Conceptually based development and evaluation of individualized materials for critical reading based on logical inference*

GEORGE F. LOWERRE
Research for Better Schools
and
JOSEPH M. SCANDURA
University of Pennsylvania

A SYSTEMATIC DIMENSIONAL ANALYSIS technique for analyzing complex tasks was developed. It was applied to the critical reading of paragraphs which require the use of logical inference rules. Five variables or dimensions were identified to specify the level of difficulty of a particular reading passage. For example, the dimension "complexity of context (including length)" can take 3 values: 1) single implication—message contains only relevant statements; 2) more than one implication—message contains only relevant statements; 3) more than one implication—message contains 3 to 5 extra statements, some of which may appear relevant. For any given logical inference rule, it was possible to determine the exact difficulty (with respect to each of the 5 dimensions) of passages in which a subject could or could not make the correct inference. Prototype materials were developed in which 2 dimensions were varied; these tested the subjects' ability to use 2 logical inference rules. The materials were used successfully in diagnostic testing and training for elementary school children reading at the third and fourth grade level.

Matériaux individualisés pour servir à la lecture critique fondée sur un procédé logique d'induction: développement et évaluation conceptuels

UNE TECHNIQUE METHODOLOGIQUE d'analyse dimensionelle a été développée pour l'étude de tâches complexes. Cette analyse a été appli-

* This report is based on a dissertation by the first author under the supervision of the second author and submitted to the University of Pennsylvania as partial fulfillment of the requirements for the degree of Doctor of Philosophy. This research was supported in part by NSF Grant GW 6796 to the second author. The materials described herein have since been expanded into a "Critical Reading Series" by members of the Mathematics Education Research Group (MERG) for production and distribution by Ann Arbor Publishers, Inc. P.O. Box 388, Worthington, Ohio 43085.

Reprinted from
READING RESEARCH QUARTERLY, Volume IX, Number 3
1973-1974

Published by the
International Reading Association, Newark, Delaware
Copyright 1973 by the International Reading Association
Desarrollo y evaluación conceptual de materiales individualizados para la lectura crítica basada en la inferencia lógica

Una técnica de análisis sistemático y dimensional ha sido desarrollada para analizar tareas complejas. Esta se aplicó a la lectura crítica de párrafos que requieren el uso de reglas de inferencia lógica. Se identificaron cinco variables o dimensiones para determinar el nivel de dificultades que presentaba un pasaje de lectura específico. Por ejemplo, la dimensión "complejidad del contexto (incluyendo la extensión)" puede involucrar 3 valores: 1) implicación singular —el mensaje sólo contiene declaraciones relevantes; 2) más de una implicación —el mensaje sólo contiene declaraciones relevantes; 3) más de una implicación —el mensaje contiene de 3 a 5 declaraciones adicionales, algunas de las cuales pueden parecer relevantes. Para cualquiera de las reglas de inferencia lógica dadas, fue posible determinar exactamente la dificultad (con respecto a cada una de las 5 dimensiones) de los pasajes en los cuales un sujeto podía, o no, hacer la inferencia correcta. Se desarrollaron materiales prototípicos en los que se variaron 2 dimensiones; éstos se aplicaron para probar la aptitud de los sujetos para utilizar 2 reglas de inferencia lógica. Los materiales se utilizaron con éxito en las pruebas de diagnóstico y entrenamiento para la lectura de los niños de las escuelas primarias en los niveles de tercero y cuarto grado.
Introduction to the problem

Most people know how to read. They can translate statements composed of graphemes (symbols) into phonemes (sounds), and they know what the statements mean. There are those, however, who are unable to think critically about written discourse. One major limitation is the inability of some people to think logically about what they read. They cannot determine which statements in a “message” must follow logically from other statements, they cannot detect incorrect applications of a logical rule, and/or they are unable to see that some statements in a message have logical implications that are directly contrary to other statements in the same message.

The present study was designed to attack the problem of producing materials for diagnostic testing and for teaching in this specific area of critical reading—logical, deductive reasoning. The task was threefold:

1. To develop a systematic procedure (systematic dimensional analysis) for constructing diagnostic tests and instructional materials for dealing with specific logical inference rules.
2. To apply the above procedure to develop prototype tests and materials, involving 2 logical inference rules, for use by children reading at a third grade level or higher.
3. To evaluate these prototype tests and materials and to test the assumptions upon which the development procedure was based.

