consistency of content operation and response conditions between the $CT_i$ and the $M_i$. Although the content operations for defining and classifying are the same, no economies in training design are

explicitly mentioned. That is, no attention is paid to possible dependencies in performances supported by the same content operation. Hence, the consistency hypothesis implies an $M_i$ for each $O_i$ that we have defined.

In reference to descriptive and productive operations at the recall and using levels, Merrill and Wood specify a set of instructional conditions that implement the most effective strategies for teaching the $O_i$. They believe that each segment of instruction should contain three presentation forms: rule, example, and practice; an adequate test consists of a series of practice displays constructed from previously unencountered instances. These prescriptions are the essential requirements for both recall and using objectives, for both types of content operations.

We applied Merrill and Wood's prescriptions to the design of our $M_i$ for the classification outcome. The criterion test for the classification outcome requires the student to use a descriptive operation in order to classify instances as examples or nonexamples of the target concept. The teaching sequence presents a rule (i.e., a definition of the target concept); example (i.e., labeled examples and nonexamples of the target concepts); and then practice (i.e., exercises where the student identifies for himself or herself which instances are examples and which are nonexamples of the target concepts). A later section of this paper describes the rationale and design of the classify teaching sequence in detail.

The consistency hypothesis and the associated prescriptions for strategy design also apply to the design of conditions for stating definitions. The instructional conditions would be the same as those for classify, except practice would be provided in recalling the definition during training, and the test (to be adequate) would require both recall and classification. While this prescription for design may indeed be effective, it reduces the chances of specific $M_i$ effects, and comes closer to being a "mixed" method of teaching rather than a purer form of matching training conditions to test conditions. And it would take more time. In view of the Johnson and Stratton results, useful amounts of learning can happen via paraphrasing definitions, without classification, and it seems that the purposes of our research are best served by testing direct and indirect effects of the simplest set of instructional conditions first. Therefore, our $M_i$ for recalling and stating definitions will involve only paraphrase and not classification.

The consistency hypothesis and related prescriptions for strategy design also apply to the design of conditions for completing sentences, another of our target $O_i$. A teaching sequence for the word sycophant would include the rule (a sycophant is not well regarded by his peers) and examples (the sycophant had to eat lunch alone because nobody would sit with him). Practice could consist of sentence completions where the ending is consistent with the proposition represented by the rule (nobody likes a

sycophant). The criterion test would also be such sentence completions. With this teaching sequence for the sentence completion criterion task, it seems probable that the outcome of learning would be competence with only this particular criterion task. That is, students may be able to complete sentences by describing their specific reactions to unpopular people in various settings, but they would by no means know why people don't like sycophants and what sycophants' characteristic behaviors are. Therefore, following the Merrill and Wood prescription for the sentence completion $O_1$ would seem to unnecessarily restrict the size of the indirect effects. Based on the considerations we have presented here, it seems reasonable to follow the Merrill and Wood prescription in only the situation where it best serves our purposes: the classification $O_1$.

In conclusion, it can be said that the consistency hypothesis and related instructional prescriptions can result in too specific teaching, too narrow a presentation of the content in the case of a productive operation. On the other hand, if we consider a definition as a descriptive operation, and desire performance at the using level, the instruction prescribed conforms nicely to our requirements for the design of the $M_1$ (i.e., to achieve similar response conditions for training and test), but also to try to achieve specific instructional effects with as little added complication as possible. By contrast, the Merrill and Wood prescription for performance at the recall level suggests that

there would need to be experience with examples. This seemed to us to be unnecessarily elaborate, given the Johnson and Stratton data and considering the fact that we learn new words every day via the dictionary, without the advantage of experience with examples and nonexamples. No compelling case can be made regarding the absolute necessity of examples for acquiring the definition of a word.

On the basis of the above considerations, we have come to the following conclusions:

(a) Response conditions during training should be consistent with the response conditions required by the test.

(b) There are no necessary conditions on the structure of the instructional content presented via the $M_i$. Different kinds of content, within reasonable informational limits, could be presented via the instructional methods and reasonably be expected to support criterion performance.

General guideline (a) is consistent with principles of transfer, and guideline (b) violates no known principles of transfer and at the same time permits us to achieve large differences in the $M_i$ (and hence achieve specific effects) without sacrificing achievement of indirect effects.

We have decided, therefore, to maximize the difference in the nature of the stimulus information presented via the $M_i$ and to follow the Johnson and Stratton pairing of the S and R, at least for the first four $O_i$ of interest.

## General Features of the $M_i$ to Be Tested

The core requirement for our research is that we build individual instructional methods that are tightly targeted to particular criterion tests. We spent the previous section defining what must be fixed and what can vary in the design of the $M_i$ to meet the criterion of being tightly targeted (but not too tightly targeted). The Johnson and Stratton strategy of pairing S and R to minimize linguistic recoding during training seems sensible to follow. Minimizing recoding has the effect of maximizing the duration of exposure to instructional content in the form it is to be learned. We believe it possible to demonstrate that under better controlled conditions than those used by Johnson and Stratton, the $M_i$ we design will have unique peaking effects on the related tests.

Following Johnson and Stratton, we will have the following $M_i$ and $CT_i$:

| | $M_i$ | | | $CT_i$ |
|---|---|---|---|---|
| | Method Name | Stimulus Content | Response Task during Training | Response Task on Related Criterion Test |
| 1. | Definitions | A definition | Paraphrasing the definition | Stating the definition |
| 2. | Synonyms | A synonym | Selecting a synonym from a large set | Selecting a synonym from a large set |
| 3. | Classify | Examples | Classification | Classification |
| 4. | Words-in context | A story | Sentence completion | Sentence completion |

Due to theoretical considerations, however, we have introduced some changes into Johnson and Stratton's original design. First, our notion of presenting a definition ($M_1$) is that, through the use of verbal summaries, we want to link new information to words and classes of experience that are already known; we want to do this in a way that illustrates differences in meaning or contextual uses between the new and old information. A "good" definition statement will be one that: (a) uses words and concept classes that are already familiar to the student (a semantic content consideration related to students' prior knowledge), (b) presents the appropriate new information, and (c) is of a structural or syntactic form that is appropriate for the form class being taught. The first consideration, (a), is a matter of good judgment; (b) is a subject matter consideration, but (c) must be discussed.

The Johnson and Stratton structure for noun definitions seems good practice to us. They used genus et differentia definitions where the unique features of meaning are explicitly stated. The definition for altercation illustrates the format: an altercation is a type of social interaction (concept class) characterized by conflict (differentia). The notion that nouns can be organized into higher order classes with special properties has empirical support (see Collins & Quillian, 1969; and Smith, Rips, & Shoben, 1974). However, we are taking broad liberties with the notion that part of a definition contains

reference to a class of which the target word is a member. Designing such definition statements required us to extend the concept classes to which the target words belong; in so doing we realized that the classes of experience to which target words belong often do not have simple lexical labels. Note that an appliance is a type of tool, while a deterrent can be a type of object or event, something that stops action because of fear. The commonality among the objects and events that are deterrents is commonality of function.

Johnson and Stratton report that they applied the genus et differentia strategy to words of other form classes. We found this difficult to do, and consequently, we revised our method for structuring definition statements for our verbs and adjectives. In an informal study where we asked adults to give adjective definitions, we found that they most frequently stated several verbal equivalents with and/or connectives either implied or explicit (e.g., opulent → fancy, rich, lush), not just one single synonym. At the present time, we are completing a study of verb definitions; the study is demonstrating that adult definitions frequently contain either verbs analogous in meaning (e.g., to break → to destroy) or they consist of a more basic word (i.e., structurally simpler, more primitive, perhaps acquired earlier) plus contextual information (e.g., to break → to fall into pieces). They also use synonyms (e.g., to break → to crack). We decided

that the structure of our definitions ought to conform to the structure found when adults give definitions. In limited empirical tests, we have found that noun definitions can be accounted for in terms of concept classes plus distinctive features; verbs by the three strategies noted, and adjectives by verbal equivalents plus reference to the domain of things that are described. The major change we made from Johnson and Stratton here was to allow ourselves to use other adjectives and other verbs in writing definitions. (It will be recalled that Johnson and Stratton did not permit synonyms in definitions.) The $M_i$ for the definition stating criterion task ($M_1$) will be comprised of an S (a definition written according to the above specifications) and an R (paraphrasing the definition presented). Additional details of this $M_i$ design are presented in a later section.

A second major change that we have made to the Johnson and Stratton procedures has been to generalize the synonym method ($M_2$). This revision was done to incorporate more recent conceptions of associative learning into the $M_i$. Recent associative learning conceptions suggest that semantic memory consists of networks of semantic relationships, only one of which is synonomy; that is, the semantic relation of means or is the same as (e.g., alacrity means eagerness) is only one of a number of relationships that could be established. In the semantic network models, there is no reason to believe that synonomy is the most prepotent aid to memory

of meaning. In planning the content of the information to be practiced in a "synonyms method", one must decide what content should be presented (the scope of the relationships presented) and how often each individual relation should be practiced (the frequency). We have decided that a wider scope than synonomy (and fewer trials on each individual relation) is preferable to a narrow one (with more trials on each). This decision was based both on the finding that variable encodings support memory and our own hypothesis that frequently changing the nature of the semantic relationship accessed would force deeper semantic contact with the target word that is presented. Further, the changes made to the definitions method also forced changes in the synonyms method.

Since we had redesigned the content to be presented in this method, we also reconsidered the design of the response task. We felt that a response task that required the student to generate the relationship between the target word and a related concept (e.g., How are flattering words related to a sycophant?) and then test it by comparing the generated relationship to the set that is then displayed, made deeper semantic contact than a response task that required selection of the appropriately related concept when the target concept and particular relation are specified (e.g., Pick the phrase that describes a tool of a sycophant: 1. brown noser 2. much dislike 3. college classroom 4. flattering words).

The change in the response task, of course, means a change in the $O_i$ statement. $O_2$ of the list presented previously now becomes:

> 2. The student can state or explain the logical or semantic relationship that exists between the target word and a related concept.

This $O_i$ is consistent with current ideas about the meaning of concepts: concepts take on meaning through their relationships with other concepts (see Rumelhart, Lindsay, & Norman, 1972). Hence, if the student can state the relationships between the target word and another concept, he "knows" the meaning of the word. The goodness of this conception of meaning content will be revealed through our empirical assessment of the direct and indirect effects of an $M_i$ designed on the basis of it. The name we will use to refer to this $M_i$ will be Word Relations.

Of great importance to the design of the Word Relations $M_i$ is the selection of relations to be taught. These relations will, of course, vary as a function of the form class of the word. There are network theories of semantic memory that can be relied upon as sources for types of relations (e.g., Rumelhart, Lindsay, & Norman, 1972). Kintsch (1972) has suggested that the meaning of a word is comprised of the lexical knowledge needed to support generation and judgment of the acceptability of sentences, a definition that highlights case grammar as a source of relations. From an inclusive list of <u>potentially</u> instructable relations, <u>really</u>

instructable ones (those that can be presented in a comprehensible way to the student) can then be selected and taught.

Concerning the third criterion test and teaching method listed at the beginning of this section, the $M_i$ targeted toward the classification $O_i$ is based on the design prescriptions of Merrill and Tennyson (1977). There are differences between our $M_i$ and the Johnson and Stratton classification method: we present a definition and we also systematically vary the arbitrary elements of the definition to generate examples and nonexamples. For example, if a deterrent is something that stops action because of fear, then the variables are the things that are used to stop the action and the nature of what it was that was stopped. In the language of concept teaching technology, what we have done is to identify the critical attributes and then vary them (see later section on the Classify Module for a more comprehensive description).

In the design of the words-in-context $M_i$, the fourth method listed at the beginning of this section, we have decided that important aspects of the meaning content presented correspond to macrosemantic relations in the text. For example, if a student <u>vociferates</u> at her parents because they will not pay her tuition for college, an aspect of meaning content that might be acquired is that one vociferates only when ones wishes to "get back" at another for a perceived injustice. To heighten the chances that this unique aspect of content gets learned (in accordance with our goal to

maximize differences in information presented and acquired), it currently seems necessary to require use of the word "because" in the sentence completion response task that is paired with this context presentation.

In addition to the four Johnson and Stratton $O_i$, our list of terminal objectives included three additional objectives: constructing new examples, using the word spontaneously in writing and/or speech, and constructing novel subject matter structures (diagrams and continua). We designed three addition $M_i$ to teach toward each of these; however, due to time and womanpower constraints, it will be necessary to delete the $M_i$ entitled Conversations which required that the target word be used in substitutions for various paraphrases found in high fidelity examples of real conversations. We will, however, retain the $M_i$ targeted toward example construction. This $M_i$ is called Create and is a discovery version of Classify, the thought being that a task requirement to induce a concept from examples might aid in the development and memory for the critical attributes (the contrast with Classify becomes clearer when the specifics of the two strategies are compared). Another $M_i$ we have designed is called Word Line and it teaches the target word in the context of other words in such a way that the context words plus the target word can be arranged along a perceptual, affective, or cognitive continuum. The roots of the content structures employed

in this $M_i$ are found in Miller's (1972) and Miller and Johnson-Laird's (1976) notions of semantic domain and in Klein and Saltz' (1976) conceptions that there are semantic dimensions along which words can be arranged or rated.

From the above descriptions of the content structures presented via the various $M_i$, the design strategy we are using to maximize differences in content between the $M_i$ should be clear. It is of interest, however, to raise a serious question regarding the need to use more complex information structures (e.g., definitions with instances, multiple semantic relations) when simpler types of content might do. To evaluate the necessity for complex structures over simpler ones, our seventh $M_i$ will be a form of comparisons instructional procedure in which a simple logic of repeated practice will be used and the content taught will be a simple phrase equivalent in meaning to the target word. The items from the pretest (see later section on pretest construction) will be used as the instructional items.

In summary, then, the $M_i$ and $CT_i$ that we will use to test our design framework, will be the following:

| | $M_i$ | | | $CT_i$ |
|---|---|---|---|---|
| | Method Name | Stimulus Content | Response Task During Training | Response Task on Related Criterion Test |
| 1. | Definitions | A definition | Paraphrasing the definition | Stating the definition |
| 2. | Word Relations | Target word and Related word | Identify the relationship | Identify the relationship |

| | $M_i$ | | | $CT_i$ |
|---|---|---|---|---|
| | Method Name | Stimulus Content | Response Task During Training | Response Task on Related Criterion Test |
| 3. | Classify | Examples | Classification | Classification |
| 4. | Words-in Context | A story | Sentence completion | Sentence completion |
| 5. | Create | Parts of examples | Generating examples | Generating examples |
| 6. | Word Line | Target word and Related words | Arrange words along continuum | Arrange words along continuum |
| 7. | Equivalents | Pretest item | Selection of correct pretest response | Selection of correct pretest response |

Taken together, the seven different $M_i$ comprise a set of instructional methods that vary on significant task dimensions - both the structure of the content presented and the nature of the response task. Our hope is that we have designed in sufficient variation in instructional conditions to significantly influence the structure of knowledge acquired from the various $M_i$ (and hence impact the levels of performance on the $CT_i$). Hopefully, this knowledge variation will permit the achievement of specific instructional effects, yet not disallow indirect effects. To further insure that differences in the content presented make real differences in the structure of the content acquired, it will be necessary to insure attention to the different elements of content; hence we will specify a response requirement for each important

element of content (e.g., in the definitions method, this means response tasks must be incorporated so that each critical attribute is responded to).