Background and related research

Many writers and researchers have attempted to list the skills involved in critical thinking (including the critical thinking that should take place while reading).¹ Some have developed long lists of skills (see, for example, McCanne, 1969; Williams, 1959; Wolf, King, and Huck, 1968). Others have offered shorter lists (see, for example, Ennis, 1962; Spache and Spache, 1969). Although these lists contain a wide variety of skills, all of them include skills in-

¹ Some writers speak of critical thinking; others use the term critical reading. We agree with Moffett (1968) that critical reading is actually critical thinking about written discourse. The ability to read per se is a prerequisite to the ability to read critically.
volving the ability to reason deductively. Thus, while experts in the field of critical thinking (and critical reading) cannot agree on a definition of exactly what it is, they do agree that logical reasoning skills are an important part of critical thinking.

Fortunately, the number of basic logical inference rules is relatively small; the others are all combinations of inference rules from the basic set. Corcoran (1968), for example, has given 24 logical inference rules which are sufficient for all proofs in first order logic. While it may not be desirable pedagogically to teach all of these rules in the form in which Corcoran presents them, they demonstrate the relatively small size of the basic set of logical inference rules.

Experimental testing has shown that many children have not mastered all of the basic inference rules. Scandura and McGee (1972), for example, found that 12 out of 16 kindergarten children could not use the rules of inference governing the words some and all.

Other research has tested the use of logical rules by somewhat older children. Suppes (1965) cites a study by Hill in which she found that a group of 8 year olds averaged 86 correct responses on a 100 item test requiring the use of logical rules. Her test, however, did not include any items where no valid conclusion could be drawn. O'Brien and Shapiro (1968) made her test more difficult by adding "Not Enough Clues" as a possible response on each item, and by making that the correct response for one-third of the items. They found that the average score dropped to 56 out of 100 for their group of 8 year olds. For items whose correct response was "Not Enough Clues" the average score dropped to approximately the chance level.

Ennis, Finkelstein, Smith, and Wilson (1969)—using only conditional statements—obtained results similar to those of O'Brien and Shapiro. They found that 62 per cent of their third grade subjects had mastered the rule for contrapositives of conditional statements. But only 45 per cent of their third grade subjects had mastered the rule for transitivity of conditionals, 31 per cent had mastered the

2. First order logic includes the logic of propositions (sentential logic) and the use of quantifiers ("for all" and "there exists") applied to members of the relevant universe. It was not felt that second order logic, in which quantifiers are applied to relations and functions, is relevant to the elementary school. Corcoran's list of 24 logical inference rules can be shortened by eliminating some of the logical connectives and defining them in terms of other logical connectives.

3. The Hill and the O'Brien and Shapiro studies each also included younger children. Only the results for the 8 year olds in both studies are quoted here.

4. Ennis, et al. also included younger children in their study. Only the results for the third graders are quoted here.
rule for using inverses of conditionals, and 7 per cent had mastered the rule for using converses of conditionals.⁵

It appears that Hill's data result from the type of test item she used, rather than from real mastery of logical rules by most third graders. The evidence is that most 8 year olds have not mastered the common rules of logical inference.

On the other hand, several earlier studies have demonstrated that it is possible to teach logic to elementary school children. Although Ennis, Finkelstein, Smith, and Ennis (1969) were not successful in teaching specific rules for conditional logic to elementary school children, Hyram (1957), Suppes and Binford (1965), and Wolf, King, and Huck (1968) were all successful to some extent in improving overall performance on a general logic test.

None of the studies in which logic was successfully taught combined teaching and diagnostic testing in a unified way. Subjects were not tested before training to determine exactly which logical rules they knew and in what contexts they could use them. Similarly, the post tests in these studies gave only a single overall score, and did not indicate which logical inference rules or types of rules each subject had or had not mastered. For example, the post test scores in these studies did not distinguish between the use of logical inference rules involving the quantifiers "for all" and "there exists" and the use of logical inference rules not involving quantifiers.⁶

**Systematic dimensional analysis**

The prototype materials developed and evaluated in the present study were designed to avoid the limitations of the above studies.

---

⁵ A conditional statement is one of the form: "If Fido is a dog, then Fido has a tail." The converse is "If Fido has a tail, then Fido is a dog." The inverse is "If Fido is not a dog, then Fido does not have a tail," and the contrapositive is "If Fido does not have a tail, then Fido is not a dog." If a conditional statement is true, then its contrapositive must also be true. But the converse and inverse of a true conditional statement may be either true or false. (For the example given, they are both false since Fido may be a cat with a tail.)