There are other design guidelines that have been applied to each of the $M_i$ - the terminal response task in the $M_i$ must be the same as the response task on the related criterion test. The feedback used in each $M_i$ (except the simple comparison $M_i$) must adequately display the reasoning, or associative connections, that support correct performance at any response request point. As a general guideline, an attempt will be made to design the $M_i$ logic so as to provide approximately three minutes of instruction for each word taught. This time guideline allows us to estimate length of instruction and degree of task variety approximately allowable across vastly different methods of teaching. The three minute approximate "limit" seems sufficient to produce a respectable level of learning (see Johnson and Stratton) for a word but not to excessively overteach so that methods effects disappear.

Each of the seven methods, Definitions, Word Relations, Classify, Words-in-Context, Create, Word Line, and Equivalents, will be used to teach students words whose meanings they do not know. Following instruction, students will take criterion tests and "experimental" tests designed to uncover what it was that was learned during the teaching episode.

## Design of the Computer Assisted Experimentation System

In pursuit of testing our decision theory framework and attempting to determine whether specific methods effects can be obtained in the area of vocabulary concept learning, we have expanded the scope of the problem beyond the initial Johnson and Stratton work. We have more outcomes, more criterion tests, more instructional methods, and lots of courseware to write. We expanded the scope of the research for the variety of reasons mentioned previously.

To review them briefly, more outcomes (seven) are represented because there are simpler ones of interest as well as more complex ones in the full range of a taxonomy of learning outcomes. Since achievement of these outcomes must be assessed, and we have added "experimental" measures to the list of criterion tests, we therefore have more tests. Since six of the outcomes are taught toward by a given method, we therefore have six methods targeted toward the criterion tests and one comparison method. The decision to test more than four methods was not based solely on our desire to include more outcomes, however. In the exploratory stages of design, as we read both design theories and theories about the psychological structure of word meaning, it became clear that more than the Johnson and Stratton four could be explored to advantage, extending the scope of the test of the decision theory approach. Then, as we translated ideas about content structure

into instructional tasks, it became clear that significant differences in instructional strategy could be retained across about seven methods (and we are using six).[8]

We are expecting, with computer instrumentation, that there will be both specific methods effects and hopefully sizeable indirect effects that will be dependent upon the $M_i$ and the particular $CT_j$ in question. The goal we have elected to try to achieve is the attainment of a decent level of performance on all the $CT_i$ using a subset (to be identified) on the $M_i$. The first stage of research requires calibrating the size of the direct and indirect effects; later stages require studying how the combined use of the $M_i$ confirms or disconfirms the decision theory equations (See Block, Note 1 for additional detail regarding this question.) The strategy of research requires that the experimentation system be designed to permit the use of the $M_i$

---

[8]The problem of translating various theories of meaning content structure into instructional tasks is an interesting one. Several times in our work, we started with a given theory i different from theory j and ended up with an instructional strategy identical to the strategy developed from theory j (or with a situation where the two strategies were different in nonessential aspects). As an example, it has been suggested that the meaning of a word is a cognitive procedure for a word's use. Therefore, to teach it one would present the student with descriptions of a situation, help him or her extract the critical aspects of the situation that are the conditions of use, present more situations, and provide guided practice in aspect extraction, then have him or her label and learn the labels for the aspects. Much of this strategy is similar to Classify, except in Classify the student is learning a concept - a classification rule based on critical attributes!

in various combinations and/or all the $M_1$ at once (the sequence of maximum cost and the upper boundary of instructional effectiveness).

Other differences in procedure that are important include the use of an on-line multiple choice pretest which allows for automatic selection of words for individual students on the basis of their pretest responses. Thus, students learn words we are certain (within limits of the inference permitted by our test) that they do not know, in contrast to the less certain use of norms. Different students will, of course, be learning different words. We are also writing instruction for a much larger corpus of words, both to increase the power of our findings, and to enable us to run in a real world setting. This setting is a College Learning Skills Center for students who need work in college reading and thereby seek improvement in vocabulary. We have also included in our instructional system an interim test option with on-line return to teaching routines for words that were missed on the interim test.

In general, the plan for one week of vocabulary instruction is that on the first day, students sign on the system, take a pretest, and automatically receive instruction on words that were missed. They return after one day's absence and take an interim posttest with automatic reteaching. Following another day's absence, they return for a final session taking the battery of tests. In addition to implementation of the instruction in the learning skills

setting, laboratory experimentation is also planned, for the simple reason that specific methods effects might not appear if the interval between instruction and test is too short or if there is an interim test with reteaching.

We also plan to add a standardized verbal ability, or a vocabulary test, to the experimental design in order to determine whether there is an interaction of ability with method reflected in the relative sizes of direct and indirect effects. We also have added to the variety of experiments we can conduct by parameterizing the $M_i$ design (see description of individual $M_i$ in following sections).

Procedures for Word Selection

One of the early decisions made during the time spanned by this report was the determination of the word pool from which words to be selected would be taught. The word pool contains "general" vocabulary words; i.e., words that appear in popularly disseminated texts, not words whose use is endemic to any one professional discipline or specialized field of knowledge.

Since there are no norms on vocabulary words known (or not known) by college students, the next reasonable substitute is word frequency, on the assumption that low frequency means words not likely to be known. This assumption was adopted by Johnson and Stratton. That 1966 study suggested to us the optimal degree of difficulty for the words to be used in the instructional modules.

This was confirmed in a pilot study conducted by our staff during the summer of 1977. From their preliminary study, Johnson and Stratton reported that words with Thorndike-Lorge (1944) frequency counts of between 11 and 14 (i.e., words occurring between 11 and 14 times per the Thorndike-Lorge sample of one million words) could be expected to be easy for college students; words with counts between 1 and 6 could be expected to be difficult. Rather than accept the Johnson and Stratton preliminary findings prima facie, and having already had experience with children's norms that are out-of-date, we conducted a pilot study of our own where college student subjects were presented with words of varying frequencies (including the Johnson and Stratton words) and were asked to perform various tasks with the words such as providing definitions, supplying synonyms, or using the words in sentences. The results of this pilot study basically corroborated the Johnson and Stratton findings concerning word difficulty for a college student population.

The college student population in our experiments, however, is likely to be divided into at least two distinctly different types of participants. Volunteers at the Learning Skills Center are usually attracted because they either need remedial skills development or because they are studying for graduate school entrance examinations. Thus, there will be some subjects who need to learn relatively basic words and other subjects who will be interested in learning relatively esoteric words. To anticipate the needs of the different types of subjects, we have compiled the word pool

in such a way that it consists half of "easy" words and half of "hard" words. In this manner, we can teach words that will likely not already be known by the subjects and yet are still within their realm of interest. Remember, each subject takes a pretest and receives instruction only on those words he or she does not already know.

The easy words. As reported in Johnson and Stratton and from our own pilot study, we knew that an "easy" word would be one which has a higher frequency count listing than a "hard" word. We used the Kučera-Francis (1967) frequency ranking of words as our source for "easy" and "hard" word rankings since it is more recent than the Thorndike-Lorge listing used by Johnson and Stratton. The Kučera-Francis rankings were obtained from a corpus of slightly more than one million words taken from 500 different samples of continuous discourse that were printed in the U.S.A. in 1961. Their samples were taken from 15 genre categories, such as press reportage, religion, skills and hobbies, and general fiction. The Kučera-Francis rankings contain not only the frequency of occurrence of the word within the total word corpus, but also the number of genre subdivisions in which the word occurs and the number of samples in which the word is found. These last two rankings seemed to us to indicate what we have come to call the "power" of a word; that is, an estimation of its ability to be used in a variety of different textual materials. We would like our word pool to contain as "powerful" words as possible.

We started with words of the highest frequencies in the Kučera-Francis listings and worked down until we came to the place where intuitively we believed we were beginning to spot words that would not likely be known by some subjects. Those words had a frequency of 10. We extracted all words with frequency rankings of 10 which were also contained in at least three different genre subdivisions (the "power" criterion); we excluded only proper nouns from this list. We followed the same process for words ranked 9, 8, and 7. Seventy-four "easy" words were determined in this manner; 20 nouns, 20 verbs, and 34 adjectives. Examples of the "easy" nouns are: consensus, deterrent, facade, hypocrisy, and vengeance; "easy" verbs are: allege, conform, entail, mandate, and speculate. Some "easy" adjectives are: callous, expedient, ingenious, naive, and potent.

The hard words. The "hard" words, which all have Kučera-Francis frequency rankings of 1-3 came from various sources: staff members were asked to generate lists of words that they thought were good candidates for instruction; words were culled from study guides for graduate school entrance examinations; words were extracted from vocabulary building quizzes that appear in newspapers, magazines, and books; the four words used in the final Johnson and Stratton study were added; and finally, words from our pilot study conducted during the summer of 1977 that had not been known by the subjects were added. The resulting list of "hard" words was

Considerably longer than the corresponding list of "easy" words. Therefore, words were randomly selected from this longer list until a final list of 74 "hard" words had been selected; it contains 20 nouns, 20 verbs, and 34 adjectives, the same part-of-speech breakdown as in the "easy" word list. Sample "hard" nouns are: antithesis, dilettante, hegemony, rancor, and sycophant; "hard" verbs are: assuage, flout, malign, pontificate, and venerate. Examples of "hard" adjectives are: apocryphal, ebullient, germane, indigenous, and salubrious.

Content contained in the instructional modules is being developed from this master list of 148 words. We do not expect all 148 words to be used in all modules; some words will not doubt fall by the wayside as they prove unsuitable for application in various instructional modules.

Materials Design: Pretest Construction

Considerations for the design of the pretest were that it (a) be in a format straightforward and familiar to students, (b) be easy to code, (c) not require much time for item construction, (d) keep testing time to a minimum, and (e) avoid similarity with the criterion tests so there would be no selective focusing effect.

As a process of narrowing down the number of possible formats, considering criterion (e) first, we knew that we should avoid definition construction, classification, construction of examples, word relationships, and sentence completion as pretest formats because they all would definitely be used as criterion measures. We had also decided, however, that whatever format we adopted as a pretest, that we would readminister the same format as a posttest using an intervening teaching sequence that teaches directly to the correct response for each target word. This would give us an assessment of the effectiveness of an unsophisticated, "quick-and-dirty" teaching method in comparison with our other more carefully designed sequences.

Considering criterion (d), keeping testing time to a minimum, implied constructing a pretest with short items and providing the responses from which students would select the correct answer. This type of format would also fulfill criterion (b) in that it would be easy to code as students selected each response from a multiple choice array. Next, considering criterion (a), that

the pretest be in a straightforward, familiar format, we realized from the remaining possibilities that selecting a word or phrase that is the best synonym or antonym of each target word would likely be familiar to students since many widely-used standarized tests of vocabulary employ these formats.

Finally, considering criterion (c), the ease of item construction, we realized that it would be more time consuming to construct an antonym pretest than a synonym pretest since it seemed that many of the words in our database do not have immediately recognizable antonyms. Using this process of elimination, a multiple choice format requiring students to select synonym phrases or words for each concept seemed to best fulfill the above mentioned criteria.

In constructing the test, we *simply followed* standard procedures in item construction, with our own specifications used for distractor construction. If the test is a good test, then the category of student response, correct or incorrect (C or I), should match each student's state of knowledge about the word - either she/he <u>really</u> <u>knows</u> the meaning of the word (RK) or she/he does not ($\overline{RK}$).

In the case in which the student really does know the meaning of the word, we want the test to be designed so as to maximize the chances that the correct response will be given, $P(C|RK)$. If the student is in state RK, then the student will not be guessing, and she/he will be testing the possible responses for their fit

to some previously-stored information about the word's meaning. In order for the correct (i.e., the "keyed" correct) answer to get selected (in preference to the distractors), it must represent the best match to stored information about the word.

Judgments regarding the fit of the possible responses to stored information can be made on an absolute basis or they can be judgments of relative goodness of fit, based on the range of deviations represented by the distractors. Regardless, in order to increase the chances that the correct answer is selected when the student is in RK, there must be one clear best fit and no distractors that would be judged to be semantically equivalent to the keyed one. There must be an acceptable semantic distance, represented among the distractors. We attempted to construct items meeting an acceptable semantic distance criterion, although one might want to perform additional empirical tests of how well we met it.

If students are in state $\overline{RK}$, they could randomly guess, or they could follow some strategy for selecting an answer, or they could respond on an associative basis. Whether they randomly guess, or not, will likely depend upon the characteristics of the word (e.g., whether it is comprised of smaller units of meaning and is amenable to morphological strategies) and upon their familiarity (prior experience) with the word. At any rate, our task in item construction is to minimize $P(C|\overline{RK})$ and maximize $P(I|\overline{RK})$. When

the student is randomly guessing, we attempted to minimize $P(C|\overline{RK})$ by using four distractors rather than the usual three.

When the student is not randomly guessing, the extent to which we have maximized $P(I|\overline{RK})$ depends upon whether our assumptions about how students select word meaning when they do not know the meaning, are actually true. Our assumptions were that, for students with no prior experience with a word, the students would follow one of three strategies for selecting an answer. The positing of these strategies reflects an assumption that verbal units (words) are encoded multidimensionally and that students are aware of this fact and use their knowledge to formulate strategies for selecting answers. The first strategy students might use would involve decomposition of a word into morphological units. Then a distractor whose meaning overlapped with the parts would be selected (e.g., living in the past is one of the distractors for the database word concurrent. By decomposing concurrent into the con - against and current - referring to the present time meaning subdivisions, we believe that this distractor might conceivably be chosen. We have evidence from our pilot study conducted during the summer of 1977 that students do perform this type of decomposition operation.)

The second strategy students with no prior experience would follow is one based on acoustic or graphic features. Students would choose a distractor on the basis that its meaning represented a word that contained phonological or orthographic similarities to the target word (e.g., for the database word dual, one of the

distractors is using shooting pistols, obviously related to the word duel; dual and duel are phonologically identical). A third strategy would involve choosing a distractor on the basis of chunks contained within the word to which meaning can be assigned (e.g., for the database word compulsive, students may focus on the puls letters of the word, make a puls - pulse association, and select the distractor beating regularly).

For students with some prior experience with a word, but who are still in state $\overline{RK}$, these students would likely choose a distractor on an associative basis. One basis might be affective reactions (e.g., students with prior experience may have negative associations with the word revulsion and might select the distractor an uncontrolled fit for that reason). The second associative basis might be reflected in the choice of a misconception (e.g., for the database word compulsory, the distractor difficult might be selected since compulsory activities, such as compulsory final exams, etc., are often viewed as difficult). A third associative basis might be the choice of a distractor on the basis of other contextual associations (e.g., for the database word fidelity, we speculate that the distractor banking policy might be selected since the word fidelity often appears in the names of banks).