⁶ An example of use of a logical inference rule involving the quantifier "for all" is the following:

- All dogs have tails.
- Fido is a dog.
- Therefore, Fido has a tail.

An example of use of a logical inference rule not involving use of a quantifier is the following:

- Fido is a dog or a cat.
- Fido is not a dog.
- Therefore, Fido is a cat.

These are examples of the 2 logical inference rules used in this study.
1. They include a pretest for each rule to determine whether and in which reading contexts each child can use it.
2. They base the initial instruction for each child directly on his performance on the pretest.
3. They permit each child to study the use of each rule in a more difficult setting only after he has mastered its use in simpler settings.
4. They give post test scores which indicate the contexts in which each child can use each rule tested.

The systematic procedure used in this study to develop the prototype diagnostic testing and teaching materials was adapted from ideas set forth in Scandura (1968). Each logical rule determines a class of stimulus-response pairs, where the stimuli consist of reading materials in which the logical rules can be applied, and the responses are logical implications of these reading materials. These reading materials may vary in many ways. Adaptation of Scandura's ideas required the identification of dimensions over which such reading materials may vary, together with levels of difficulty along each dimension. The difficulty levels depend on the amount of information which must be processed and how consistent the information is with what is commonly known.

Five dimensions were identified, and levels of difficulty along each of them were determined. These are given in Figure 1. A serious attempt was made to make these dimensions exhaustive and essentially independent. It should be emphasized that these dimensions were determined by strictly analytical means. There was no use of factor analysis or other statistical procedures.

For each logical inference rule it is possible to construct reading passages at predetermined levels of difficulty along each dimension. For example, a passage might be at level 2 of dimension A, level 3 of dimension B, and so on. The testing procedure depends on the assumption that success in using a logical rule at any level of difficulty along any dimension implies success at the less difficult level(s). Conversely, failure at any level of difficulty along a dimension is assumed to imply failure to use the logical rule correctly at the more difficult level(s).

Studies by Scandura, Woodward, and Lee (1967) and Scandura and Durnin (1968) lend support for these assumptions. They
Dimension A: Relation of statements in message to reality

<table>
<thead>
<tr>
<th>Difficulty Levels</th>
<th>1. Statements which agree with commonly known facts</th>
<th>2. (Neutral) Statements which neither agree nor contradict commonly known facts</th>
<th>3. Statements which contradict commonly known facts</th>
</tr>
</thead>
</table>

Dimension B: Complexity of context (including length)

<table>
<thead>
<tr>
<th>Difficulty Levels</th>
<th>1. Simple: single implication; message contains only relevant statements</th>
<th>2. More than one implication; message contains only relevant statements</th>
<th>3. More than one implication; message contains 3 to 5 extra statements, one or 2 of which may appear to be relevant</th>
</tr>
</thead>
</table>

Dimension C: Availability of premises in message

<table>
<thead>
<tr>
<th>Difficulty Levels</th>
<th>1. All relevant premises present and clearly stated</th>
<th>2. Nuance: premise determined from context</th>
<th>2. Some premise is missing but implied by the context</th>
</tr>
</thead>
</table>

Dimension D: Required length of chain inference

<table>
<thead>
<tr>
<th>Difficulty Levels</th>
<th>1. Single rule application</th>
<th>2. Two rule applications</th>
<th>3. Three rule applications</th>
<th>4. Four rule applications</th>
<th>...</th>
</tr>
</thead>
</table>

Dimension E: Terminology used

<table>
<thead>
<tr>
<th>Difficulty Levels</th>
<th>1. Most common English terminology (e.g., If A, then B)</th>
<th>2. Variations from common terminology (e.g., A only if B, or B is necessary for A)</th>
</tr>
</thead>
</table>

Figure 1
Dimensions over which reading settings may vary with difficulty levels along them (No ordering of difficulty among dimensions implied.)

studied rules which define classes of tasks and which may be thought of as varying in generality along ordered dimensions. Success in using a general rule (which was more difficult to learn) implied success in using a form of the rule which was specifically restricted along one or more dimensions (easier to learn). Similarly, failure to use the restricted rule correctly implied failure to use the more general rule. It seems reasonable that similar results should be obtained using logical rules where the reading contexts (which define classes of tasks) vary in difficulty along ordered dimensions.
Part of the experimental procedure involved testing of the assumptions on which the testing procedure was based. Using these assumptions, one may determine the most difficult level at which a child can use a logical rule along one dimension (holding the difficulty levels along the other dimensions constant) as follows:

1. Test for use of the logical inference rule at the least difficult level on that dimension.
2. Test for use of the logical inference rule at the most difficult level on that dimension.
3. Give test items which require use of the logical inference rule at successively less difficult levels along that dimension.