Our detection of state $\overline{RK}$ is in error to the extent that the correct response can be selected for the wrong reasons. To avoid the chances that an orthographic strategy could result in a correct selection, for database words where the correct synonym resembles the target word orthographically, we included as distractors

other possible responses that resemble the target word (e.g., for the database word monastic, the correct response is living like a monk, which has some orthographic similarity to the database word. For this reason, we included as distractors craving money, behaving like a monkey, and happening on Monday, because they also orthographically resemble the database word, but retain acceptable semantic distance.)

In addition to scoring responses as correct and incorrect, we have elected to take another measure of student response during the pretest. As noted previously, a given alternative can be selected for a variety of reasons. Assuming that the category of a response, C or I, is a valid index of knowledge state, responses that are incorrect and fast are probably due to guessing; responses that are incorrect, yet slow, reflect a strategy or partial prior knowledge (and are separable in part due to the method for generating distractors). Responses that are correct and fast reflect a well designed item whose distractors enable quick rejection and/or whose correct alternative is a good match to the meaning store. Responses that are correct, yet slow, may mean a poorly designed item. We would, of course, have tried to use confidence ratings to get a measure of prior learning or response strategy, but such measures are obtrusive (making the criterion for judgments unstable because of interruptions); they are tedious for the students (a possible source of invalidity); and they add to the time required for the test.

Of interest in our research will be the extent to which meanings for words that we provide during instruction are remembered on the posttest to a greater or lesser degree depending upon initial response speed and, of course, category of distractor selected. We are therefore assuming that quick incorrects mean guessing which in turn means no prior knowledge of the word. A hypothesis proposed by Underwood and Postman (1960) suggests that a deliberative attempt to figure out or reconstruct word meaning (prior to receiving instruction in the word), which in turn leads to the generation of some inacceptable response (shown by slow incorrects on our pretest), may have negative influence on what students remember from instruction. Underwood and Postman point out that in order to learn the new associations, the student probably has to unlearn the prior associations. They speculate that during a rest interval, spontaneous recovery of the unlearned prior associations will decrease the probability of recalling the associations taught during instruction. A similar hypothesis about prior associations and amount of recall can be evaluated in our research.

## Components of the $M_1$

At this point, it is necessary to establish some terminology to enable more precise description of the instructional sequences we have designed. We change from the term instructional method,

which in the Johnson and Stratton sense meant one of four particular types of pairings of stimulus form and type of response. We introduce the term module which is a higher level descriptor for a sequence of more elementary instructional events, the most important of which are instructional routines that serve different functions within a module. These routines can be response contingent or not response contingent. They are designed to facilitate various stages of cognitive learning and also to facilitate perceptual or motor learning in some cases. A module is a higher level controlling program that governs the entry into particular routines and selects the value of the instructional parameters to be implemented in that routine. Modules are what we have been referring to as $M_i$. Hence, at a general level of description the task dimensions that differentiate the $M_i$ are the type of content structure[9] taught and the nature of the terminal response task.

The term, instructional routine, refers to a sequence of instructional events consisting of the presentation of a stimulus, a response request, a response evaluation performed by the computer (yes, no, incorrect, etc.), and an automatic response correction message, in the case of incorrect responses. Modules also contain

---

[9] There is a need to clearly differentiate the term content structure, which is the instructional developer's structuring of instructional content, from the term knowledge structure, which is a theoretical description of the structuring of information in memory. A content structure is an interpretation and extrapolation of a knowledge structure.

other routines that are not response contingent such as directions to pronounce a word, type it from memory, or compare a constructed response to a model, for which there is no external check on response adequacy. Many of these events directed more to perceptual or motor rather than cognitive learning are included to provide for response familiarization.

Our $M_i$ descriptions consist of brief descriptions of the events of instruction as sequences of instructional routines. To date, only two modules, CLASSIFY and DEFINE have been designed. Work is underway on the other $M_i$. The next two sections of this paper present brief descriptions of the CLASSIFY and DEFINE $M_i$. When the design of all $M_i$ is completed, the designs will be thoroughly documented using detailed flow chart descriptions accompanied by the complete scripts of the instructional database.

## The CLASSIFY Module: Rationale and Design

The rationale for the CLASSIFY Module is based rather closely on the design guidelines specified by Merrill and Tennyson (1977) for teaching concepts. Their guidelines are probably the most detailed specification of a technology of concept teaching extant in current instructional design literature. The Merrill and Tennyson guidelines have evolved from their considerable research on strategies for teaching "classroom like" concepts; Merrill and Tennyson note, in fact, that each of the design prescriptions have been tested and validated in carefully controlled

experimental research studies. (See Merrill & Tennyson, 1977, for an elaboration of their relevant research on concept instruction.) Thus, of our seven modules, each containing different instructional strategies, CLASSIFY is the one module whose development required the least number of design decisions. This is because Merrill and others have targeted their instructional research on the decisions designers face and they have synthesized the findings of that research into instructional design guidelines for concept teaching. Hence, CLASSIFY is the one module that adheres most strictly to the direct application of research findings that reflect the current state of the art in concept teaching.

Merrill and Tennyson's prescriptions for concept teaching facilitate the development of classification behavior in the student; that is, when given the general name of the class (i.e., the target word from the instructional database) and shown representations of specific instances of this and other classes (i.e., situations representing examples and nonexamples of the class), the student will be able to identify those situations which are members of the class and those which are not.

Four major instructional events, Rule, Example, Practice, and Test comprise the main facets of an instructional episode.

Simply stated, the Merrill and Tennyson "recipe" for teaching concepts is as follows:

RULE

a) Present a definition; i.e., a statement identifying each critical attribute of the concept and how the critical attributes are combined.

EXAMPLE

a) Provide an expository presentation of a set of matched example/nonexample pairs in which subsequent examples are divergent from the preceding examples and which range in difficulty from easy to hard.

b) Provide attribute isolation help for each example and nonexample in the expository presentation.

PRACTICE

a) Provide an inquisitory practice presentation of newly-encountered examples and nonexamples arranged in random sequence.

b) Provide feedback after the response that includes the correct answer and attribute isolation help that focuses attention on the critical attributes.

TEST

a) Test correct classification by means of an inquisitory test presentation consisting of a

> sufficient number and variety of randomly-sequenced, newly-encountered instances to allow reliable and valid inference about subsequent classification behavior of yet-to-be encountered instances of the concept being taught.

The ways in which we have adhered to and deviated from this "recipe" in the design of our CLASSIFY Module will become apparent in the CLASSIFY Routine Composition and Sequence section that follows shortly. In sum, our reasons for selecting a Merrill and Tennyson classification type module are: 1) theirs is currently the most detailed statement of a concept teaching technology; and 2) as a subpoint, Merrill and Tennyson are very much proselytizers of their concept teaching approach and insist that instruction for many $O_i$ requires classification. Merrill and Tennyson believe that they have indeed specified an efficient and effective means of teaching concepts; however, a variety of other teaching methods can also be compared and contrasted to theirs. Most of their research and related development work, in fact, has been performed using technical concepts as exemplars. By following their guidelines, we will attempt to ascertain whether they will apply equally well to the less highly structured "general" vocabulary concepts in our instructional database.

## CLASSIFY Routine Composition and Sequence

The CLASSIFY Module comprises five different exercise routines. Here we will describe the tasks in order by routine. At the end

of this section, we will comment on how closely the five routines match the Merrill and Tennyson instructional guidelines for teaching concepts.

Routine 1: Pronounce. This routine identifies the target word and provides a phonetic pronunciation key, with the direction to the student to pronounce the word.

Routine 2: Definition. A definition of the target word is then presented, emphasizing its critical attributes by separating them visually from the rest of the display. This was done to facilitate encoding; research by Markle and Tiemann (Markle, 1975) has shown that simple spatial separation of critical attributes enhances the learning of them. The student is then instructed to read the definition.

A sample CRT screen display for this routine using the word consensus would be:

```
Read the definition:

  When two or more people meet to make a decision, the people can
agree or disagree about what they believe or what they want to
do.  When most of the people agree on what they believe or what
they want to do, this event is called -- a consensus.

  Thus, a consensus is a type of situation
                       characterized by agreement in opinion or belief
                       among the members of a group.

Press return when you are finished reading >
```

The key elements of content in the CLASSIFY module are the critical attributes of the definition. The routines using

examples and nonexamples require specific attention to these elements.

Routine 3: Matched examples and nonexamples. A situational example of the target word is presented, labeled as an EXAMPLE. A yes-no question, focusing on one of the target word's critical attributes, is presented along with the example. The student is asked to respond to the question by typing a Y for a yes answer or an N for a no answer. Feedback after the response alerts the student to the correct response and provides attribute isolation help. (That is, the feedback assists the student in encoding the example in terms of the critical attribute.) A matched situational nonexample of the target word is then presented, labeled as NOT AN EXAMPLE. The matched nonexample is visually paired on the screen with the example. The same yes-no question as above that focuses on one of the target word's critical attributes is presented along with the nonexample. The student is asked to respond in the same way as above. Once again, feedback after the response alerts the student to the correct response and provides attribute isolation help. This matched example/nonexample pairing with yes-no questions on a critical attribute is repeated for the number of critical attributes contained in the target word.

Routine 4: Forced choice of example. Next, a forced choice exercise is presented to the student; i.e., a matched example/nonexample pair of situations is presented and the student is instructed that only one is an example of the target word. The

student is asked to identify the number of the instance that is the example. Feedback after the response alerts the student to the correct response and identifies the specific wording in the example that "gives it away" as the example. The number of times that this routine is repeated is also dependent upon the number of critical attributes in the target word.

Routine 5: Classification. In the final routine of Module CLASSIFY, the student is presented with one instance at a time with the direction to identify the instance as an example or not as an example of the target word. This routine is repeated a number of times, depending upon the difficulty of the word being taught. Feedback for a correct response confirms the response as correct. Feedback for an incorrect response constructs a matched example/nonexample pair out of the instance being presented and labels for the student which instance of the pair is the EXAMPLE and which is NOT THE EXAMPLE.

The criterion task for this module is the classification task specified by Merrill and Tennyson, i.e., when given the target word and shown representations of specific instances representing examples and nonexamples of the class, the student will identify those situations which are members of the class and those which are not. The criterion task is the same as the exercise presented in Routine 5, using different instances and eliminating the feedback.

One other important point to note here is that the CLASSIFY Module is being designed in such a way as to allow for variability

along what we believe to be important instructional parameters. For instance, the module has been specified so that the definition of the target word can be either present or absent when exercises in the subsequent routines are being performed by the student. This will hopefully enable us to identify whether we have indeed isolated some important instructional parameters, a situation we can determine by examining how level of achievement varies with particular parameter sets.

In comparing our CLASSIFY module with the Merrill and Tennyson guidelines for teaching concepts, it should be apparent that we have followed their specifications rather closely. Our deviations from their "recipe" are as follows: 1) Our Routine 1, Pronunciation is presented as an addition to Merrill and Tennyson's suggestions for teaching concepts. We believed here that facility with the oral production of a word will help to develop and reinforce memory for the word itself. 2) Our Routine 2 is identical to the RULE presentation suggested by Merrill and Tennyson. 3) Our Routines 3 and 4 are modified versions of the Merrill and Tennyson EXAMPLE presentation. Merrill and Tennyson suggest that the EXAMPLE presentation should be done as exposition. We decided here that we wanted to apply a particular principle of instructional programming; i.e., to use active responding whenever possible, to keep the student attending to our instructional display. We therefore used an inquisitive rather than an expository presentation here. Routines 3 and 4 seemed to us to shape the terminal

behavior (i.e., classification) in a logical manner. 4) Our Routine 5 is identical to the PRACTICE guidelines laid out by Merrill and Tennyson. As mentioned previously, our criterion task is also the one that they specified as their test.

The DEFINE Module: Rationale and Design

The content structure taught in the DEFINE $M_1$ is a definition and the terminal response task is being able to construct a paraphrase of the definition. In designing this $M_1$, we worked with more elementary instructional events than did Johnson and Stratton in their definitions method. We used our more fine-grained control over student learning processes to segment the events in learning and make it a serial process. We designed the task conditions to focus attention on the key elements of content, the attributes of a definition; to control the number of rehearsals or encounters with each critical attribute; and to institute a form of retrieval practice (in which students must construct a definition from memory). The conceptual framework that rationalizes what we have done is the Frase (1975) model of prose processing in which there is an attempt to relate instructional programming characteristics to various prose learning processes.

DEFINE Routine Composition and Sequence

The DEFINE module comprises seven different exercise routines. As we describe the tasks in order by routine, we will also note

where the routines are similar to or begin to differ from those that make up the CLASSIFY module. The reader will see the opportunities for student rehearsal, relation, and retrieval of the definition of the target word as the routines are described.

Routine 1: Pronounce. This routine identifies the target word and provides a phonetic pronunciation key, with the direction to the student to pronounce the word. This routine is identical to the first routine in the CLASSIFY module.

Routine 2: Definition. A definition of the target word is then presented, emphasizing its critical attributes by separating them visually from the rest of the display. For noun definitions, we emulated the type of definition provided by Johnson and Stratton in that our definitions also characterize each word in a specific way and place each in a higher order class. Definitions for verbs and adjectives permit use of other verbs and adjectives. We also supplement definitions for verbs by describing conditions of use and for adjectives by describing the domain of application. Following the presentation, the student is instructed to read the definition. This routine is identical to the second routine in the CLASSIFY module; i.e., in both DEFINE and CLASSIFY the student is presented with this same type of definition in the second routine.

Routine 3: Type word from memory. In this routine, with the target word as the stimulus, the student is instructed to look carefully at the word until she/he is ready to type it from memory.

As a response, the student types the word after the stimulus has been removed. Feedback here either confirms a correct response or presents the word again, which is then retyped by the student, with the word on view.

This routine is not specifically contained in CLASSIFY because its task is actually subsumed in another of the CLASSIFY routines; as the student performs classification exercises in that module, she/he types the target word from memory at that time.

In the remaining routines of module DEFINE, there is total deviation from CLASSIFY as the routines turn to shaping the specific criterion behavior associated with this module.

Routine 4: Definition with attribute labeling. The definition is presented again in this routine, but this time making the critical attributes even more apparent. First, the form class (part of speech) of the target word is identified with a general statement alerting the student to the way a definition of a word of this form class should look. Then the target word's definition is presented with its critical attributes labeled.

A sample CRT screen display for this routine using the word deterrent would be:

```
Deterrent is a Noun.
   A Noun definition names a general class of things plus
   some special qualities.
   A deterrent is something (what it is)
                  that makes you stop what you are doing (Special Quality 1)
                  because you fear the results (Special Quality 2)
   Press return to answer questions>
```

Routine 5: Multiple choice paraphrase of each critical attribute of definition. Next, a multiple choice exercise is presented to the student. Part of a definition paraphrased from the original presentation appears on the screen; it is incomplete because one of the critical attributes appears as a blank. In order to complete the definition, the student selects from three possible choices; the correct selection is a paraphrase of the critical attribute missing from the partial definition. Feedback for a correct response confirms the response and retypes the complete paraphrased definition. Feedback for an incorrect response tells the student to try again to pick the phrase that says the correct idea in different words. This routine is repeated for the number of critical attributes contained in the target word.