For a single dimension with n levels of difficulty, this testing procedure may be diagrammed as follows:

<table>
<thead>
<tr>
<th>Level of Difficulty</th>
<th>Simplest</th>
<th>Slightly less simple</th>
<th>. . .</th>
<th>Slightly less difficult</th>
<th>Most difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Item Number</td>
<td>1</td>
<td>n</td>
<td>. . .</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

When a child can solve test items at one level, the first assumption allows us to conclude that he can do all the problems requiring use of the logical rule at less difficult levels along the dimension being tested. Hence, when one tests in the sequence above and a child succeeds on the test items at any level beyond the simplest one, further testing along that dimension is unnecessary. Also, failure on the test items at the simplest level implies, on the basis of the second assumption, that the child cannot use the logical rule at the other levels, so that testing stops. Finally, success on the test items at the simplest level followed by failure on the test items at all other levels, implies that the child can use the logical rule only in the simplest setting along that dimension.

This procedure works with any logical inference rule, and may be extended to test for use of the rule along 2 or more dimen-
sions. To determine the most difficult levels at which a child can apply a given logical rule along two or more dimensions, the procedure is as follows:

1. Hold the level of difficulty on all dimensions fixed but one, and test as just described to find the maximum level of difficulty at which the child can apply the logical rule along that dimension.

2. Hold the level of difficulty on the first dimension fixed at the level just determined, and repeat the procedure with the second dimension. (The level of difficulty on the first dimension is held fixed at the level just determined because the dimensions may not be independent for a child who is applying a logical rule. An attempt was made to get dimensions which are independent for purposes of classifying reading settings, but no claim is made regarding their behavioral independence.)

3. Repeat the process, one dimension at a time, to determine the maximum level at which the child can apply the logical rule on succeeding dimensions.

Instruction on a logical inference rule begins after the most difficult levels, along each dimension at which a child can use that logical rule, have been determined. The prototype instructional materials developed for this study, like the testing materials, use the systematic dimensional analysis based on Scandura (1968). Specifically, the instructional materials were designed to start at the levels at which the child can use a logical rule, and gradually to increase the difficulty level along one dimension at a time.

A child may be said to know how to use a logical rule when he can apply it in 3 basic ways: 1) to make or recognize a valid inference, 2) to detect logical incompatibilities, and 3) to detect invalid uses of the rule. The prototype materials were designed to test and teach all 3 applications of each logical rule.

**Prototype materials**

The prototype materials developed for this study were based on only 2 logical rules, and the stimuli reading materials were allowed to vary over only 2 of the 5 dimensions. The 2 logical inference rules may be represented:
Rule for “Or” Elimination
From the premise P or Q
and the premise not P (or not Q)
we conclude Q (or P).

Rule for “All” Elimination
From the premise All P’s are Q’s (have the property Q)
and the premise x is a P
we conclude x is a Q (has the property Q).

The materials developed for the “all” elimination rule include sentences using a pattern which appears similar to the valid one above:

All P’s are Q’s (have the property Q)
x is a Q (has the property Q).

But in this case, no valid conclusion is possible. Any child who draws the invalid conclusion, “x is a P,” is said to have used the fallacy of “all” conversion.8

These 2 rules were chosen because they are among the most basic logical inference rules and because one of them (“all” elimination) involves the use of a quantifier and the other does not. The materials were designed for use by children reading at the third or fourth grade levels.

Dimensions B (Complexity of Context) and C (Availability of Premises) were allowed to vary. The testing procedure varied only one of these dimensions at a time. Dimension A was held fixed at level 2 (the “neutral” level) and dimensions D and E were held fixed at level 1 (the simplest level).

The following type of test item was used. The child was instructed to circle TRUE if the statement must be true (based on the preceding paragraph), to circle FALSE if the statement must be false (based on the preceding paragraph), and to circle DON’T KNOW if the paragraph did not give enough information to determine whether the statement is true or false. A sample item follows:

7. The names used for different logical rules vary among authors. These names are a variation of the relatively simple naming system used by Corcoran (1968).
8. This name is adapted from Ennis, Finkelstein, Smith, and Wilson (1969). They consider use of the converse of a conditional to be the use of a separate (invalid) rule. In these materials, the fallacy of “all” conversion is considered to be an invalid application of the rule for “all” elimination.
All pro football linemen weigh over 200 pounds. Tom Smith weighs over 200 pounds. Fred Jones is a pro football lineman. Paul Franks is a pro football lineman, too.