An interesting point here is the design of the module in such a way that the definition presented in Routine 4 can be either present or absent while the multiple choice paraphrasing occurs in Routine 5. This allows for variability in what we believe to be important instructional parameters. By designing the modules with variability, we can take a single module and isolate parameters that theoretically might produce significant differences in achievement as we vary those conditions. This enables us to do research within modules as well as comparisons across modules.

Routine 6: Multiple choice paraphrase of all critical attributes of definition. Similar to Routine 5, this routine presents a situation where the student selects a correct paraphrase of the definition's critical attributes. In this routine, however, all critical attributes for the definition are selected within one exercise, not individually as in Routine 5. This routine, therefore, begins to approximate definition construction. The student response and feedback are similar to those in Routine 5.

Routine 7: Definition construction. In the final routine of the module, the student is instructed to type in a definition of the target word. She/he is cued to use the target word in the definition and to phrase the definition in a manner to make it read appropriately for a word of its form class. After making the response, as feedback the student compares his or her definition with the original presentation of the definition.

The criterion test for this module is definition construction. It is identical to the exercise in Routine 7, eliminating the feedback.

Test Development

Tests will be developed to measure performance on each $O_i$ and will be administered in such a way as to assess both direct and indirect effects. As noted previously, we will also use "experimental" tests, in addition to the criterion tests, in order to determine aspects of the structure of information acquired during learning.

Although we have not yet finalized the formats for all of the criterion tests, we have been pondering their design decisions. Having spent a great deal of time in attempting to piece together the Johnson and Stratton teaching strategies and related testing items, we are profoundly aware that such tests of performance can be designed either to be extremely difficult or quite easy to the student, depending upon the content of the stimuli. From the Johnson and Stratton methodology, we cannot ascertain whether they followed either of these extremes or whether they reached a desirable medium in test construction. For most of our criterion tests, the performance task will be closely related to the terminal task (i.e., the stimulus and response) in the associated instructional module, but employing different items of approximately equivalent difficulty where appropriate.

The experimental tests. Stemming from theories underlying the WORD RELATIONS and WORD LINE modules, we are developing three experimental measures that we hope will reflect differences in what is learned and how it is remembered, depending upon the instructional method.

Rumelhart, Lindsay, and Norman (1972) and others (Fillmore, 1968; Kintsch, 1972; Schank, 1970) have proposed a process model for long term memory in which information is stored as a semantic network. In their model, the network of concepts is interconnected by relations. Meanings are given through the relations which hold among the various concepts. Rumelhart, Lindsay, and Norman propose that every definable piece of information in the memory system, then, is encoded in the format of a node and its relations. Together all these nodes and relations form the total body of knowledge that a person has at any given time.

Our WORD RELATIONS module is based upon an application of their model. In WORD RELATIONS, the target word (a node) is presented along with another word that is related in some way (another node). From a multiple choice selection, the student determines what the relationship is between the two words. Our exercises then are a directed pathfinding strategy between our target word and other words that would likely be nearby nodes in the memory representation. Because the criterion test that accompanies the WORD RELATIONS module (where, similar to the exercises in the teaching sequence, the students identify relationships between words) is based upon this model of memory, we are considering that criterion test also to be an experimental test. It will be of great interest to us to determine whether students receiving instruction through the WORD RELATIONS module reflect a difference in cognitive structure from those students receiving instruction through the other modules.

In conjunction with our WORD LINE module, also based upon the theory that target words should be learned in conjunction with other words that comprise a structure of the subject matter, we are devising two experimental tests that will attempt to measure the residue structure in a student's memory after learning. The correspondence between a subject matter structure and its representation in the student's memory can then be determined. Shavelson and Stanton (1975) presented three methods for examining representations of a subject matter structure in students' memories; we have adopted two of their methods, word association and card sorting to use as experimental tests. The word association test will present our target words and synonyms of the target words and will measure the associative overlap between the target words and their synonyms. For the card sorting test, the students will sort concepts presented on cards into piles along certain boundaries yet to be established.

We hope that these experimental tests will help us to evaluate the extent to which our teaching methods in WORD RELATIONS and WORD LINE communicate the structure they are being developed to communicate.

Summary of the Research Endeavor to Date

The main accomplishments to date of this research endeavor are summarized within the present paper. They include:

1. A formalization of the decision theory framework that guides this research;
2. Descriptions of the seven $M_i$ at a general level and a rationale for each;

3. The introduction of a cognitive psychology perspective on the design research so as to aid understanding of learning;

4. The rationale and design of two of the $M_i$; and

5. The rationale and design of the pretest.

The main tasks that have yet to be accomplished include:

1. Completion of the design of the $M_i$.

2. Development of the instructional database.

3. Design of the criterion tests.

4. Design of the experimental tests.

5. Completion of the design of the computer based instructional research system.

Work is currently in progress on the above tasks.

## Reference Notes

1. Block, K. K. Making teaching of concepts more efficient. University of Pittsburgh, Learning Research and Development Center. (Research Grant Proposal Submitted to the Defense Advanced Research Projects Agency, February 1977; Contract awarded June 1977, No. MDA903-77-C-0208, ARPA Order No. 3395)

2. Voss, J. F. Cognition and instruction: Toward a cognitive theory of learning. Paper presented at the meeting of the NATO Conference, Amsterdam, June 1977.

## References

Beck, I. L., & McCaslin, E. S. An analysis of dimensions that affect the development of code-breaking ability in eight beginning reading programs. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, in press.

Becker, W. C., Englemann, S., & Thomas, D. R. Teaching: A course in applied psychology. Chicago: Science Research Associates, 1971.

Block, K. K. Vocabulary development: A problem in learning and instruction. International Reading Association Monograph Series, in press.

Bloom, B. S., (Ed.). Taxonomy of educational objectives. The classification of educational goals. Handbook I: Cognitive domain. New York: Longman, 1956.

Bourne, L. E. Human conceptual behavior. Boston: Allyn and Bacon, 1966.

Collins, A. M., & Quillian, M. R. Retrieval times from semantic memory. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 240-247.

Craig, R. C. The transfer value of guided learning. New York: Teachers College, Columbia University, 1953.

Cronbach, L. J. An analysis of techniques for diagnostic testing. Journal of Educational Research, 1942, 36, 206-217.

Cronbach, L. J. Measuring knowledge of precise word meanings. Journal of Educational Research, 1943, 36, 528-534.

Ellis, H. C. The transfer of learning. New York: Macmillan, 1965.

Ellson, D. G. Tutoring. In N. L. Gage (Ed.), The psychology of teaching methods. Chicago: The National Society for the Study of Education, 1976.

Eustace, B. W. Learning a complex concept at differing hierarchical levels. Journal of Educational Psychology, 1969, 6, 440-452.

Ferguson, R. L., & Hsu, T. The application of item generators for individualizing measurement. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1971. (Publication No. 1971/14)

Fillmore, C. J. The case for case. In E. Bach & R. G. Harms (Eds.), Universals in linguistic theory. New York: Holt, 1968.

Frase, L. T. Prose processing. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 9). New York: Academic Press, 1975.

Frase, L. T. The demise of generality in measurement and research methodology. Paper presented at the Center for the Study of Evaluation, Winter Invitational Conference on Measurement and Methodology, Los Angles, January 1978.

Frase, L. T., & Silbiger, F. Some adaptive consequences of searching for information in a text. American Educational Research Journal, 1970, 7, 553-560.

Gage, N. L., & Berliner, D. C. Educational psychology. Chicago: Rand McNally, 1975.

Gagné, R. M. The conditions of learning (3rd ed.). New York: Holt, Rinehart, & Winston, 1977.

Glaser, R. Concept learning and concept teaching. In R. M. Gagné & W. J. Gephart (Eds.), Learning research and school subjects. Itasca, Illinois: R. E. Peacock Publishers, 1968.

Humes, A. A design for a vocabulary development program (SWRL Technical Note No. TN-3-76-09). Los Alamitos, California: Southwest Regional Laboratory, 1976.

Johnson, D. M., & O'Reilly, C. A. Concept attainment in children: Classifying and defining. Journal of Educational Psychology, 1965, 55, 71-74.

Johnson, D. M., & Stratton, R. P. Evaluation of five methods of teaching concepts. Journal of Educational Psychology, 1966, 57, 48-53.

Kintsch, W. Notes on the structure of semantic memory. In E. Tulving & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Klausmeier, H. J., Ghatala, E. S., & Frayer, D. A. Conceptual learning and development: A cognitive view. New York: Academic Press, 1974.

Klein, K., & Saltz, E. Specifying the mechanisms in a levels-of-processing approach to memory. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 671-679.

Kučera, H., & Francis, W. N. Computational analysis of present-day American English. Providence, Rhode Island: Brown University Press, 1967.

Markle, S. M. They teach concepts, don't they? Educational Researcher, 1975, 4, 3-9.

Melton, A. W. (Ed.) Categories of human learning. New York: Academic Press, 1964.

Merrill, M. D., Olsen, J. B., & Coldeway, N. A. Research support for the Instructional strategy Diagnostic Profile (Courseware Technical Report No. 3). Orem, Utah: Courseware, Inc., 1976.

Merrill, M. D., & Tennyson, R. D. Teaching Concepts: An instructional design guide. Englewood Cliffs, New Jersey: Educational Technology Publications, 1977.

Merrill, M. D., & Wood, N. D. Rules for effective instructional strategies (Courseware Instructional Design Series). Orem, Utah: Courseware, Inc., 1975.

Miller, G. A. English verbs of motion: A case study in semantics and lexical memory. In A. W. Melton & E. Martin (Eds.), Coding processes in human memory. New York: John Wiley & Sons, 1972.

Miller, G. A., & Johnson-Laird, P. N. Language and perception. Cambridge: Harvard University Press, 1976.

Resnick, L. B., with Ford, W. W. The analysis of tasks for instruction: An information-processing approach. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1976. (Publication No. 1976/4)

Rumelhart, D. E., Lindsay, P. H., & Norman, D. A. A process model of long-term memory. In E. Tulving & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Schank, R. Semantics in conceptual analysis. Stanford Artificial Intelligence Project Memo AIM-122, 1970.

Shavelson, R. J., & Stanton, G. C. Construct validation: Methodology and application to three measures of cognitive structure. Journal of Educational Measurement, 1975, 12, 67-85.

Smith, E. E., Rips, L. J., & Shoben, E. J. Semantic memory and psychological semantics. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 8). New York: Academic Press, 1974.

Thorndike, E. L., & Lorge, I. The teacher's word book of 30,000 words. New York: Columbia University Teachers College Press, 1944.

Underwood, B. J., & Postman, L. Extra-experimental sources of interference in forgetting. Psychological Review, 1960, 67, 73-95.

Werner, O. H. The influence of the study of modern foreign language on the development of desirable abilities in English. Studies in Modern Language Teaching, 1930, 17, 97-145.

# APPENDIX A

## Numerical Examples of the Decision Theory Framework

### EXAMPLE 1

**Three Outcomes** $(O_1, O_2, O_3)$

| | | | |
|---|---|---|---|
| Cost | $c_1 = \$12$ | $c_2 = \$12$ | $c_3 = \$12$ |
| | $\Delta_1 = .75$ | $\Delta_2 = .70$ | $\Delta_3 = .73$ |
| | $\sigma_{12} = .05$ | $\sigma_{21} = .04$ | $\sigma_{32} = .05$ |
| | $\sigma_{13} = .02$ | $\sigma_{23} = .04$ | $\sigma_{31} = .02$ |

| | | 1 | 2 | 3 | Maximum | Minimum | Average |
|---|---|---|---|---|---|---|---|
| $M_1, M_2, M_3$ used: | levels | .82 | .78 | .80 | | .78 | |
| $M_1, M_2$ used: | levels | .80 | .74 | .07 | .80 | .07 | $\frac{1.61}{3}$ |
| $M_1, M_3$ used: | levels | .77 | .08 | .75 | .77 | .08 | $\frac{1.60}{3}$ |
| $M_2, M_3$ used: | levels | .07 | .74 | .80 | .80 | .07 | $\frac{1.61}{3}$ |
| $M_1$ used: | levels | .75 | .04 | .05 | .75 | .04 | $\frac{.84}{3}$ |
| $M_2$ used: | levels | .05 | .70 | .05 | .70 | .05 | $\frac{.80}{3}$ |
| $M_3$ used: | levels | .02 | .04 | .73 | .33 | .02 | $\frac{.79}{3}$ |

1. **Problem**: Budget \$30 or less

   1a. Find combination of $M_1, M_2, M_3$ which produces the maximum - minimum level.

   **SOLUTION**: Use $M_1, M_3$

   1b. Find combination of $M_1, M_2, M_3$, which produces highest average $\frac{1.61}{3}$ which is $M_1, M_2$, or $M_2, M_3$.

   1c. Maximize maximum $\rightarrow\rightarrow\rightarrow\rightarrow\rightarrow$ $M_1, M_2$, or $M_2, M_3$

## Decision Theory Framework

### EXAMPLE 2

**Three Outcomes** ($O_1$, $O_2$, $O_3$)

| | | | |
|---|---|---|---|
| Cost | $c_1 = \$12$ | $c_2 = \$12$ | $c_3 = \$12$ |
| | $\Delta_1 = .30$ | $\Delta_2 = .29$ | $\Delta_3 = .29$ |
| | $\sigma_{12} = .29$ | $\sigma_{21} = .29$ | $\sigma_{31} = .27$ |
| | $\sigma_{13} = .28$ | $\sigma_{23} = .30$ | $\sigma_{32} = .28$ |

| | | 1 | 2 | 3 | Minimum | Maximum | Average | Cost |
|---|---|---|---|---|---|---|---|---|
| $M_1$, $M_2$, $M_3$ used: | levels | .87 | .91 | .74 | .74 | .91 | $\frac{2.52}{3}$ | \$36 |
| $M_1$, $M_2$ used: | levels | .59 | .61 | .55 | .55 | .61 | $\frac{1.75}{3}$ | \$24 |
| $M_1$, $M_3$ used: | levels | .58 | .59 | .56 | .56 | .59 | $\frac{1.73}{3}$ | \$24 |
| $M_2$, $M_3$ used: | levels | .57 | .62 | .57 | .57 | .62 | $\frac{1.76}{3}$ | \$24 |
| $M_1$ used: | levels | .30 | .29 | .27 | .27 | .30 | $\frac{.86}{3}$ | \$12 |
| $M_2$ used: | levels | .29 | .32 | .28 | .28 | .37 | $\frac{.89}{3}$ | \$12 |
| $M_3$ used: | levels | .28 | .30 | .29 | .28 | .30 | $\frac{.87}{3}$ | \$12 |

**\$30 or less:** **Solutions**

1a) Maximize minimum → $M_2$, $M_3$
1b) Maximize average → $M_2$, $M_3$
1c) Maximize maximum → $M_2$, $M_3$

**Cost Effectiveness Given By:**

$$\frac{\text{Minimum}}{\text{Cost}} \text{ or } \frac{\text{Maximum}}{\text{Cost}} \text{ or } \frac{\text{Average}}{\text{Cost}}$$