A. Tom Smith is a pro football lineman.  
     TRUE   FALSE   DON'T KNOW
B. Paul Franks weight over 200 pounds.  
     TRUE   FALSE   DON'T KNOW
C. Fred Jones weighs less than  
     200 pounds.  
     TRUE   FALSE   DON'T KNOW
D. Tom Smith lives near Paul Franks.  
     TRUE   FALSE   DON'T KNOW
E. Fred Jones weighs over 200 pounds.  
     TRUE   FALSE   DON'T KNOW

Such an item tests a child's ability to apply a logical rule in all 3 basic ways:

1. Sentences which should be marked TRUE (B and E in the sample item) test his ability to make or to recognize a valid inference.
2. Sentences which should be marked FALSE (C in the sample item) test his ability to detect statements which are incompatible with the rest of the message.
3. Sentences which should be marked DON'T KNOW (A and D in the sample item) test his ability to recognize statements which, while not incompatible with the rest of the message, do not follow from it by use of the logical rule.⁹ (All of the test items were written so that such sentences did not follow from the use of any other logical rule.)

Two tests were written for each rule with 12-16 items at each pair of difficulty levels. They were made as nearly equivalent as possible. Both used essentially the same stories; but names, colors, dates, and order of sentences were changed. Each test was 20 pages long, including 2 practice pages. The pages were grouped in pairs, with all problems on each pair of pages being at the same level along both dimensions. At each level on the "all" elimination tests there were at least 3 sentences (responses) which could be marked using the similar appearing, but invalid fallacy of "all" conversion. Mas-

⁹. It must be recognized that this response may be chosen not only in the case where a child recognizes an invalid application of the logical rule, but also in the case where he does not know whether a logical rule applies or not. Complete inability to use a logical rule will result in correct responses in cases where DON'T KNOW is the correct answer.
tery, or criterion, at each (pair of) levels was set at a maximum of 2 errors out of a total of 12 to 16 responses.\textsuperscript{10}

The instructional materials consisted of 10 workbooks for each logical rule, with a pre-recorded cassette tape accompanying each workbook. The tapes provided hints and diminishing amounts of help as the instruction progressed, together with answers for checking. The workbooks were from 5 to 7 pages long, and they required from 37 to 62 student responses. The workbooks were sequenced as follows:

<table>
<thead>
<tr>
<th>Workbook Number</th>
<th>Entering Level</th>
<th>Level After Completing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2</td>
<td>0.0\textsuperscript{11}</td>
<td>1.1</td>
</tr>
<tr>
<td>3, 4</td>
<td>1.1</td>
<td>2.1</td>
</tr>
<tr>
<td>5, 6</td>
<td>2.1</td>
<td>3.1</td>
</tr>
<tr>
<td>7, 8</td>
<td>3.1</td>
<td>3.2</td>
</tr>
<tr>
<td>9, 10</td>
<td>3.2</td>
<td>3.3</td>
</tr>
</tbody>
</table>

The even numbered workbooks were used as backups for the child who had not mastered the use of the rule at the higher level by the time he had completed the preceding odd-numbered workbook. For example, he used workbook 6 only if he had not mastered use of the rule at level 3-1 by the end of workbook 5.

The starting workbook for each child was determined by the pretest. These were as follows:

<table>
<thead>
<tr>
<th>Maximum Level at which child demonstrated mastery on pretest</th>
<th>Beginning workbook</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>3</td>
</tr>
<tr>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>5</td>
</tr>
<tr>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>7</td>
</tr>
<tr>
<td>3.2</td>
<td>9</td>
</tr>
<tr>
<td>3.3</td>
<td>none</td>
</tr>
</tbody>
</table>

\textsuperscript{10} It was felt that this would allow a large enough number of careless mistakes for a child who could really use the rule at the given level. It was also small enough that any child who systematically used the invalid fallacy of “all” conversion, and who therefore did not fully understand the rule for “all” elimination, would not meet criterion.

\textsuperscript{11} The first number represents the level of dimension B; the second number represents the level of dimension C.
The child then progressed through the workbooks in numerical order, skipping the even-numbered workbooks whenever possible. Notice that some of the starting workbooks were grouped (for levels 1, 2, and 3 of dimension C). This was done in order to keep the number of workbooks relatively small. Although it would have been possible to provide a separate starting point for each pair of levels on dimensions B and C, this would have increased the total number of workbooks to a level which was felt to be impractical for school use.

The instructional materials utilized a guided discovery format. That is, each workbook contained marked examples. Spoken comments on the corresponding tape explained the examples and gave relevant ideas. Then the child did some examples. These were explained on the tape, more examples were done and the answers explained and so on. The tapes were completely self-contained. They included directions for how to mark sentences, when to stop and start the tape player, and what workbook to use next. Once a child knew how to operate the tape player, he received all of his directions, as well as explanations of the correct responses for the workbook sentences, from the tape and he had no need for outside help.

All of the written materials, both the tests and the workbooks, were tested for readability using the Spache (1953) readability formula with the Stone (1956) revised word list. Because these materials were designed for children reading at the third or fourth grade levels, a cut-off level of 2.5 was used. Any test items that gave a reading level higher than 2.5 were rewritten or deleted.

No formal method was used to evaluate the language used on the tapes, although an effort was made to keep the vocabulary as simple as possible.

**Empirical test: subjects, experimental design, and procedure**

The empirical study was designed primarily to test the efficacy of the prototype materials. In addition, it was to test the basic assumptions underlying the testing procedure that success in using a logical rule at level 3 along either dimension implied success at level 2, and that failure at level 1 along both dimensions implied failure at all higher levels.

A simple pretest-post test control group design was used. The pretests for both rules were first administered to 45 children from 9 different classrooms in the second, third, and fourth grades in the
Henry C. Lea School in West Philadelphia. All of the children were volunteers reading at the third or fourth grade level. The reading levels had been determined by their teachers using a group informal reading inventory, following procedures outlined by the School District of Philadelphia for that purpose.

The pretest was given to groups ranging from 3 to 6 subjects each. There was no time limit for taking the test. After the maximum levels along both dimensions at which a child could use either logical rule had been determined, additional items from the test, which had not been required to determine the maximum levels along dimensions, were assigned to test the basic testing assumptions.

Each child took the pretest for the "or" elimination rule first and took the "all" elimination pretest between one and 5 days later.

Five of the original children were excluded from further participation for technical reasons. This left 40 subjects in the experiment. Of these, 17 scored at level 3 of both dimensions on the test for the rule for "or" elimination; none of the subjects scored at level 3 on both dimensions on the test for the rule for "all" elimination.

The 40 subjects were randomly assigned to the experimental and control groups. The experimental group had to be kept slightly smaller \( n = 19 \) than the control group \( n = 21 \) because of scheduling limitations at the Lea School. All of the experimental subjects who had not scored at level 3 on both dimensions received the instruction for the "or" elimination rule. All but 3 experimental subjects received instruction for the "all" elimination rule. (One of those 3 moved out of the school district; the other 2 were absent for an extended period of time.) No instruction was provided to any of the control subjects.

The experimental subjects received the instruction 2 at a time (except when one subject scheduled for a given time was absent). Each was furnished with a cassette tape player and workbook, and proceeded at his own pace. Most subjects received instruction 2 times per week, although a subject occasionally worked with the materials either one or 3 times in a given week. All instruction took place in carrels in the instructional materials center at the Lea School.

The subjects worked by themselves. The experimenter was available to handle unforeseen difficulties. Any subject who failed to meet criterion at any level both on the original (odd-numbered) workbook, and on the back-up (even numbered) workbook, used the original workbook again. If he met criterion on that try, he continued in
the usual manner.\textsuperscript{12} If he again failed to meet criterion, instruction on that rule was halted.\textsuperscript{13}

Those experimental subjects who did not score at level 3 on both dimensions on the pretest for the "or" elimination rule, received instruction for the rule for "or" elimination first. Then they received the post test on that rule. Next they received the instruction on the rule for "all" elimination, followed by the post test for that rule. Those in the experimental group who scored at level 3 on each dimension on the "or" elimination pretest, went directly to instruction on the rule for "all" elimination.

The subjects in the control group were given the post tests at the same time as the last members of the experimental group were completing the post test on the "all" elimination rule. All subjects received the post tests on separate days. The elapsed time between pretests and post tests for the control group varied from 50 days to 64 days, with a median of 54 days.

\textit{Test validation results}

The test data were first analyzed to determine whether the assumptions on which the testing procedure was based were valid. In general these hierarchal assumptions were supported. Subjects who met criterion at level 3 along a dimension also tended to meet criterion at level 2 along that dimension. And subjects who failed at level 0-0 also tended to fail at all other levels tested.\textsuperscript{14}

There was one exception. In 10 out of 51 tests (counting pretests and post tests together) on the rule for "or" elimination, a subject succeeded at level 3 of dimension C but failed at level 2. Re-examination of the test items at level 3 of dimension C (regardless of the level of dimension B) showed that these items required the child to find a missing premise, but did not require him to use the "or" elimination rule after finding the missing premise. Therefore, all materials related to level 3 of dimension C for "or" elimination were discarded.

\textsuperscript{12} This happened twice (both times on the "all" elimination rule).
\textsuperscript{13} This happened 3 times (2 times with the "or" elimination rule and one time with the "all" elimination rule).
\textsuperscript{14} Because of time limitations, the only additional levels tested in this case were levels 1-2, 1-3, 2-1, and 3-1. For subjects who failed to meet criterion at all of these levels it seems safe to assume they would also fail to meet criterion at levels 2-2, 3-2, 2-3, and 3-3.
With level 3 of dimension C discarded, the testing assumptions were upheld in a total of 68 out of 72 instances (94 per cent) on the "or" elimination pretest and in 30 out of 33 instances (91 per cent) on the "or" elimination post test. The testing assumptions were upheld on the "all" elimination pretest in a total of 128 out of 152 instances (84 per cent), and on the "all" elimination post test in a total of 78 out of 83 instances (94 per cent).\(^{15}\)

Because the pretests and post tests each involve what is essentially a series of subtests, it is not appropriate to compute ordinary reliability coefficients. Indeed, there is, to our knowledge no accepted method for calculating such reliabilities on sequential tests. Nonetheless, for comparative purposes, reliability coefficients were computed for each subtest which was taken by at least 10 subjects. For this purpose, use was made of the Livingston (1972) criterion-referenced reliability formula, based on norm referenced reliabilities determined by Kuder-Richardson formula 20. Twenty reliability coefficients were computed. Eight ranged from .84 to .90; 7 from .77 to .81; 3 from .73 to .75; and 2 from .52 to .53. The last 2 reliabilities were unusually low because almost all of the subjects scored at the criterion level. (This results from an artifact of the formulas used.) Although quite high considering the lengths of the subtests, such reliabilities in sequential tests are of secondary importance relative to the above percentages concerned with hierarchical validity.

**Instructional results**

The results of the instruction are summarized in Tables 1 and 2. Table 1 compares the number in the experimental and control groups who improved their scores between the pretest and post test (that is, those who learned or expanded the use of each rule). For the "or" elimination rule, the differences in number of subjects who improved their scores approaches significance (.05 < p < .10). For the "all" elimination rule, the differences are highly significant (p < .0001).

\(^{15}\) Most of the cases where the testing assumptions were not upheld on the pretest involved level 3 of dimension C. However, at level 3 of dimension C on the pretest, the child had only to find the missing premise; he did not have to use the "all" elimination rule. This was changed on the post test. All items at level 3 of dimension C on the post test required the child to find the missing premise and then to use the "all" elimination rule to find a conclusion (or to recognize that no valid conclusion was possible). This accounts for the higher percentage on the "all" elimination post test.
Table 2 compares the changes in performance between pretest and post test. The change in performance for each subject was obtained by adding the changes along both dimensions. For example, a change from level 2-1 on the pretest to level 3-2 on the post test is a change of $+2$. According to this criterion, too, the experimental subjects learned significantly more, both for the "or" elimination rule ($p < .05$) and for the "all" elimination rule ($p < .0001$).

The differences associated with "or" elimination are not as large as those for "all" elimination. Many of the control subjects did better on the "or" elimination post tests than they did on the pretest, while very few control subjects did better on the "all" elimination post test than they did on the pretest. "Or" elimination seems to have

Table 1  Relative gains in expanding use of rules

<table>
<thead>
<tr>
<th>Pretest Level</th>
<th>Post test Level</th>
<th>&quot;Or&quot; Elimination</th>
<th>&quot;All&quot; Elimination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>n</td>
<td>P*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exper.</td>
<td>Control</td>
</tr>
<tr>
<td>0-0</td>
<td>Higher</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Same or lower</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1-1 or higher</td>
<td>Higher</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Same or lower</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>All levels</td>
<td>Higher</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Same or lower</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

* Using Fisher's exact test.
** Any subject who was able to use the rule at the highest level on the pretest was not included here.

Table 2  Comparison of change of level (gain) from pretest to post test

<table>
<thead>
<tr>
<th></th>
<th>&quot;Or&quot; Elimination</th>
<th>&quot;All&quot; Elimination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exper.</td>
<td>Control</td>
</tr>
<tr>
<td>n*</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>mean</td>
<td>2.50</td>
<td>1.23</td>
</tr>
<tr>
<td>S. D.</td>
<td>1.12</td>
<td>1.25</td>
</tr>
<tr>
<td>sum of ranks</td>
<td>86.5</td>
<td>189.5</td>
</tr>
<tr>
<td>p**</td>
<td>&lt; .025</td>
<td></td>
</tr>
</tbody>
</table>

* Any subject who was able to use the rule at the highest level on the pretest was not included here.
** Using a one-tailed Mann Whitney U test.

16. Because levels 0-1 and 1-0 do not exist (level 0-0 indicates complete inability to use the rule, and level 1-1 means the child can use the rule at the lowest level along each dimension), a change from level 0-0 to level 1-1 was considered a change of $+1$. 

---

202  READING RESEARCH QUARTERLY • NUMBER 2, 1973-1974 IX/2
been a relatively easy rule, and the data suggest that some subjects learned to use it while taking the pretest or during the pretest to post test period. "All" elimination was distinctly harder, probably because of the fallacy of "all" conversion, and therefore the effects of instruction were more apparent.

In summary, all of the experimental results favored the experimental group. Perhaps most important, every experimental subject scored better on both post tests than on the pretests (excepting, of course, those 9 subjects who scored at the top levels (3-2) on the "or" elimination pretest). This was not true for the control subjects.

Discussion and implications

Within the bounds set for these prototype materials—which dealt with only 2 of 5 dimensions and with only 2 logical rules—the systematic dimensional analysis was demonstrated to be a good technology. The testing assumptions were shown to be valid, and the tests were shown to be efficient instruments for measuring the reading contexts in which children can use a logical rule. Basing instruction directly on pretest performance and sequencing instruction by expanding reading contexts along one dimension at a time appears to be a sound procedure. As with all studies of this type, of course, the usual limitations apply (e.g., those pertaining to the particular subject population).

Further developmental work using these procedures to improve logical reasoning using written materials therefore seems justified. Such development should involve all 5 dimensions and a wider variety of logical inference rules. Indeed, a "Logic Series" based on these ideas has been published commercially for use with students reading at a third grade reading level or higher (See footnote 1, p. 188). Besides additional materials on general topics, as in the prototype materials, tapes and workbooks might be developed to accompany curriculum materials in particular subjects, such as mathematics or social studies. Additional research to answer 6 questions would aid such development:

1. How should logical rules be matched with grade or reading level for teaching purposes? For example, the data suggest that "or" elimination may be less appropriate for third and fourth graders than "all" elimination.
2. What is the role of fallacious rules, such as the fallacy of "all" conversion? Should fallacious rules be taught as misapplications of valid rules, or should separate materials be devoted just to them? Comparing the results of this study with the results of Ennis, Finkelstein, Smith, and Wilson (1969), where invalid converses and inverses of conditionals were taught as separate rules, it would appear that it is better to teach fallacious rules as misapplications of valid ones. But this is not clearly established.

3. More basically, what is it that allows children to distinguish between information that is relevant to the use of a logical rule, and information that is irrelevant? How do children distinguish between applications of valid rules (such as the rule for "all" elimination) and similar appearing situations which suggest the use of invalid rules (such as the fallacy of "all" conversion)?

4. What can be done to make materials of this kind more interesting? Although only a subjective observation, it appeared that some of the experimental subjects found the materials difficult and tiring. If different types of materials will improve motivation and/or interest, what form should they take? In answering this question, one must keep in mind that simple scoring would appear to be essential in order to maintain the self-instructional feature.

5. How can one be sure that a child will use a logical rule which he has mastered in a paper-and-pencil test situation in non-test situations? Out of class use of what is learned in class is, of course, the goal of all education. Can methods be developed for measuring the use of logical rules in other classroom subjects and in non-school reading and other activities?

6. Finally, it should be noted that the use of the inductive method to teach the use of logical rules with written materials requires a wide variety of stories. If efficiency is to be increased further, it will be necessary to identify more precisely the various underlying rules. (Specifying the general form of each logical rule is just a first step in this direction.) To the extent that this can be accomplished, the present inductive approach can be replaced by more direct instruction.
REFERENCES


